Prioritizing 2nd and 3rd order interactions via support vector ranking using sensitivity indices on static Wnt measurements †

shriprakash sinha*

It is widely known that the sensitivity analysis plays a major role in computing the strength of the influence of involved factors in any phenomena under investigation. When applied to expression profiles of various intra/extracellular factors that form an integral part of a signaling pathway, the variance and density based analysis yields a range of sensitivity indices for individual as well as various combinations of factors. These combinations denote the higher order interactions among the involved factors that might be of interest in the working mechanism of the pathway. For example, DACT3 is known to be a epigenetic regulator of the Wnt pathway in colorectal cancer and subject to histone modifications. But many of the n\textsuperscript{th} ≥ 2 order interactions of DACT3 that might be influential have not been explored/tested. In this work, after estimating the individual effects of factors for a higher order combination, the individual indices are considered as discriminative features. A combination, then is a multivariate feature set in higher order (≥ 2). With an excessively large number of factors involved in the pathway, it is difficult to search for important combinations in a wide search space over different orders. Exploiting the analogy of prioritizing webpages using ranking algorithms, for a particular order, a full set of combinations of interactions can then be prioritized based on these features using a powerful ranking algorithm via support vectors. The computational ranking sheds light on unexplored combinations that can further be investigated using hypothesis testing based on wet lab experiments. Here, the basic framework and results obtained on 2\textsuperscript{nd} and 3\textsuperscript{rd} order interactions for members of family of DACT, SFRP, DKK (to name a few) in both normal and tumor cases is presented using a static data set.

Significance

The search and wet lab testing of unknown biological hypotheses in the form of combinations of various intra/extracellular factors that are involved in a signaling pathway, costs a lot in terms of time, investment and energy. To reduce this cost of search in a vast combinatorial space, a pipeline has been developed that prioritises these list of combinations so that a biologist can narrow down their investigation. The pipeline uses kernel based sensitivity indices to capture the influence of the factors in a pathway and employs powerful support vector ranking algorithm. The generic workflow and future improvements are bound to cut down the cost for many wet lab experiments and reveal unknown/untested biological hypothesis.

1 Introduction

1.1 A short review

Sharma\textsuperscript{1}'s accidental discovery of the Wingless played a pioneering role in the emergence of a widely expanding research field of the Wnt signaling pathway. A majority of the work has focused on issues related to • the discovery of genetic and epigenetic factors affecting the pathway (Thorstensen \textit{et al.}\textsuperscript{2} & Baron and Kneissl\textsuperscript{3}), • implications of mutations in the pathway and its dominant role on cancer and other diseases (Clevers\textsuperscript{4}), • investigation into the pathway's contribution towards embryo development (Sokol\textsuperscript{5}), homeostasis (Pinto \textit{et al.}\textsuperscript{6}, Zhong \textit{et al.}\textsuperscript{7})
1.2 Canonical Wnt signaling pathway

Before delving into the problem statement, a brief introduction to the Wnt pathway is given here. From the recent work of Sinha\cite{17}, the canonical Wnt signaling pathway is a transduction mechanism that contributes to embryo development and controls homeostatic self renewal in several tissues\cite{8}. Somatic mutations in the pathway are known to be associated with cancer in different parts of the human body. Prominent among them is the colorectal cancer case\cite{18}. In a succinct overview, the Wnt signaling pathway works when the Wnt ligand gets attached to the Frizzled (FZD)/LRP coreceptor complex. FZD may interact with the Dishevelled (DVL) causing phosphorylation. It is also thought that Wnts cause phosphorylation of the LRP via casein kinase 1 (CK1) and kinase GSK3. These developments further lead to attraction of Axin which causes inhibition of the formation of the degradation complex. The degradation complex constitutes of AXIN, the β-catenin transportation complex APC, CK1 and GSK3. When the pathway is active the dissolution of the degradation complex leads to stabilization in the concentration of β-catenin in the cytoplasm. As β-catenin enters into the nucleus it displaces the GROUCHO and binds with transcription cell factor TCF thus instigating transcription of Wnt target genes. GROUCHO acts as lock on TCF and prevents the transcription of target genes which may induce cancer. In cases when the Wnt ligands are not captured by the coreceptor at the cell membrane, AXIN helps in formation of the degradation complex. The degradation complex phosphorylates β-catenin which is then recognized by FBOX/WD repeat protein β-TRCP. β-TRCP is a component of ubiquitin ligase complex that helps in ubiquitination of β-catenin thus marking it for degradation via the proteasome. Cartoons depicting the phenomena of Wnt being inactive and active are shown in figures 1(A) and 1(B), respectively.

2 Problem statement

It is widely known that the sensitivity analysis plays a major role in computing the strength of the influence of involved factors in any phenomena under investigation. When applied to expression profiles of various intra/extracellular factors that form an integral part of a signaling pathway, the variance and density based analysis yields a range of sensitivity indices for individual as well as various combinations of factors. These combinations denote the higher order interactions among the involved factors. Computation of higher order interactions is often time consuming but they give a chance to explore the various combinations that might be of interest in the working mechanism of the pathway. For example, in a range of fourth order combinations among the various factors of the Wnt pathway, it would be easy to assess the influence of the destruction complex formed by APC, AXIN, CSKI and GSK3 interaction. Unknown interactions can be further investigated by and apoptosis\cite{8} and safety and feasibility of drug design for the Wnt pathway\cite{9,10,11,12,13}. Approximately forty years after the discovery, important strides have been made in the research work involving several wet lab experiments and cancer clinical trials\cite{9,13} which have been augmented by the recent developments in the various advanced computational modeling techniques of the pathway. More recent informative reviews have touched on various issues related to the different types of the Wnt signaling pathway and have stressed not only the activation of the Wnt signaling pathway via the Wnt proteins\cite{14} but also the on the secretion mechanism that plays a major role in the initiation of the Wnt activity as a prelude\cite{15}.

The work in this paper investigates some of the current aspects of research regarding the pathway via sensitivity analysis while using static\cite{16} data retrieved from colorectal cancer samples.
transforming biological hypothesis regarding these interactions in vitro, in vivo or in silico. But to mine these unknown interactions it is necessary to search a wide space of all combinations of input factors involved in the pathway.

In this work, after estimating the individual effects of factors for a higher order combination, the individual indices are considered as discriminative features. A combination, then is a multivariate feature set in higher order (>2). With an excessively large number of factors involved in the pathway, it is difficult to search for important combinations in a wide search space over different orders. Exploiting the analogy with the issues of prioritizing webpages using ranking algorithms, for a particular order, a full set of combinations of interactions can then be prioritized based on these features using a powerful ranking algorithm via support vectors (Joachims). The computational ranking sheds light on unexplored combinations that can further be investigated using hypothesis testing based on wet lab experiments. In this manuscript both local and global SA methods are used.

Similar higher order interactions can be computed and prioritized. This gives a chance to examine the biological hypothesis of interest regarding the positive/negative roles of unexplored combinations of the various factors involved in the Wnt pathway, in tumor and normal cases. Using recordings in time, it is possible to find how the prioritization of a specific combination of involved factors is changing in time. This further helps in revealing when an important combination like destruction complex formed by APC, AXIN, CSKI and GSK3 interaction could be influenced via an intervention. Thus a powerful computational mechanism is presented to explore the interactions involved in Wnt pathway. The framework can be used in any other pathway also.

3 Sensitivity analysis

Seminal work by Russian mathematician Sobol's lead to development as well as employment of SA methods to study various complex systems where it was tough to measure the contribution of various input parameters in the behaviour of the output. A recent unpublished review on the global SA methods by Looss and Lemaître categorically delineates these methods with the following functionality • screening for sorting influential measures (Morris method, Group screening in Moon et al., & Dean and Lewis, Iterated factorial design in Andres and Hajas, Sequential bifurcation in Bettonvil and Kleijn, and Cotter design), • quantitative indices for measuring the importance of contributing input factors in linear models (Christensen, Saltelli et al., Helton and Davis, and McKay et al.) and nonlinear models (Homma and Saltelli, Sobol, Saltelli, Saltelli et al., Saltelli et al., Cukier et al., Saltelli et al., Janon et al., Owen, Tissot and Prieur, Da Veiga and Gamboa, Archer et al., Tarantola et al., Saltelli and Annoni, and Jensen) and • exploring the model behaviour over a range on input values (Storlie and Helton, and Da Veiga et al., Li et al., and Hajikolaei and Wang). Looss and Lemaître also provide various criteria in a flowchart for adapting a method or a combination of the methods for sensitivity analysis. Figure 3 shows the general flow of the mathematical formulation for computing the indices in the variance based Sobol method. The general idea is as follows - A model could be represented as a mathematical function with a multidimensional input vector where each element of a vector is an input factor. This function needs to be defined in a unit dimensional cube. Based on ANOVA decomposition, the function can then be broken down into \( f_0 \) and summands of different dimensions, if \( f_0 \) is a constant and integral of summands with respect to their own variables is 0. This implies that orthogonality follows in between two functions of different dimensions, if at least one of the variables is not repeated. By applying these properties, it is possible to show that the function can be written into a unique expansion. Next, assuming that the function is square integrable variances can be computed. The ratio of variance of a group of input factors to the variance of the total set of input factors constitute the sensitivity index of a particular group. Detailed derivation is present in the Appendix.

Besides the above Sobol’s variance based indices, more recent developments regarding new indices based on density, derivative and goal-oriented can be found in Borgonovo, Sobol and Kucherenko, and Fort et al., respectively. In a more recent development, Da Veiga propose new class of indices based on density ratio estimation (Borgonovo) that are special cases of dependence measures. This in turn helps in exploiting measures like distance correlation (Székely et al.) and Hilbert-Schmidt independence criterion (Gretton et al.) as new sensitivity indices. The basic framework of these indices is based on use of Csiszár et al. f-divergence, concept of dissimilarity measure and kernel trick Aizerman et al. Finally, Da Veiga propose feature selection as an alternative to screening methods in sensitivity analysis. The main issue with variance based indices (Sobol’s) is that even though they capture importance information regarding the contribution of the input factors, they do not handle multivariate random variables easily and are only invariant under linear transformations. In comparison to these variance methods, the newly proposed indices based on density estimations (Borgonovo) and dependence measures are more robust. Figure 4 shows the general flow of the mathematical formulation for computing the indices in the density based HSIC method. The general idea is as follows - The sensitivity index is actually a distance correlation which incorporates the kernel based Hilbert-Schmidt Information Criterion between two input vectors in higher dimension. The criterion is nothing but the Hilbert-Schmidt norm of cross-covariance operator which generalizes the covariance matrix by representing higher order correlations between the input
Sensitivity analysis computes the strength of the influence of involved factors in any phenomena under investigation.

A combination denotes higher order interaction within the pathway.

Generate range of sensitivity indices for individual and combination of factors.

Excessively large number of factors involved in the pathway - difficult to search for important combinations in a wide search space over different orders.

