1 Types of Analytics

- Which are the different types of analytics?

Analytics is the discovery, interpretation, and communication of meaningful patterns in data. When talking about analytics, we can differentiate between three major areas of study: descriptive, predictive and prescriptive.

- Descriptive Analytics can be used to describe the past, where data structures are used to efficiently save and extract data.
- Predictive Analytics, uses past data dot create predictive models.
- Prescriptive Analytics uses heuristic models, optimizations, and simulations to prescribe optimal actions.

- What are the question that the different types of analytics want to answer?

Descriptive Analytics are less difficult and provide less value, help to answer questions like: "what has happened?", "why did it happen?", or "what is happening now?". Predictive Analytics stay in between information and optimization purpose, help to answer question like: "what will happen?", or "why will it happen?". Finally Prescriptive Analytics are the most difficult but provide the most value, help to answer questions like: "what should I do?", or "why should I do it?".

- Provide an explanation of Prescriptive.

It determines what should happen and how to make it happen by identifying those factors that contribute do desirable/undesirable outcomes. An example cold be finding a set of prices and and advertising frequency to maximize economic return.

2 Types of Decision-Making Problems

- Describe the type of decision making problems and the relative decisions.

Decision-Making is a goal-oriented process. It is divided in: Unstructured problems and Structured problems.

- Unstructured Problems: are problems that are new, or unusual, and for which information is ambiguous or incomplete, they require ad-hoc solutions. They have "Non-Programmed Decisions", mining that the decisions are unique, non-recurring and generate unique responses.
- Structured Problems: are problems where the objectives are clear, they have occurred before, and they are completely defined. They have "Programmed Decisions", meaning that the decision are repetitive and can be managed by a routine-based approach, the difficulty of this process is related to the number of decisions.

Structured problems are based on multiple short-terms solutions, or if tactical on medium-term solutions, while Unstructured problems are characterized by few possible long-term solutions.

- Talk about certainty and uncertainty

Certainty is referred to a situation in which the results of all possible decisions are known, and we can make accurate decisions. Uncertainty is when we have no knowledge of the outcome of the various possible choices. In case of uncertainty the probabilistic estimation of the result allows to estimate the probability of the results of the choices, called "risk".

- Why we need for decision models?

S decision model is a symbolic representation of assumptions about reality. The foal of optimization is to maximize or minimize a certain performance function relative to some possible solutions that meet a set of constraints. Models are important to avoid anchoring effect and framing effect. Anchoring effect is when trivial factors influence the initial reasoning on a certain problem, it result that some problems and their solutions are underestimated. The framing effect refers to a win-loss prospective to see a problem, that allow irrational choices. Decision Models help in making good decisions but cannot guarantee that good outcomes will always occur as a result of those decisions.

3 Linear programming problems formulation

- Explain the formulation of a LP problem.

A linear problem is an optimization problem, meaning that we want to maximize or minimize a function following some constraints. An optimization problem has the following characteristics:

- Objectives: an objective function expresses the main aim of the model which is either to be minimized or maximized.
- Decisions: a set of variables which control the value of the objective function.
- Constraints: a set of constraints that allow the unknowns to take on certain values but exclude others.

In a LP both objective functions and the constraint functions are linear.

4 Graphical solution of linear programming problems

- Explain how to implement the graphical solution of LP problems.

For bi-variate LP problems it is simple to identify the feasible region through graphs. The feasible region is iteratively defined by applying the constrains, which will progressively shrink the final region. After having defined the feasible region, the objective function is plotted and the vertex that maximize (or minimize) the region is identified as the optimal solution. The optimal solution will always occur at a vertex of the feasible region. Another way to identify the optimal solution is by enumerating the extreme points and identify the one that maximize/minimize the objective function.

5 Types of solutions (unboundedness, multiple solutions, unfeasibility)

- Illustrate the special conditions in LP Models.

In LP we can encounter various anomalies:

- Alternate optimal solutions: when one of the side of the feasible region coincides with part of the objective function, all points in that segment are an optimal solution
- Redundant constraints: when there are constraints that do not contribute to define the feasible region.
- Unbounded solutions: when the feasible region extends into an infinite space, it is not possible to determine optimal vertices, there are infinite solutions.
- Infeasibility: when the feasible region is empty, there are no possible solutions.