Exploiting the analogy of prioritizing webpages using ranking algorithms, for a particular order, a full set of combinations of interactions can then be prioritized based on these features using a powerful ranking algorithm via support vectors. Ranking points to important interactions in Wnt-Tumor case.

RANK | HSIC rbf | HSIC laplace | HSIC linear
--- | --- | --- | ---
4 | DACT3SFRP1LEF1 SFRP1SFRP2SFRP4 | DKK3-1SFRP2SFRP4 | DKK3-1SFRP2SFRP4
3 | DACT3SFRP1SFRP4 DACT3SFRP1SFRP4 | DKK3-1SFRP1SFRP4 | DKK3-1SFRP1SFRP4
2 | DACT3SFRP1SFRP2 DACT3SFRP1LEF1 | DACT3SFRP1LEF1 | DACT3SFRP1SFRP2
1 | DACT3SFRP1SFRP3 DACT3SFRP1SFRP3 | DKK3-1DACT3SFRP4 | DKK3-1DACT3SFRP4

Fig. 2 A graphical view of the general idea behind the current work. (1) Sensitivity indices capture the influence of involved factors in a pathway. (2) Generate sensitivity indices for individual and combinations of involved factors. Note that for a combination, the indices for the involved factors in the combination are generated separately. (3) Vectorize the indices per combination. (4) Rank these combinations based on the sensitivity indices using support vector ranking as web pages are ranked using a ranking algorithm. (5) Obtained is a prioritized list of interactions that could point to important interactions in the pathway in cancer cases.

It is this strength of the kernel methods that HSIC is able to capture the deep nonlinearities in the biological data and provide reasonable information regarding the degree of influence of the involved factors within the pathway. Improvements in variance based methods also provide ways to cope with these nonlinearities but do not exploit the available strength of kernel methods. Results in the later sections provide experimental evidence for the same.

3.1 Relevance in systems biology

Recent efforts in systems biology to understand the importance of various factors apropos output behaviour has gained prominence. Sumner et al.\textsuperscript{61} compares the use of Sobol's variance based indices versus Morris\textsuperscript{22} screening method which uses a One-at-a-time (OAT) approach to analyse the sensitivity of GSK3 dynamics to uncertainty in an insulin signaling model. Similar efforts, but on different pathways can be found in Zheng and Rundell\textsuperscript{62} and Marino et al.\textsuperscript{63}.

SA provides a way of analyzing various factors taking part in a biological phenomena and deals with the effects of these factors on the output of the biological system under consideration. Usually, the model equations are differential in nature with a set of
Fig. 3 Computation of variance based sobol sensitivity indices. For detailed notations, see appendix.

inputs and the associated set of parameters that guide the output. SA helps in observing how the variance in these parameters and inputs leads to changes in the output behaviour. The goal of this manuscript is not to analyse differential equations and the parameters associated with it. Rather, the aim is to observe which input genotypic factors have greater contribution to observed phenotypic behaviour like a sample being normal or cancerous in both static and time series data. In this process, the effect of fold changes in time is also considered for analysis in the light of the recently observed psychophysical laws acting downstream of the Wnt pathway (Goentoro and Kirschner64).

There are two approaches to sensitivity analysis. The first is the local sensitivity analysis in which if there is a required solution, then the sensitivity of a function apropos a set of variables is estimated via a partial derivative for a fixed point in the input space. In global sensitivity, the input solution is not specified. This implies that the model function lies inside a cube and the sensitivity indices are regarded as tools for studying the model instead of the solution. The general form of g-function (as the model or output variable) is used to test the sensitivity of each of the input factor (i.e expression profile of each of the genes). This is mainly due to its non-linearity, non-monotonicity as well as the capacity to produce analytical sensitivity indices. The g-function takes the form

\[ f(x) = f_0 + \sum f_j(x_j) = f_0 + \sum_{i<j} f_{ij}(x_i, x_j) + \ldots + f_{1,2\ldots n}(x_1, x_2, \ldots, x_n) \]

f_0 is constant & integral of summands w.r.t their own variables is 0

\[ f_0 = \int \chi^n f(x) dx \]

Assuming f(x) is square integrable the constants are termed variances

\[ \int \chi^n f^2(x) dx - f_0^2 = D \]

Due to this, we have

\[ f(x)dx = f_0 + f_1(x_1) + f_2(x_2) + \ldots + f_n(x_n) \]

\[ \int_0^1 f(x) dx/dx_i = f_0 + f_1(x_1) + f_2(x_2) + f_n(x_n) \]

due to which

\[ \int_0^1 \ldots \int_0^1 \frac{f(x)dx}{dx_i} = f_0 + f_1(x_1) + f_2(x_2) + f_n(x_n) \]

\[ \int_0^1 \ldots \int_0^1 f_j(x_1, x_2, \ldots, x_i) dx_i = \frac{D_{i,j}}{D} \]

For the static (time series) data d = 18 (factors affecting the pathway). The value of d varies from 2 to 17, depending on the order of the combination one might be interested in. Thus the expression profiles of the various genetic factors in the pathway are considered as input factors and the global analysis conducted. Note that in the predefined dataset, the working of the signaling pathway is governed by a preselected set of genes that affect the pathway. For comparison purpose, the local sensitivity analysis method is also used to study how the individual factor is behaving with respect to the remaining factors while working of the pathway is observed in terms of expression profiles of the various factors.

Finally, in context of Goentoro and Kirschner64’s work regarding the recent development of observation of Weber’s law working downstream of the pathway, it has been found that the law is governed by the ratio of the deviation in the input and the absolute input value. More importantly, it is these deviations in input that are of significance in studying such a phenomena. The current manuscript explores the sensitivity of deviation in the fold changes between measurements of fold changes at consecutive time points to explore in what duration of time, a particular factor is affecting the pathway in a major way. This has deeper implications in the fact that one is now able to observe when in time an intervention can be made or a gene be perturbed to study the behaviour of the pathway in tumorous cases. Thus sensitivity analysis of deviations in the mathematical formulation of the psychophysical law can lead to insights into the time period based values of αi indicate higher importance of Xi. In our formulation, we randomly assign values of Xi ∈ [0, 1]. For the static (time series) data d = 18 (factors affecting the pathway). The value of d varies from 2 to 17, depending on the order of the combination one might be interested in. Thus the expression profiles of the various genetic factors in the pathway are considered as input factors and the global analysis conducted. Note that in the predefined dataset, the working of the signaling pathway is governed by a preselected set of genes that affect the pathway. For comparison purpose, the local sensitivity analysis method is also used to study how the individual factor is behaving with respect to the remaining factors while working of the pathway is observed in terms of expression profiles of the various factors.

Finally, in context of Goentoro and Kirschner’s work regarding the recent development of observation of Weber’s law working downstream of the pathway, it has been found that the law is governed by the ratio of the deviation in the input and the absolute input value. More importantly, it is these deviations in input that are of significance in studying such a phenomena. The current manuscript explores the sensitivity of deviation in the fold changes between measurements of fold changes at consecutive time points to explore in what duration of time, a particular factor is affecting the pathway in a major way. This has deeper implications in the fact that one is now able to observe when in time an intervention can be made or a gene be perturbed to study the behaviour of the pathway in tumorous cases. Thus sensitivity analysis of deviations in the mathematical formulation of the psychophysical law can lead to insights into the time period based
influence of the involved factors in the pathway. Thus, both global and local analysis methods are employed to observe the entire behaviour of the pathway as well as the local behaviour of the input factors with respect to the other factors, respectively, via analysis of fold changes and deviations in fold changes, in time.

Given the range of estimators available for testing the sensitivity, it might be useful to list a few which are going to be employed in this research study. Also, a brief introduction into the fundamentals of the derivation of the three main indices and the choice of sensitivity packages which are already available in literature, has been described in the Appendix.

4 Ranking Support Vector Machines

Learning to rank is a machine learning approach with the idea that the model is trained to learn how to rank. A good introduction to this work can be found in Li.65 Existing methods involve pointwise, pairwise and listwise approaches. In all these approaches, Support Vector Machines (SVM) can be employed to rank the required query. SVMs for pointwise approach build various hyperplanes to segregate the data and rank them. Pairwise approach uses ordered pair of objects to classify the objects and then utilize the classifier to rank the objects. In this approach, the group structure of the ranking is not taken into account. Finally, the listwise ranking approach uses ranking list as instances for learning and prediction. In this case the ranking is straightforward and the group structure of ranking is maintained. Various different designs of SVMs have been developed and the research in this field is still in preliminary stages. In context of the gene expression data set employed in this manuscript, the objects are the genes with their recorded expression values for normal and tumor cases. Both cases are treated separately.

Note that rankings algorithms have been developed to be employed in the genomic datasets but to the author’s awareness, these algorithms do not rank the range of combinations in a wide combinatorial search space in time. Also, they do not take into account the ranking of unexplored biological hypothesis which are assigned to a particular sensitivity value or vector that can be used for prioritization. For example, Kolde et al.66 presents a ranking algorithm that betters existing ranking model based on the assignment of $P$-value. As stated by Kolde et al.66 it detects genes that are ranked consistently better than expected under null hypothesis of uncorrelated inputs and assigns a significance score for each gene. The underlying probabilistic model makes the algorithm parameter free and robust to outliers, noise and errors. Significance scores also provide a rigorous way to keep only the statistically relevant genes in the final list. The proposed work here develops on sensitivity analysis and computes the influences of the factors for a system under investigation. These sensitivity indices give a much realistic view of the biological influence than the proposed $P$-value assignment and the probabilistic model. The manuscript at the current stage does not compare the algorithms as it is a pipeline to investigate and conduct a systems wide study. Instead of using SVM-Ranking it is possible to use other algorithms also, but the author has restricted to the development of the pipeline per se. Finally, the current work tests the effectiveness of the variance based (SOBOL) sensitivity

\[
S_{\text{HSIC}}(X, U)_{\alpha, \beta} = R(X_k, U)_{\alpha, \beta}
\]

The cross-covariance operator generalizes the covariance matrix by representing higher order correlations between $X$ and $U$ through nonlinear kernels. For every linear operator $C : B \to A$ and provided the sum converges, the Hilbert-Schmidt norm of $C$ is given by

\[
\|C\|_{\text{HS}}^2 = \sum_{k,l} \langle a_k, C b_l \rangle_{\alpha, \beta}
\]

\[a, b \text{ are orthonormal bases of } A \text{ and } B, \text{ respectively}\]

\[
\text{HSIC}_{\text{HS}}(X, U)_{\alpha, \beta} = \frac{1}{n^2} \text{Tr}(K_{X,H}K_{U,H})
\]

for $n$-dimensional vectors

\[
R^2(X, U)_{\alpha, \beta} = \frac{\text{HSIC}(X, U)_{\alpha, \beta}}{\sqrt{\text{HSIC}(X, X)_{\alpha, \beta} \text{HSIC}(U, U)_{\alpha, \beta}}}
\]

\[\text{HSIC is then defined as Hilbert-Schmidt norm of cross-covariance operator}\]

\[
\text{HSIC}(X, U)_{\alpha, \beta} = \left\{ \begin{array}{cl}
|C_{XU}|^2_{\text{HS}} & = \beta_{X,k}X^\beta_k \beta_{X',k'}X'^{\beta_k} k_{X}(U,U') + \beta_{X,k}X^\beta_k \beta_{X',k'}X'^{\beta_k} \beta_{U,k}U^\beta_k k_{U}(U,U') - 2 \beta_{X,k}X^\beta_k \beta_{X',k'}X'^{\beta_k} \beta_{U,k}U^\beta_k k_{X}(U,U') \\
\end{array} \right.
\]

Fig. 4 Computation of density based hsic sensitivity indices. For detailed notations, see appendix.
indices apropos the density and kernel based (HSIC) sensitivity indices. Finally, Blondel et al.\textsuperscript{67} provides a range of comparison for 10 different regression methods and a score to measure the models. Compared to the frame provided in Blondel et al.\textsuperscript{67}, the current pipeline takes into account biological information an converts into sensitivity scores and uses them as discriminative features to provide rankings. Thus the proposed method is algorithm independent.

5 Description of the dataset & design of experiments

A simple static dataset containing expression values measured for a few genes known to have important role in human colorectal cancer cases has been taken from Jiang et al.\textsuperscript{16}. Most of the expression values recorded are for genes that play a role in Wnt signaling pathway at an extracellular level and are known to have inhibitory affect on the Wnt pathway due to epigenetic factors. For each of the 24 normal mucosa and 24 human colorectal tumor cases, gene expression values were recorded for 14 genes belonging to the family of SFRP, DKK, WIFI, and DACT. Also, expression values of established Wnt pathway target genes like LEF1, MYC, CD44, and CCND1 were recorded per sample.

Note that green (red) represents activation (repression) in the heat maps of data in Jiang et al.\textsuperscript{16}. For the static data, only the scaled results are reported. This is mainly due to the fact that the measurements vary in a wide range and due to this there is often an error in the computed estimated of these indices. Bootstrapping without replicates on a smaller sample number is employed to generate estimates of indices which are then averaged. This takes into account the variance in the data and generates confidence bands for the indices.

GENERAL ISSUES - • Here the input factors are the gene expression values for both normal and tumor cases in static data. For the case of time series data, the input factors are the fold change (deviations in fold change) expression values of genes at different time points (periods). Also, for the time series data, in the first experiment the analysis of a pair of the fold changes recorded at to different consecutive time points i.e $t_i$ & $t_{i+1}$ is done. In the second experiment, the analysis of a pair of deviations in fold changes recorded at $t_i$ & $t_{i+1}$ and $t_{i+1}$ & $t_{i+2}$. In this work, in both the static and the time series datasets, the analysis is done to study the entire model/pathway rather than find a particular solution to the model/pathway. Thus global sensitivity analysis is employed. But the local sensitivity methods are used to observe and compare the affect of individual factors via 1\textsuperscript{st} order analysis w.r.t total order analysis (i.e global analysis). In such an experiment, the output is the sensitivity indices of the individual factors participating in the model. This is different from the general trend of observing the sensitivity of parameter values that affect the pathway based on differential equations that model a reaction. Thus the model/pathway is studied as a whole by observing the sensitivities of the individual factors.

Note that the 24 normal and tumor cases are all different from each other. The 18 genes that are used to study in\textsuperscript{16} are the input factors and it is unlikely that there will be correlations between different patients. The phenotypic behaviour might be similar at a grander scale. Also, since the sampling number is very small for a network of this scale, large standard deviations can be observed in many results, especially when the Sobol method is used. But this is not the issue with the sampling number. By that analysis, large deviations are not observed in kernel based density methods. The deviations are more because of the fact that the nonlinearities are not captured in an efficient way in the variance based Sobol methods. Due to this, the resulting indices have high variance in numerical value. For the same number of samplings, the kernel based methods don't show high variance.

The procedure begins with the listing of all $C_k^n$ combinations for $k$ number of genes from a total of $n$ genes. $k$ is $\geq 2$ and $\leq (n-1)$. Each of the combination of order $k$ represent a unique set of interaction between the involved genetic factors. While studying the interaction among the various genetic factors using static data, tumor samples are considered separated from normal samples. For the experiments conducted here on a toy example, 20 bootstrap replicates were generated for each of normal and tumor samples without replacement. For each bootstrap replicate, the normal and tumor samples are divided into two different sets of equal size. Next the datasets are combined in a specified format which go as input as per the requirement of a particular sensitivity analysis method. Thus for each $p^{th}$ combination in $C_k^n$ combinations, the dataset is prepared in the required format from both normal and tumor samples (See \textup{R} code in main script-1.1.R in google drive and step 3 in figure 5). After the data has been transformed, vectorized programming is employed for density based sensitivity analysis and looping is employed for variance based sensitivity analysis to compute the required sensitivity indices for each of the $p$ combinations. Once the sensitivity indices are generated for each of the $p^{th}$ combination, for every bootstrap replicate in normal and tumor cases, confidence intervals are estimated for each sensitivity index. This procedure is done for different kinds of sensitivity analysis methods.

After the above sensitivity indices have been stored for each of the $p^{th}$ combination, each of the sensitivity analysis method for normal and tumor cases per bootstrap replicates, the next step in the design of experiment is conducted. Here, for a particular $k^{th}$ order of combination, a choice is made regarding the number of sample size (say $p$), where $2 \leq p \leq noObs - 1$ (noObs is the number of observations i.e 20 replicates). Then for all sample sets of order $p$ in $C_p^{noObs}$, generate training index set of order $p$ and test index set of order $noObs-p$. For each of
**Fig. 5** A cartoon of experimental setup. IMPORTANT NOTE - In this figure, \( G_x \), \( G_y \) and \( G \) represent a combination. Step - (1) Segregation of data into normal and tumor cases. (2) Further data division per case and bootstrap sampling with no repetitions for different iterations. (3) Assembling bootstrapped data and application of SA methods. (4) Generation of SI’s for normal and tumor case per gene per iteration. (5) Generation of averaged SI and confidence bands per case per gene.

the sample set, considering the sensitivity index for each individual factor of a gene combination in the previous step as described in the foregoing paragraph, a training and a test set is generated. Thus an observation in a training and a test set represents a gene combination with sensitivity indices of involved genetic factors as feature values. For a particular gene combination there are \( p \) training samples \( \text{noObs} = p \) test samples. In total there are \( C_{\text{noObs}}^p \) training sets and corresponding test sets. Next, \( SVM_{\text{Rank}} \) (Joachims\textsuperscript{19}) is used to generate a model on default value \( C \) value of 20. In the current experiment on toy model \( C \) value has not been tuned. The training set helps in the generation of the model as the different gene combinations are numbered in order which are used as rank indices. The model is then used to generate score on the observations in the testing set using the \( SVM_{\text{Rank}} \) model. Next the scores are averaged across all \( C_{\text{noObs}}^p \) test samples. The experiment is conducted for normal and tumor samples separately. This procedure is executed for each and every sensitivity analysis method. Finally, for each sensitivity analysis method, for all \( k^{th} \) order combinations, the mean across the averaged \( p \) scores is computed. This is followed by sorting of these scores along with the rank indices already assigned to the gene combinations. The end result is a sorted order of the gene combinations based on the ranking score learned by the \( SVM_{\text{Rank}} \) algorithm. These steps are depicted in figure 6.

Note that the following is the order in which the files should be executed in R, in order, for obtaining the desired results (Note that the code will not be explained here)
- • use source("mainScript-1-1.R") with arguments for Static data • use source("Combine-Static-files.R")
- • source("Store-Results-S.R")
- • source("SVMRank-Results-S.R") again this needs to be done separately for different kinds of SA methods and • source("Sort-n-...-R.R")
For all k no. of genes compute nCk combinations
n - Total No. of genes
k - number genes >=2 & <=(n-1)
for all noObsCp combinations of samples,
noObs - total no. of observations, p - >= 2 & (noObs-1)

jth combination
for k = 2, SA1

SA1

SA2

SAm

Gx

Normal

Tumor

Gy

Normal

Tumor

Gx

Normal

Tumor

Gy

Normal

Tumor

Gx

Normal

Tumor

Gy

Normal

Tumor

Gx

Normal

Tumor

Gy

Normal

Tumor

Gx

Normal

Tumor

Gy

Normal

Tumor

Gx

Normal

Tumor

Gy

Normal

Tumor

Gx

Normal

Tumor

Gy

Normal

Tumor

jth combination
for k = 2, SA2

jth combination
for k = 2, SAn

jth combination
for k = 2, SAe

Plot-S.R”) to sort the interactions.

6 Results and Discussion

Initial results on prioritization of 2nd order interactions learnt from support vector ranking using these sensitivity indices were plotted for visualization. For a training sample size of 4 and test sample size of 16, from a total of 20 bootstraps for each of the interactions, in normal and tumor cases, and the sensitivity index using $F_d/uniEBF8/uniEBE9 - Ch/uniEBE9^2$, the computed sensitivity indices were sorted in increasing order of value and are shown in figure 7 and 8, respectively. The $y-axis$ represents the sensitivity indices and the $x-axis$ represents the values of the different factors involved under investigation (here $2^{nd}$ order combinations). Since the indices have been sorted based on the ranking score, the combinations are now ordered according to the strength of their significance. The combinations with low significance or sensitivity index is placed towards the left most end and vice versa. Visualizations for other indices were also generated and are relegated to the appendix. Since various indices have been tested, it would be important to see how good the rankings tally across the different sensitivity methods. To tally the rankings visually, the top and the bottom 10 combinations were tabulated in tables 1, 2, 5, 3 and 4. The top and the bottom 10 rankings in figures 7 and 8 can be found in the second columns of tables 1 and 2.
In tumor cases (table 1), for most of the sensitivity indices we find that SFRP1-SFRP3 has the highest priority in the data. This is followed by various others interactions like DACT3-SFRP3 and DACT3-SFRP4. For other kinds of the interactions we find varying kinds of rankings. It is important to note that the rankings are relative but they do give a glimpse of the importance of the interactions in a vast combinatorial space. These rankings are of importance in a particular phenomena of investigation. A particular colour across the columns, more specifically the sensitivity methods, represent the relative positioning of a particular interaction. The different colours in a row of a column do not vary in a row of a column do not
Table 1 Ranking of second order interactions in Tumor case using density and variance based sensitivity indices. Here 1 has high priority and 153 has low priority.

<table>
<thead>
<tr>
<th>Rank</th>
<th>chi2</th>
<th>hellinger</th>
<th>kl</th>
<th>tv</th>
<th>rbf</th>
<th>laplace</th>
</tr>
</thead>
<tbody>
<tr>
<td>High priority interactions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>SFRP1LEF1</td>
<td>SFRP1MYC</td>
<td>SFRP5LEF1</td>
<td>SFRP5LEF1</td>
<td>SFRP2SFRP4</td>
<td>SFRP1MYC</td>
</tr>
<tr>
<td>9</td>
<td>SFRP1SFRP5</td>
<td>DACT3LEF1</td>
<td>DACT3SFRP5</td>
<td>DACT3SFRP5</td>
<td>SFRP2SFRP3</td>
<td>SFRP5LEF1</td>
</tr>
<tr>
<td>8</td>
<td>DACT3SFRP2</td>
<td>SFRP1SFRP2</td>
<td>SFRP5MYC</td>
<td>SFRP5MYC</td>
<td>SFRP1LEF1</td>
<td>DACT3LEF1</td>
</tr>
<tr>
<td>7</td>
<td>SFRP1SFRP3</td>
<td>SFRP1LEF1</td>
<td>SFRP1MYC</td>
<td>SFRP1MYC</td>
<td>SFRP1LEF1</td>
<td>DACT2SFRP3</td>
</tr>
<tr>
<td>6</td>
<td>SFRP2SFRP4</td>
<td>SFRP2SFRP4</td>
<td>DACT3SFRP4</td>
<td>DACT3SFRP4</td>
<td>SFRP1SFRP4</td>
<td>SFRP1SFRP4</td>
</tr>
<tr>
<td>5</td>
<td>SFRP1SFRP4</td>
<td>SFRP2SFRP3</td>
<td>DACT3MYC</td>
<td>DACT3MYC</td>
<td>SFRP1LEF1</td>
<td>DACT2SFRP2</td>
</tr>
<tr>
<td>4</td>
<td>SFRP2SFRP3</td>
<td>SFRP1SFRP4</td>
<td>DACT3LEF1</td>
<td>DACT3LEF1</td>
<td>SFRP1SFRP4</td>
<td>SFRP1SFRP2</td>
</tr>
<tr>
<td>3</td>
<td>DACT3SFRP4</td>
<td>DACT3SFRP4</td>
<td>SFRP1LEF1</td>
<td>SFRP1LEF1</td>
<td>DACT3SFRP2</td>
<td>DACT3SFRP3</td>
</tr>
<tr>
<td>2</td>
<td>DACT3SFRP3</td>
<td>DACT3SFRP3</td>
<td>DACT3SFRP3</td>
<td>DACT3SFRP3</td>
<td>DACT3SFRP5</td>
<td>SFRP1LEF1</td>
</tr>
<tr>
<td>1</td>
<td>SFRP1SFRP3</td>
<td>SFRP1SFRP3</td>
<td>SFRP1SFRP3</td>
<td>SFRP1SFRP3</td>
<td>SFRP1SFRP3</td>
<td>SFRP1SFRP3</td>
</tr>
</tbody>
</table>

| Low priority interactions |
| 153 | DKK3-1SFRP1 | DKK3-1SFRP1 | DKK3-1SFRP1 | DKK3-1SFRP1 | DKK4DACT3 | DKK4DACT3 |
| 152 | DKK3-1DACT3 | DKK3-1DACT3 | DKK3-1DACT3 | DKK3-1DACT3 | DKK2DACT3 | DKK3-1SFRP5 |
| 151 | DKK3-1SFRP2 | DKK3-1SFRP2 | DKK3-1SFRP2 | DKK3-1SFRP2 | DKK3-1SFRP5 | DKK4DACT3 |
| 150 | DKK3-1SFRP5 | DKK3-1SFRP5 | DKK3-1SFRP5 | DKK3-1SFRP5 | DKK1DACT3 | DKK2DACT3 |
| 149 | SFRP3SFRP5 | SFRP3SFRP5 | SFRP3SFRP5 | SFRP3SFRP5 | DKK1SFRP1 | DKK3-1SFRP1 |
| 148 | SFRP4SFRP5 | SFRP4SFRP5 | DACT1SFRP1 | DACT1SFRP1 | DACT1SFRP1 | DACT2SFRP1 |
| 147 | DKK4SFRP1 | DACT1SFRP1 | SFRP4SFRP5 | SFRP4SFRP5 | DKK2DACT3 | DKK2SFRP1 |
| 146 | DKK4DACT3 | DACT1SFRP2 | DACT1DACT3 | DACT1DACT3 | DKK2SFRP1 | DKK4SFRP2 |
| 145 | DACT1SFRP1 | DACT1DACT3 | DACT1SFRP5 | DACT1SFRP5 | DACT2SFRP1 | DACT1SFRP1 |
| 144 | DKK3-2SFRP5 | DKK4SFRP1 | DKK2DACT3 | DKK2DACT3 | DKK4SFRP1 | DKK2SFRP1 |

mean any thing. What has been found is that DACT3 is implicated in the Wnt β-catenin signaling in colorectal cancer through epigenetic modifications and is a histone modification therapeutic target Jiang et al.16. As a confirmatory result, the pipeline indicates the priority that DACT3 gets along with the members of the SFRP family in the tumor. This high priority also indicates the efficacy of the designed pipeline. Combinations of the members of the SFRP family also show at the high priority list. More specifically SFRP1SFRP3, SFRP2SFRP3 and SFRP1SFRP2.

On the contrary, when we look at the set of low priority interactions, there is a range of combinations involving the family of DKK and members of DACT, SFRP. Surprisingly, specific interactions of within family members also show up at this stage. This is specially seen for DACT1-DACT3, DACT2-DACT3, SFRP3-SFRP5 and SFRP4-SFRP5. These indicate the non-effectiveness of the 2nd order interactions among a family of protein in the tumor case. Based on these rankings it might be possible to narrow down the search for important interactions in the vast combinatorial search space. We also generated the similar rankings for the interactions between the same set of factors in normal case. These are represented in table 2.

In comparison to the tables 1 and 2, which take into account the density based sensitivity methods, we also generated the rankings in tumor and normal cases using variance based sensitivity indices (table 5 in appendix). What we find is that there are interactions that get high priority in both the tumor and normal case. For example, the cases of DKK3−1-WIF1 and DKK3−2-LEF1, show up with high priority in both the normal and tumor case. This indicates the inability of the variance based sensitivity indices to capture crucial influence of the participating factors or combination of the factors, properly. Clearly, what we find is that the density based methods show superior selection and capture of important nonlinear biological information to provide proper indices that can be ranked later on.

6.2 3rd Order Interactions

We now turn our attention to the 3rd order combinations which open a deeper and a greater space of investigation as compared to the space of the second order combinations. Tables 3 and 4 show the rankings for the 3rd order interactions in tumor and normal cases, respectively. A major finding is that DACT3 combination with members of the family of SFRP show very high ranking in tumor cases. Note that we could only generate the rankings for the different kernels in the HSIC method as other density based
methods were computationally intensive and costly in time. On the other hand, the combinations of the family of DKK with other factors showed very low rankings in the tumor case, thus indicating their non-participation or downregulated participation in the tumor cases. The most beautiful aspect of these rankings is that one can specifically see which members of a family are having a potent role in the tumor and which ones are not. For example, DACT 3 plays a very important role along with various members of the SFRP family but the combination DACT 3-SFRP 3-SFRP 5 shows very low participation in tumor cases. This potential of the search engine to specifically rank the different combinations gives the biologists deep insights into the interactions of members of a single family of proteins also. A similar ranking profile was generated for the normal case.

Code Availability Code has been made available on Google drive at https://drive.google.com/folderview?id=0B7Kkv8whPU-V1Fkd1dMSTd5ak0&usp=sharing

### 7 Conclusions

A workflow has been presented that can prioritize the entire range of interactions among the constituent or subgroup of intr/extracellular factors affecting the pathway by using a powerful algorithm of support vector ranking on interactions that have sensitivity indices of the involved factors as features. These sensitivity indices compute the influences of the factors on the pathway and represent nonlinear biological relations among the factors that are captured using kernel methods. SVM ranking then scores the testing data which can be sorted to find the highly prioritized interactions that need further investigation. Using this efficient workflow, it is possible to analyze any combination of involved factors in a signaling pathway. Here, we show confirmatory results for members of family of DACT, SFRP (DKK) to have highly prioritized role in tumor (normal) cases, in a single experimental setup. These preliminary results show the efficacy of the pipeline in providing a ranked list of interactions which the biologists can narrow down to in a vast search space of combinations, thus cutting down the cost of terms of time, investment and energy. Additionally, the pipeline also generates unknown/untested biological hypothesis regarding the combinations involved in the pathway in both tumor and normal cases. This opens up the potential for the biologists to work on recommendations from this pipeline in a vast combinatorial forest of combinations, when many a times, it is not possible to know where to start the search.

<table>
<thead>
<tr>
<th>Rank</th>
<th>chi2</th>
<th>hellinger</th>
<th>kl</th>
<th>tv</th>
<th>rbf</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>DACT3CCND1</td>
<td>DKK2LEF1</td>
<td>DKK2DDK4</td>
<td>DACT3MYC</td>
<td>DKK2SFRP5</td>
</tr>
<tr>
<td>9</td>
<td>SFRP5WIF1</td>
<td>DKK2SFRP5</td>
<td>DKK2LEF1</td>
<td>DKK2WIF1</td>
<td>DKK2SFRP5</td>
</tr>
<tr>
<td>8</td>
<td>WIF1LEF1</td>
<td>DACT3SFRP2</td>
<td>DKK2DACT2</td>
<td>DACT3SFRP2</td>
<td>SFRP5LEF1</td>
</tr>
<tr>
<td>7</td>
<td>DKK2MYC</td>
<td>SFRP5LEF1</td>
<td>DKK2MYC</td>
<td>SFRP5LEF1</td>
<td>DKK2DACT2</td>
</tr>
<tr>
<td>6</td>
<td>DKK1DKK4</td>
<td>DACT3CCND1</td>
<td>DACT3LEF1</td>
<td>DACT3CCND1</td>
<td>DKK2DACT2</td>
</tr>
<tr>
<td>5</td>
<td>DACT2MYC</td>
<td>DKK1MYC</td>
<td>DACT3CCND1</td>
<td>DACT3LEF1</td>
<td>SFRP5CCND1</td>
</tr>
<tr>
<td>4</td>
<td>SFRP5LEF1</td>
<td>DKK2MYC</td>
<td>SFRP5LEF1</td>
<td>DKK2MYC</td>
<td>DKK2DACT2</td>
</tr>
<tr>
<td>3</td>
<td>DKK2DACT2</td>
<td>DKK2DACT2</td>
<td>DACT3SFRP2</td>
<td>DKK2DACT2</td>
<td>DKK2DACT2</td>
</tr>
<tr>
<td>2</td>
<td>DKK2LEF1</td>
<td>DKK2LEF1</td>
<td>DKK2WIF1</td>
<td>DKK2DACT2</td>
<td>DKK2DACT2</td>
</tr>
<tr>
<td>1</td>
<td>DKK2DDK4</td>
<td>DKK2DDK4</td>
<td>DACT3MYC</td>
<td>DKK2DDK4</td>
<td>DKK1MYC</td>
</tr>
</tbody>
</table>

Table 2: Ranking of second order interactions in Normal case using density and variance based sensitivity indices. Here 1 has high priority and 153 has low priority.
The author thanks Mr. Prabhat Sinha and Mrs. Rita Sinha for financially supporting the project.

References

### Table 4

Ranking of third order interactions in Normal case using variance based sensitivity indices. Here 1 has high priority and 816 has low priority.

<table>
<thead>
<tr>
<th>Ranking</th>
<th>HSIC</th>
<th>rbf</th>
<th>laplace</th>
<th>linear</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High priority interactions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>DKK2WIF1MYC</td>
<td>DACT3SFRP5MYC</td>
<td>DKK4SFRP4CCND1</td>
<td>DKK4DACT2LEF1</td>
</tr>
<tr>
<td>9</td>
<td>DKK2DKK4WIF1</td>
<td>DKK3-1DACT3CCND1</td>
<td>DKK1DKK4MYC</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>DKK1LEF1MYC</td>
<td>DKK2DACT2LEF1</td>
<td>SFRP5WIF1CCND1</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>DKK3-1LEF1CCND1</td>
<td>DACT3SFRP5CCND1</td>
<td>WIF1LEF1CCND1</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>DKK2DKK4MYC</td>
<td>DACT3SFRP5CCND1</td>
<td>DACT2LEF1CCND1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>DKK2DACT2LEF1</td>
<td>DACT3SFRP5CCND1</td>
<td>DKK1LEF1CCND1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>DACT3SFRP5CCND1</td>
<td>DACT3SFRP5CCND1</td>
<td>DACT2LEF1CCND1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>DACT3SFRP5CCND1</td>
<td>DACT3SFRP5CCND1</td>
<td>DKK1LEF1CCND1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>DKK2DKK4DACT2</td>
<td>DKK2DKK4DACT2</td>
<td>SFRP2LEF1CCND1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>DKK2DKK4LEF1</td>
<td>DKK2DKK4LEF1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low priority interactions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>816</td>
<td>DKK1DKK2SFRP5</td>
<td>DKK4SFRP1LEF1</td>
<td>DKK4DACT2LEF1</td>
<td></td>
</tr>
<tr>
<td>815</td>
<td>DKK1DKK2DKK4</td>
<td>DKK1DKK3-2SFRP5</td>
<td>DKK4WIF1LEF1</td>
<td></td>
</tr>
<tr>
<td>814</td>
<td>DKK4SFRP1LEF1</td>
<td>DKK1DKK3-2SFRP2</td>
<td>DKK1DKK2LEF1</td>
<td></td>
</tr>
<tr>
<td>813</td>
<td>DKK4WIF1LEF1</td>
<td>DKK1DKK3-2SFRP4</td>
<td></td>
<td>DKK1DKK2DKK4</td>
</tr>
<tr>
<td>812</td>
<td>DKK4SFRP1CCND1</td>
<td>DKK1DKK3-2DACT2</td>
<td></td>
<td>DACT2WIF1LEF1</td>
</tr>
<tr>
<td>811</td>
<td>DKK4DACT3SFRP1</td>
<td>DKK1DKK3-2CCND1</td>
<td>DKK1DKK2WIF1</td>
<td></td>
</tr>
<tr>
<td>810</td>
<td>DKK4SFRP1WIF1</td>
<td>DKK4SFRP1MYC</td>
<td>DACT2DACT3SFRP5</td>
<td></td>
</tr>
<tr>
<td>809</td>
<td>DKK4SFRP1SFRP2</td>
<td></td>
<td>DKK1DKK2DACT2</td>
<td></td>
</tr>
<tr>
<td>808</td>
<td>DKK4SFRP3LEF1</td>
<td>DKK1SFRP3SFRP4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>807</td>
<td>DKK1SFRP3CCND1</td>
<td>DKK4SFRP1WIF1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

105, 2811–2828
Appendix

9 Sensitivity indices

9.1 Variance based indices

The variance based indices as proposed by Sobol’ prove a theorem that an integrable function can be decomposed into summands of different dimensions. Also, a Monte Carlo algorithm is used to estimate the sensitivity of a function apropos arbitrary group of variables. It is assumed that a model denoted by function \( u = f(x) \), \( x = (x_1, x_2, \ldots, x_n) \), is defined in a unit \( n \)-dimensional cube \( K^n \) with \( u \) as the scalar output. The requirement of the problem is to find the sensitivity of function \( f(x) \) with respect to different variables. If \( u^* = f(x^*) \) is the required solution, then the sensitivity of \( u^* \) apropos \( x_k \) is estimated via the partial derivative \( (\partial u/\partial x_k)_{x=x^*} \). This approach is the local sensitivity. In global sensitivity, the input \( x = x^* \) is not specified. This implies that the model \( f(x) \) lies inside the cube and the sensitivity indices are regarded as tools for studying the model instead of the solution. Detailed technical aspects with examples can be found in Homma and Saltelli and .

Let a group of indices \( i_1, i_2, \ldots, i_s \) exist, where \( 1 \leq i_1 < \ldots < i_s \leq n \) and \( 1 \leq s \leq n \). Then the notation for sum over all different groups of indices is -

\[
\Sigma_{i_1, i_2, \ldots, i_s} = \Sigma_{i_1} \Sigma_{i_2} \ldots \Sigma_{i_s} = \Sigma_{i_1} \Sigma_{i_2} \ldots \Sigma_{i_s}
\]

Then the representation of \( f(x) \) using equation 2 in the form -

\[
f(x) = f_0 + \Sigma_{i_1, i_2, \ldots, i_s} f_{i_1, i_2, \ldots, i_s}
\]

is called ANOVA-decomposition from Archer et al. or expansion into summands of different dimensions, if \( f_0 \) is a constant and integrals of the summands \( f_{i_1, i_2, \ldots, i_s} \) with respect to their own variables are zero, i.e,

\[
f_0 = \int_{K^n} f(x) dx
\]

\[
\int_{0}^{1} f_{i_1, i_2, \ldots, i_s}(x_{i_1}, x_{i_2}, \ldots, x_{i_s}) dx_{i_k} = 0, 1 \leq k \leq s
\]

It follows from equation 4 that all summands on the right hand side are orthogonal, i.e if at least one of the indices in \( i_1, i_2, \ldots, i_s \) and \( j_1, j_2, \ldots, j_l \) is not repeated i.e

\[
\int_{0}^{1} f_{i_1, i_2, \ldots, i_s}(x_{i_1}, x_{i_2}, \ldots, x_{i_s}) f_{j_1, j_2, \ldots, j_l}(x_{j_1}, x_{j_2}, \ldots, x_{j_l}) dx = 0
\]

Sobol’ proves a theorem stating that there is an existence of a unique expansion of equation 4 for any \( f(x) \) integrable in \( K^n \). In brief, this implies that for each of the indices as well as a group of indices, integrating equation 4 yields the following -

\[
\int_{0}^{1} \int_{0}^{1} f(x) dx_{i} dx_{j} = f_0 + f_{i}(x_i)
\]

\[
\int_{0}^{1} \int_{0}^{1} f(x) dx_{i} dx_{j} = f_0 + f_{i}(x_i) + f_{j}(x_j) + f_{i,j}(x_i, x_j)
\]

were, \( dx_{i} dx_{j} = \prod_{i \in \{1, \ldots, n\}, j \in \{1, \ldots, n\}} dx_i \) and \( dx_{i} dx_{j} \) is \( \prod_{i \in \{1, \ldots, n\}, j \in \{1, \ldots, n\}} dx_i dx_j \). For higher orders of grouped indices, similar computations follow. The computation of any summand \( f_{i_1, i_2, \ldots, i_s}(x_{i_1}, x_{i_2}, \ldots, x_{i_s}) \) is reduced to an integral in the cube \( K^n \). The last summand \( f_{1, 2, \ldots, n}(x_{1}, x_{2}, \ldots, x_{n}) \) is \( f(x) - f_0 \) from equation 4. Homma and Saltelli stresses that use of Sobol sensitivity indices does not require evaluation of any \( f_{i_1, i_2, \ldots, i_s}(x_{i_1}, x_{i_2}, \ldots, x_{i_s}) \) nor the knowledge of the form of \( f(x) \) which might well be represented by a computational model i.e a function whose value is only obtained as the output of a computer program.

Finally, assuming that \( f(x) \) is square integrable, i.e \( f(x) \in L^2 \), then all of \( f_{i_1, i_2, \ldots, i_s}(x_{i_1}, x_{i_2}, \ldots, x_{i_s}) \in L^2 \). Then the following constants

\[
\int_{K^n} f^2(x) dx - f_0^2 = D
\]

\[
\int_{0}^{1} \int_{0}^{1} f^2(x_{i_1}, x_{i_2}, \ldots, x_{i_s}) dx_{i_1} dx_{i_2} \ldots dx_{i_s} = D_{i_1, i_2, \ldots, i_s}
\]

are termed as variances. Squaring equation 4, integrating over \( K^n \) and using the orthogonality property in equation 6, \( D \) evaluates to -

\[
D = \Sigma D_{i_1, i_2, \ldots, i_s}
\]

Then the global sensitivity estimates is defined as -

\[
S_{i_1, i_2, \ldots, i_s} = D_{i_1, i_2, \ldots, i_s} / D
\]

It follows from equations 11 and 12 that

\[
\Sigma S_{i_1, i_2, \ldots, i_s} = 1
\]

Clearly, all sensitivity indices are non-negative, i.e an index \( S_{i_1, i_2, \ldots, i_s} = 0 \) if and only if \( f_{i_1, i_2, \ldots, i_s} \equiv 0 \). The true potential of Sobol indices is observed when variables \( x_1, x_2, \ldots, x_n \) are divided into \( m \) different groups with \( y_1, y_2, \ldots, y_m \) such that \( m < n \). Then \( f(x) \equiv f(y_1, y_2, \ldots, y_m) \). All properties remain the same for the computation of sensitivity indices with the fact that integration with respect to \( y_k \) means integration with respect to all the \( x_i \)’s in \( y_k \). Details of these computations with examples can be found in . Variations and improvements over Sobol indices
have already been stated in section 3.

9.2 Density based indices

As discussed before, the issue with variance based methods is the high computational cost incurred due to the number of interactions among the variables. This further requires the use of screening methods to filter out redundant or unwanted factors that might not have significant impact on the output. Recent work by Da Veiga\cite{dfdiv} proposes a new class of sensitivity indices which are a special case of density based indices Borgonovo\cite{ddiv}. These indices can handle multivariate variables easily and relies on density ratio estimation. Key points from Da Veiga\cite{dfdiv} are mentioned below.

Considering the similar notation in previous section, \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) (\( u = f(x) \)) is assumed to be continuous. It is also assumed that \( X_k \) has a known distribution and are independent. Baucells and Borgonovo\cite{dfdiv} state that a function which measures the similarity between the distribution of \( U \) and that of \( U|X_k \) can define the impact of \( X_k \) on \( U \). Thus the impact is defined as -

\[
S_{X_k} = \mathcal{E}(d(U, U|X_k)) \tag{14}
\]

were \( d(\cdot, \cdot) \) is a dissimilarity measure between two random variables. Here \( d \) can take various forms as long as it satisfies the criteria of a dissimilarity measure. Csiszár et al.\cite{ddiv}’s \( f \)-divergence between \( U \) and \( U|X_k \) when all input random variables are considered to be absolutely continuous with respect to Lebesgue measure on \( \mathbb{R} \) is formulated as -

\[
d_F(U||U|X_k) = \int_{\mathbb{R}} F(\frac{p_U(u)}{p_{U|X_k}(u)}) p_{U|X_k}(u) du \tag{15}
\]

were \( F \) is a convex function such that \( F(1) = 0 \) and \( p_U \) and \( p_{U|X_k} \) are the probability distribution functions of \( U \) and \( U|X_k \). Standard choices of \( F \) include Kullback-Leibler divergence \( F(t) = -\log p(t) \), Hellinger distance \( (\sqrt{t} - 1)^2 \), Total variation distance \( F(t) = |t-1| \), Pearson \( \chi^2 \) divergence \( F(t) = t^2 - 1 \) and Neyman \( \chi^2 \) divergence \( F(t) = (1-t^2)/t \). Substituting equation 15 in

\[
\begin{array}{|c|c|c|c|c|c|c|}
\hline
\text{Tumor Case} & \text{Normal Case} \\
\hline
\text{Rank} & \text{2007} & \text{jansen} & \text{martinez} & \text{2007} & \text{jansen} & \text{martinez} \\
\hline
10 & DACT3WIF1 & SFRP3LEF1 & SFRP3LEF1 & DKK3-2CD44 & SFRP3LEF1 & DKK3-1DACT2 \\
9 & SFRP1LEF1 & DKK2LEF1 & DKK3-1MYC & SFRP2LEF1 & DKK2LEF1 & SFRP3WIF1 \\
8 & DACT3LEF1 & DACT1LEF1 & DKK3-1DACT2 & DKK3-2WIF1 & DACT1LEF1 & SFRP3LEF1 \\
7 & SFRP4LEF1 & DKK3-1WIF1 & DKK2LEF1 & DKK3-2DACT2 & DKK2DDK4 & ACT1LEF1 \\
6 & DKK3-1WIF1 & DKK2DDK4 & DACT1LEF1 & WIFI1MYC & SFRP4LEF1 & DKK2DDK4 \\
5 & DACT1LEF1 & DKK3-2DDK4 & SFRP5LEF1 & DKK4MYC & DKK3-2DDK4 & SFRP5LEF1 \\
4 & DKK3-2LEF1 & SFRP4LEF1 & DKK3-1LEF1 & DKK3-2DDK4 & DKK3-1DDK4 & DKK2LEF1 \\
3 & SFRP4WIF1 & DKK3-2LEF1 & DKK3-1WIF1 & DKK3-2MYC & DKK3-1WIF1 & DKK3-1LEF1 \\
2 & DACT2LEF1 & DKK3-1DDK4 & DKK2DDK4 & LEF1MYC & DKK3-2LEF1 & DKK3-1WIF1 \\
1 & SFRP5LEF1 & SFRP5LEF1 & DKK3-1DDK4 & DKK3-2LEF1 & SFRP5LEF1 & DKK3-1DDK4 \\
\hline
\end{array}
\]

Table 5 Ranking of second order interactions in Tumor & Normal case using density and variance based sensitivity indices. Here 1 has high priority and 153 has low priority.
equation 14, gives the following sensitivity index -

\[ S_{X_k}^F = \int_{\mathcal{R}} d\mathcal{R}(U|U|X_k)p_{X_k}(x)dx \]

\[ = \int_{\mathcal{R}} \int_{\mathcal{R}} F\left(\frac{p_{U}(u)}{p_{U}(X_k)}\right)p_{U}(X_k)dxdu \]

\[ = \int_{\mathcal{R}^2} F\left(\frac{p_{U}(u)p_{X_k}(x)}{p_{U}(X_k)}\right)p_{U}(X_k)dxdu \]

\[ = \int_{\mathcal{R}^2} F\left(\frac{p_{U}(u)p_{X_k}(x)}{p_{U}(X_k)}\right)p_{X_k,u}(x,u)dxdy \quad (16) \]

were \( p_{X_k} \) and \( p_{X_k,u} \) are the probability distribution functions of \( X_k \) and \((X_k, U)\), respectively. Csiszár et al.\(^{59}\) f-divergences imply that these indices are positive and equate to 0 when \( U \) and \( X_k \) are independent. Also, given the formulation of \( S_{X_k}^F \), it is invariant under any smooth and uniquely invertible transformation of the variables \( X_k \) and \( U \) (Kraskov et al.\(^{65}\)). This has an advantage over Sobol sensitivity indices which are invariant under linear transformations.

By substituting the different formulations of \( F \) in equation 16, Da Veiga\(^{56}\)'s work claims to be the first in establishing the link that previously proposed sensitivity indices are actually special cases of more general indices defined through Csiszár et al.\(^{59}\)'s f-divergence. Then equation 16 changes to estimation of ratio between the joint density of \((X_k, U)\) and the marginals, i.e.

\[ S_{X_k}^F = \int_{\mathcal{R}^2} F\left(\frac{1}{r(x,u)}\right)p_{X_k,u}(x,u)dxdy = \varepsilon(x,u)F\left(\frac{1}{r(x,u)}\right) \quad (17) \]

where, \( r(x,y) = (p_{X_k,u}(x,u))/(p_{U}(u)p_{X_k}(x)) \). Multivariate extensions of the same are also possible under the same formulation.

Finally, given two random vectors \( X \in \mathcal{R}^p \) and \( Y \in \mathcal{R}^q \), the dependence measure quantifies the dependence between \( X \) and \( Y \) with the property that the measure equates to 0 if and only if \( X \) and \( Y \) are independent. These measures carry deep links (Sejdinovic et al.\(^{70}\)) with distances between embeddings of distributions to reproducing kernel Hilbert spaces (RKHS) and here the related Hilbert-Schmidt independence criterion (HSIC by Gretton et al.\(^{58}\)) is explained.

In a very brief manner from an extremely simple introduction by Daumé III\(^{71}\) - "We first defined a field, which is a space that supports the usual operations of addition, subtraction, multiplication and division. We imposed an ordering on the field and described what it means for a field to be complete. We then defined vector spaces over fields, which are spaces that interact in a friendly way with their associated fields. We defined complete vector spaces and extended them to Banach spaces by adding a norm. Banach spaces were then extended to Hilbert spaces with the addition of a dot product." Mathematically, a Hilbert space \( \mathcal{H} \) with elements \( r, s \in \mathcal{H} \) has dot product \( \langle r, s \rangle _{\mathcal{H}} \) and \( r \cdot s \). When \( \mathcal{H} \) is a vector space over a field \( \mathcal{F} \), then the dot product is an element in \( \mathcal{F} \). The product \( \langle r, s \rangle _{\mathcal{H}} \) follows the below mentioned properties when \( r, s, t \in \mathcal{H} \) and for all \( a \in \mathcal{F} \).

- Associative: \( (ar) \cdot s = a(r \cdot s) \)
- Commutative: \( r \cdot s = s \cdot r \)
- Distributive: \( r \cdot (s + t) = r \cdot s + r \cdot t \)

Given a complete vector space \( V \) with a dot product \( \langle \cdot, \cdot \rangle \), the norm on \( V \) defined by \( ||r||_V = \sqrt{\langle r, r \rangle} \) makes this space into a Banach space and therefore into a full Hilbert space.

A reproducing kernel Hilbert space (RKHS) builds on a Hilbert space \( \mathcal{H} \) and requires all Dirac evaluation functionals in \( \mathcal{H} \) are bounded and continuous (on implies the other). Assuming \( \mathcal{H} \) is the \( L_2 \) space of functions from \( X \) to \( \mathcal{R} \) for some measurable \( X \). For an element \( x \in \mathcal{X}, \) a Dirac evaluation functional at \( x \) is a functional \( \delta_x \in \mathcal{H} \) such that \( \delta_x(g) = g(x) \). For the case of real numbers, \( x \) is a vector and \( g \) a function which maps from this vector space to \( \mathcal{R} \). Then \( \delta_x \) is simply a function which maps \( g \) to the value \( g \) has at \( x \). Thus, \( \delta_x \) is a function from \((\mathcal{R}^n \rightarrow \mathcal{R})\) into \( \mathcal{R} \).

The requirement of Dirac evaluation functions basically means (via the Riesz\(^{72}\) representation theorem) if \( \Phi \) is a bounded linear functional (conditions satisfied by the Dirac evaluation functionals) on a Hilbert space \( \mathcal{H} \), then there is a unique vector \( f \in \mathcal{H} \) such that \( \Phi g = \langle g, f \rangle _\mathcal{H} \) for all \( f \in \mathcal{H} \). Translating this theorem back into Dirac evaluation functionals, for each \( \delta_x \) there is a unique vector \( k_x \in \mathcal{H} \) such that \( \delta_x g = g(x) = \langle g, k_x \rangle _\mathcal{H} \). The reproducing kernel \( \mathcal{K} \) for \( \mathcal{H} \) is then defined as \( \mathcal{K}(x, x') = \langle k_x, k_{x'} \rangle _\mathcal{H} \), were \( k_x \) and \( k_{x'} \) are unique representatives of \( \delta_x \) and \( \delta_{x'} \). The main property of interest is \( \langle g, \mathcal{K}(x, x') \rangle _\mathcal{H} = g(x') \). Furthermore, \( k_x \) is defined to be a function \( f \rightarrow \mathcal{K}(x, f) \) and thus the reproducibility is given by \( \mathcal{K}(x, \cdot), K(y, \cdot) \mathcal{H} = \mathcal{K}(x, y) \).

Basically, the distance measures between two vectors represent the degree of closeness among them. This degree of closeness is computed on the basis of the discriminative patterns inherent in the vectors. Since these patterns are used implicitly in the distance metric, a question that arises is, how to use these distance metric for decoding purposes?

The kernel formulation as proposed by Aizerman et al.\(^{60}\), is a solution to our problem mentioned above. For simplicity, we consider the labels of examples as binary in nature. Let \( x_i \in \mathcal{R}^n \), be the set of \( n \) feature values with corresponding category of the example label \( (y_i) \) in data set \( D \). Then the data points can be mapped to a higher dimensional space \( \mathcal{H} \) by the transformation \( \Phi \):

\[ \Phi : x_i \in \mathcal{R}^n \rightarrow \Phi(x_i) \in \mathcal{H} \quad (18) \]
This \( \mathcal{H} \) is the *Hilbert Space* which is a strict inner product space, along with the property of completeness as well as separability. The inner product formulation of a space helps in discriminating the location of a data point w.r.t a separating hyperplane in \( \mathcal{H} \). This is achieved by the evaluation of the inner product between the normal vector representing the hyperplane along with the vectorial representation of a data point in \( \mathcal{H} \). Thus, the idea behind equation (18) is that even if the data points are non-linearly clustered in space \( \mathbb{R}^n \), the transformation spreads the data points into \( \mathcal{H} \), such that they can be linearly separated in its range in \( \mathcal{H} \).

Often, the evaluation of dot product in higher dimensional spaces is computationally expensive. To avoid incurring this cost, the concept of kernels in employed. The trick is to formulate kernel functions that depend on a pair of data points in the space \( \mathbb{R}^n \), under the assumption that its evaluation is equivalent to a dot product in the higher dimensional space. This is given as:

\[
\kappa(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle
\]  

(19)

Two advantages become immediately apparent. First, the evaluation of such kernel functions in lower dimensional space is computationally less expensive than evaluating the dot product in higher dimensional space. Secondly, it relieves the burden of searching an appropriate transformation that may map the data points in \( \mathbb{R}^n \) to \( \mathcal{H} \). Instead, all computations regarding discrimination of location of data points in higher dimensional space involves evaluation of the kernel functions in lower dimension. The matrix containing these kernel evaluations is referred to as the *kernel matrix*. With a cell in the kernel matrix containing a kernel evaluation between a pair of data points, the kernel matrix is square in nature.

As an example in practical applications, once the kernel has been computed, a pattern analysis algorithm uses the kernel function to evaluate and predict the nature of the new example using the general formula:

\[
f(z) = \langle w, \phi(z) \rangle + b
\]

\[
= \sum_{i=1}^{N} \alpha_i y_i \phi(x_i), \phi(z) > + b
\]

\[
= \sum_{i=1}^{N} \alpha_i y_i \langle \phi(x_i), \phi(z) \rangle > + b
\]

\[
= \sum_{i=1}^{N} \alpha_i y_i \kappa(x_i, z) + b
\]  

(20)

where \( w \) defines the hyperplane as some linear combination of training basis vectors, \( z \) is the test data point, \( y_i \) the class label for training point \( x_i \), \( \alpha_i \) and \( b \) are the constants. Various transformations to the kernel function can be employed, based on the properties a kernel must satisfy. Interested readers are referred to Taylor and Cristianini\(^ {73} \) for description of these properties in detail.

The Hilbert-Schmidt independence criterion (HSIC) proposed by Gretton et al.\(^ {58} \) is based on kernel approach for finding dependences and on cross-covariance operators in RKHS. Let \( X \in \mathcal{X} \) have a distribution \( P_X \) and consider a RKHS \( \mathcal{A} \) of functions \( \mathcal{X} \rightarrow \mathbb{R} \) with kernel \( k_X \) and dot product \( \langle \cdot, \cdot \rangle_A \). Similarly, Let \( U \in \mathcal{Y} \) have a distribution \( P_U \) and consider a RKHS \( \mathcal{B} \) of functions \( \mathcal{U} \rightarrow \mathbb{R} \) with kernel \( k_U \) and dot product \( \langle \cdot, \cdot \rangle_B \). Then the cross-covariance operator \( C_{X,U} \) associated with the joint distribution \( P_{X,U} \) of \( (X, U) \) is the linear operator \( \mathcal{B} \rightarrow \mathcal{A} \) defined for every \( a \in \mathcal{A} \) and \( b \in \mathcal{B} \) as :

\[
\langle a, C_{X,U} b \rangle_A = \mathcal{E}_X [a(X), b(U)] - \mathcal{E}_X a(X) \mathcal{E}_U b(U)
\]  

(21)

The cross-covariance operator generalizes the covariance matrix by representing higher order correlations between \( X \) and \( U \) through nonlinear kernels. For every linear operator \( C : \mathcal{B} \rightarrow \mathcal{A} \) and provided the sum converges, the Hilbert-Schmidt norm of \( C \)
Finally, assuming were the equality in terms of kernels is proved in Gretton et al.\textsuperscript{58}. The dissimilarity between density functions is measured with Csiszar f-divergences. Estimation is performed through kernel density estimation and the function kde of the package ks. (Borgonovo\textsuperscript{53}, Da Veiga\textsuperscript{56})

\section{Choice of sensitivity indices}

The sensitivity package (Faivre et al.\textsuperscript{74} and Iooss and Lemaitre\textsuperscript{21}) in R language provides a range of functions to compute the indices and the following indices will be taken into account for addressing the posed questions in this manuscript.

1. \textsc{sensiFdiv} - conducts a density-based sensitivity analysis where the impact of an input variable is defined in terms of dissimilarity between the original output density function and the output density function when the input variable is fixed. The dissimilarity between density functions is measured with Csiszar f-divergences. Estimation is performed through kernel density estimation and the function kde of the package ks. (Borgonovo\textsuperscript{53}, Da Veiga\textsuperscript{56})

2. \textsc{sensiHSIC} - conducts a sensitivity analysis where the impact of an input variable is defined in terms of the distance between the input/output joint probability distribution and the product of their marginals when they are embedded in a Reproducing Kernel Hilbert Space (RKHS). This distance corresponds to HSIC proposed by Gretton et al.\textsuperscript{58} and serves as a dependence measure between random variables.

3. \textsc{soboljansen} - implements the Monte Carlo estimation of the Sobol indices for both first-order and total indices at the same time (all together 2p indices), at a total cost of (p+2) × n model evaluations. These are called the Jansen estimators. (Jansen\textsuperscript{48} and Saltelli et al.\textsuperscript{40})

4. \textsc{sobol2002} - implements the Monte Carlo estimation of the Sobol indices for both first-order and total indices at the same time (all together 2p indices), at a total cost of (p+2) × n model evaluations. These are called the Saltelli estimators. This estimator suffers from a conditioning problem when estimating the variances behind the indices computations. This can seriously affect the Sobol indices estimates in case of largely non-centered output. To avoid this effect, you have to center the model output before applying "sobol2002". Functions "soboljansen" and "sobolmartinez" do not suffer from this problem. (Saltelli\textsuperscript{34})

5. \textsc{sobol2007} - implements the Monte Carlo estimation of the Sobol indices for both first-order and total indices at the same time (all together 2p indices), at a total cost of (p+2) × n model evaluations. These are called the Mauntz estimators. (Saltelli and Annoni\textsuperscript{37})

6. \textsc{sobolmartinez} - implements the Monte Carlo estimation of the Sobol indices for both first-order and total indices using correlation coefficients-based formulas, at a total cost of (p + 2) × n model evaluations. These are called the Martinez estimators.

7. \textsc{sobol} - implements the Monte Carlo estimation of the Sobol sensitivity indices. Allows the estimation of the indices of the variance decomposition up to a given order, at a total cost of (N + 1) × n where N is the number of indices to estimate. (Sobol\textsuperscript{20})

\section{10 Optimization and Support Vector Machines}

Aspects of SVMs from Sinha\textsuperscript{75} are reproduced for completeness.
10.1 Optimization Problems

10.1.1 Introduction

The main focus in this section is optimization problems, the concept of Lagrange multipliers and KKT conditions, which will be later used to explain the details about the SVMs.

10.1.2 Mathematical Formulation

Optimization problems arise in almost every area of engineering. The goal is to achieve an almost perfect and efficient result, while carrying out certain procedures of optimization. Our main source of reference on this topic derives from Bletzinger carrying out certain procedures of optimization. We will be introducing in order to solve the problem. Thus, the Lagrangian formulation, for Eqn. 28 becomes:

\[ L(x, \lambda, \mu) = f(x) + \sum_{j=1}^{p} \lambda_j g_j(x) + \sum_{j=1}^{q} \mu_j h_j(x) \]  

(30)

where \( L \) is the Lagrangian, \( \lambda \) and \( \mu \) are the vectors of the Lagrange multipliers for inequality and equality constraints, respectively.

Next comes the solving of the Lagrangian. We try to derive a solution in terms of variables used and show that the final solution achieved by Equ. 28 and Eqn. 30 remains the same. For the sake of derivation, we assume that each of the vectors \( x, \lambda \) and \( \mu \) have a single element and also there exists a single optimal solution. We will then generalize the solution to vectors containing various elements. Let \( x^*, \lambda^* \) and \( \mu^* \) be the optimal solution for the Lagrangian. Let \( x^1 \) be the optimal solution for \( f(x) \). To begin with, our Lagrangian has the form:

\[ L(x, \lambda, \mu) = f(x) + \lambda g(x) + \mu h(x) \]  

(31)

Derivation:

- **Step 1**: Differentiate the Lagrangian in Eqn. 31 w.r.t \( x \) and equate it to zero.

  \[ \frac{\partial L}{\partial x} = \frac{\partial f}{\partial x} + \lambda \frac{\partial g}{\partial x} + \mu \frac{\partial h}{\partial x} = 0 \]  

(32)

- **Step 2**: Find \( x \) in terms of \( \lambda \) and \( \mu \), such that \( x = x(\lambda, \mu) \).

- **Step 3**: Differentiate the Lagrangian in Eqn. 31 w.r.t \( \lambda \) and equate it to zero.

  \[ \frac{\partial L}{\partial \lambda} = g(x) = 0 \]  

(33)

- **Step 4**: Differentiate the Lagrangian in Eqn. 31 w.r.t \( \mu \) and equate it to zero.

  \[ \frac{\partial L}{\partial \mu} = h(x) = 0 \]  

(34)

- **Step 5**: Substitute \( x(\lambda, \mu) \) in Eqn. 33 and Eqn. 34 to get two equations in two unknowns \( \lambda \) and \( \mu \) and solve to get...
Lastly, since $g(x)$ in Eqn. 33 is an inequality constraint, we have:

$$\lambda g(x) = 0 \quad \lambda \geq 0 \quad (39)$$

10.1.4 Dual Functions

For sake of simplicity, let us for a moment ignore the equality constraint. Then the Lagrangian becomes:

$$L(x, \lambda, \mu) = f(x) + \lambda g(x). \quad (40)$$

It is sometimes easy to transform the Lagrangian into a simpler form, in order to find an optimal solution. We can represent the Lagrangian as a dual function in such a manner that the optimal solution defined as minimum of $L(x, \lambda^*)$ w.r.t $x$ where $\lambda = \lambda^*$, can be represented as the maximum of dual function $D(\lambda)$ w.r.t $\lambda$. For a given $\lambda$, the dual is evaluated by finding the minimum of $L(x, \lambda)$ w.r.t $x$. Thus to find the optimal point we evaluate:

$$\max_{\lambda} D(\lambda) = \min_x L(x, \lambda^*) \quad (41)$$

So the basic steps to solve the dual problem are as follows: **Step 1:** Minimize $L(x, \lambda)$ w.r.t $x$, and find $x$ in terms of $\lambda$. **Step 2:** Substitute $x(\lambda)$ in $L$ s.t. $D(\lambda) = L(x(\lambda), \lambda)$. **Step 3:** Maximize $D(\lambda)$ w.r.t $\lambda$.

10.1.5 Karush Kuhn Tucker Conditions

The derivation in last part (Eqn. 31 to Eqn. 39) gives us a set of equations that need to be evaluated along with the consideration of constraints present. These set of equations and constraints in terms of the Lagrangian, form the Karush Kuhn Tucker Conditions. We give here the generalized KKT conditions and explain the necessary details.

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_{j=1}^{p} \lambda_j \frac{\partial g_j}{\partial x_i} + \sum_{j=1}^{p} \mu_j \frac{\partial h_j}{\partial x_i} = 0 \quad i = 1, \ldots, n \quad (42)$$

where $L$ is Eqn. 30.

The KKT conditions specify a few points which are as follows:

1. The first line states that the linear combination of objective and constraint gradients vanishes.

2. A prerequisite of the KKT conditions is that the gradients of the constraints must be continuous (evident from second and third lines in Eqn. 42).

3. The last two lines in Eqn. 42 state that at optimum either the constraints are active or the constrains are inactive.

10.2 Support Vector Machines

Armed with the knowledge of optimization problems and concept of Lagrange multipliers, we now delve into the workings of support vector machines. Burges provides a good introduction to SVMs and is our main reference. Interested readers should refer to Cristianini and Shawe-Taylor, Schölkopf and Smola, and Vapnik and Vapnik for detailed references.

10.2.1 Separable Case

Let us suppose that we are presented with a data set that is linearly separable. We assume that there are $m$ examples of data in the format $\{x_i, y_i\}$, s.t. $x_i \in \mathbb{R}^n$; $i = 1, \ldots, m$, where $y_i \in \{-1, 1\}$ is the corresponding true label of $x_i$. We also suppose there is an existence of a linear hyperplane in the $n$ dimensional space that separates the positively labeled data from the negatively labeled data. Let this separating hyperplane be given by

$$w \cdot x + b = 0 \quad (43)$$

where, $w$ is the normal vector $\perp$ to the hyperplane and $|b|/||w||$ is the shortest perpendicular distance of the hyperplane to the origin. $||w||$ is the Euclidean norm of $w$. The margin of a hyperplane is then defined as the minimum of the distance of the positively and negatively labeled examples, to the hyperplane. For the linear case, the SVM searches for the hyperplane with largest margin. We now have three conditions, based on the location of
an example \( \mathbf{x}_i \) w.r.t the hyperplane:

\[
\mathbf{w} \cdot \mathbf{x}_i + b = 0 \quad : \quad \text{example lying on the hyperplane.} \quad (44)
\]
\[
\mathbf{w} \cdot \mathbf{x}_i + b \geq +1 \quad : \quad \text{positively labeled example.} \quad (45)
\]
\[
\mathbf{w} \cdot \mathbf{x}_i + b \leq -1 \quad : \quad \text{negatively labeled example.} \quad (46)
\]

Combining the equality and the two inequalities we have:

\[
y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0 \quad \text{Eqn. 47}
\]

Since the SVMs search for the largest margin, we now try to find a mathematical expression of the margin. Considering the examples that satisfy equality in Eqn. 45, the distance of the closest positive example can be expressed as \( |1-b|/||\mathbf{w}|| \). Similarly, considering the negative examples that satisfy equality in Eqn. 46, the distance of the closest negative example can be expressed as \( |-1-b|/||\mathbf{w}|| \). On summation of the two shortest distances, we get the margin of the hyperplane as \( 2/||\mathbf{w}|| \). Since the labels are \( \{-1, 1\} \), no example lies inside the hyperplanes representing the margin in this case. Taking into account that the SVM searches for the largest margin, we can say that it can be achieved by minimizing \( ||\mathbf{w}||^2 \), subject to the constraints in Eqn. 47. Examples lying on the hyperplanes of the margins are termed support vectors, as their removal would change the margin and thus the solution. Figure 11 represents the conceptual points about separating hyperplanes.

10.3 Lagrangian Representation: Separable Case

Clearly, the previous paragraph shows that finding the margin is a problem of optimization as the goal is to minimize \( ||\mathbf{w}||^2 \) subject to constraints in Eqn. 47. Employing the ideas of Chapter 3, the Lagrangian for the above problem, is:

\[
L(\mathbf{w}, b, \alpha) = \frac{1}{2} ||\mathbf{w}||^2 - \sum_{i=1}^{m} \alpha_i y_i (\mathbf{x}_i \cdot \mathbf{w} + b) + \sum_{i=1}^{m} \alpha_i \quad (48)
\]

where \( \frac{1}{2} ||\mathbf{w}||^2 \) is the objective function, \( \alpha \) is the Lagrangian multiplier and the Eqn. 47 is the inequality constraint. Since the minimization of the objective function is required, we employ the ideas of the derivation of KKT conditions (Eqn. 42) to Eqn. 48.

In short, we would require the \( L(\mathbf{w}, b, \alpha) \) to be minimized w.r.t \( \mathbf{w} \) and \( b \) and also require its derivative w.r.t all \( \alpha_i \)'s to vanish. Thus the KKT conditions take the form:

\[
\frac{\partial L}{\partial \mathbf{w}} = w_i - \sum_{i=1}^{m} \alpha_i y_i \mathbf{x}_i = 0 \quad : \quad j = 1,...,n
\]
\[
\frac{\partial L}{\partial b} = -\sum_{i=1}^{m} \alpha_i y_i = 0 \quad : \quad i = 1,...,m
\]
\[
\frac{\partial L}{\partial \alpha_i} = y_i (\mathbf{x}_i \cdot \mathbf{w} + b) + \sum_{i=1}^{m} \alpha_i = 0 \quad : \quad i = 1,...,m
\]
\[
\alpha_i \geq 0 \quad : \quad i = 1,...,m
\]

Thus solving the SVMs is equivalent to solving the KKT conditions. While \( \mathbf{w} \) is determined by the training set, \( b \) can be found by solving the penultimate equation in Eqn. 49 for which \( \alpha_i \neq 0 \). Also note that examples that have \( \alpha_i \neq 0 \) form the set of support vectors.

The dual problem for the same Lagrangian is:

\[
D = \sum_{i} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \quad (50)
\]

Solving for Eqn. 50 requires maximization of \( D \) w.r.t \( \alpha_i \), subject to second line of Eqn. 49 and positivity of \( \alpha_i \), with the solution given by first line of Eqn. 49.

To classify or predict the label of a new example \( \mathbf{x}_{\text{new}} \), the SVM has to evaluate \( (\mathbf{x}_{\text{new}} \cdot \mathbf{w} + b) \) and check the sign of the evaluated value. A positive sign would lead to assignment of a +1 label and a negative sign to -1.

10.4 Nonseparable Case

For many classification problems, the data present is nonseparable. To extend the idea to nonseparable case, some amount of cost is added, which takes care of particular cases of examples. This is achieved by introducing slack in the constraints Eqn. 45 and Eqn. 46 (Burges, Vapnik and Vapnik). The equations

\[
\sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \quad (50)
\]
then becomes

\[ w \cdot x_i + b \geq 1 - \xi_i : \text{ positively labeled example. (51)} \]

\[ w \cdot x_i + b \leq -1 + \xi_i : \text{ negatively labeled example. (52)} \]

For an error to occur, the \( \xi_i \) value must exceed unity. To take care of the cost of errors, a penalty is introduced which changes the objective function from \( ||w||^2 / 2 \) to \( ||w||^2 / 2 + C(\sum_i \xi_i)^k \). Thus \( \sum_i \xi_i \) represents the upper bound on the training error. For quadratic problems, \( k \) can be 1 or 2.

### 10.5 Lagrangian Representation: Nonseparable Case

Since the formulation of the Lagrangian and its dual follow the same procedure, as mentioned before, we only mention the equations. The Lagrangian for nonlinear nonseparable case is:

\[
L(w, b, \alpha, \xi) = \frac{1}{2}||w||^2 + C \sum_{i=1}^{m} \xi_i - \sum_{i=1}^{m} \alpha_i \{y_i(x_i \cdot w + b) - 1 + \xi_i\} = \sum_{i=1}^{m} \mu_i \xi_i
\]

The corresponding KKT conditions are:

\[
\begin{align*}
\frac{\partial L}{\partial w} &= w_j - \sum_{i=1}^{m} \alpha_i y_i x_{ij} = 0 & j &= 1, \ldots, n \\
\frac{\partial L}{\partial \alpha_i} &= y_i(x_i \cdot w + b) - 1 + \xi_i = 0 & i &= 1, \ldots, m \\
\frac{\partial L}{\partial \xi_i} &= C - \alpha_i - \mu_i = 0 & i &= 1, \ldots, m \\
\alpha_i \geq 0 & i &= 1, \ldots, m \\
\xi_i \geq 0 & i &= 1, \ldots, m \\
\mu_i \geq 0 & i &= 1, \ldots, m
\end{align*}
\]

The dual formulation \( k = 1 \) for the Lagrangian just discussed is:

\[
D = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j
\]

All the previous conditions remain same, except that the Lagrangian multiplier \( \alpha_i \) now has a upper bound of value \( C \). The solution for the dual is given by \( w = \sum_{i=1}^{N_S} \alpha_i y_i x_i \). \( N_S \) is the number of support vectors. Figure 12 depicts the nonseparable case.

### 10.6 Kernels and Space Dimensionality Transformation

The above cases were for linear separating hyperplanes. In order to generalize for nonlinear cases, Boser et al.\(^{81}\) employed the idea of Aizerman\(^{60}\) as follows: Since the Dual in Eqn. 55 and its corresponding constraint equations employ the dot product of the examples, \( x_i \cdot x_j \), it was proposed to map the data in a higher dimensional space using a function \( \phi \) s.t. the algorithm would depend only on dot products in the higher space. Next, the existence of a function called kernel, dependent on \( x_i \) and \( x_j \), was assumed s.t. the value reported by the kernel was equal to the value resulting from the dot product in the higher space. A mathematical representation of the above concept is:

\[
\phi : \mathbb{R}^n \rightarrow H
\]

\[
K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)
\]

where \( H \) is a higher dimensional space.

This technique drastically reduces the amount of work required while dealing with nonlinear separating hyperplanes, concerning search for appropriate \( \phi \). Instead, one only works with \( K(x_i, x_j) \), in place of \( x_i \cdot x_j \). For classification purpose, where the sign of the function \( (x_{new} \cdot w + b) \) is evaluated, the formulation employing kernels become:

\[
f(x_{new}) = (w \cdot x_{new} + b) = \sum_{i=1}^{N_S} \alpha_i y_i \phi(s_i) \cdot \phi(x_{new}) + b
\]

where \( s_i \) are the support vectors.
Fig. 13 $F_{\text{div Hellinger}}$; Training sample size - 4; Test sample size - 16; Case - Normal
Fig. 15 FdivTV; Training sample size - 4; Test sample size - 16; Case - Normal
Fig. 17 HSIClaplace; Training sample size - 4; Test sample size - 16; Case - Normal
Fig. 19 HSIC/rbf; Training sample size = 4; Test sample size = 16; Case = Normal