- Which are the assumptions in LP?

The linearity of LP objective function has two implications:

- Proportionality: contribution to the objective function from each decision variable is proportional to the value of that variable (not valid if non linear).
- Additive: contribution to the objective function value form any decision variable is independent of the values of the other decision variables.

The linearity of constraints has two implications:

- Proportionality: the contribution of each decision variable to LHS of each constraint is proportional to the value of the variable.
- Additivity: the contribution of a decision variable to LHS of a constraint is independent of the values of the other decision variables.

In general in LP we make two important assumptions:

- Divisibility Assumption: each decision variable is allowed to assume fractional values.
- Certainty Assumption: each parameter is known with certainty.

6 Sensitivity analysis

- Explain the Simplex Method.

Simplex method is an algorithm to solve LP problems. It works by moving iteratively form one vertex of the feasible region to another along edges, until the optimal solution is found. It is highly efficient, especially for large-scale problems. We first need to convert all inequalities to equalities by adding "slack variables" to $\leq$ constraints and subtracting them to $\geq$ ones. Slack variables with coefficient -1 are called "surplus variables". The goal of Simplex Method is to find the intersection of the lines that define the vertices of the feasible region and then find the optimal vertex.

Since with Slack Variables we have more variables, some of them are set to zero ("non-basic variables") and the system is solved with only the "basic variables" (number equations).

Moving form an extreme point to the adjacent one involve switching one of the basic variables with one of the nonbasic variables. The algorithm stop when no adjacent extreme point yields to a better objective function value.

It is not guaranteed that the algorithm follow the best path, but the convergence is guaranteed.

- What is the purpose of Sensitivity Analysis?

Sensitivity analysis helps answer questions about how sensitive the optimal solution is to changes in various coefficients in a model. We can answer questions about:

- amount by which objective function coefficients can change without changing the optimal solution.
- the impact on the optimal objective function value of changes in constrained resources.
- The impact on the optimal objective function value of forced changes in decision variables.
- The impact changes in constraint coefficients will have on the optimal solution.

Any change on a "binding constraint" (constraints whose line rest on the optimal solution) will also change the optimal solution. Binding constraints have slack variables to 0. If slack variable of a c constraint is positive it represent the difference between the LHS and RHS of the constraint (unused resources).

Changes in the coefficients of the objective function can change the slope, and so the optimal solution.

- Explain the values that appear in the Sensitivity Report.

- Optimality Ranges: indicate the amount by which and objective function coefficient can change without changing the optimal solution (assuming all other coefficients constant). If the range is 0 it means that "alternate optimal solution" exist (if we are not in degenerate solution case).

- Shadow Price: is relative a constraint, int indicates the amount by which the objective function value changes given a unit increase in RHS value of the constraint (assuming all other coefficients constant). Shadow prices are valid only if RHS changes fall in a range, for non binding constraints they are always zero.

- Reduced Cost: for each product is equal its per-unit marginal profit minus the per-unit value of the resources it consumes (priced at their shadow prices).

Shadow prices of resources equate the marginal value of the resources consumed with the marginal benefit of the goods being produced. Resources in excess supply have a shadow price of zero. The reduced cost of a product is the difference between its marginal profit and the marginal value of the resources it consumes. Products whose marginal profits are less then the marginal value of the foods required for the production will not be produced in an optimal solution.

- What happen in the case of Simultaneous changes in objective function coefficients?

If simultaneous changes in objective function coefficients occur we apply the "100% rule" to determine if the optimal solution changes. Two cases can occur:

- All variables with changed objective function coefficients have non-zero reduced cost.

- At least one variable with changed objective function coefficients have zero reduced cost.

In first case the solution remain optimal, if the changes fall in the optimality ranges.

In second case it is needed a an algorithm to study the impact on the solution. We calculate the percentage of change with respect to the related admissible range. So each variable compute the objective function coefficients:

$$\begin{cases} \frac{\Delta c_j}{I_j} & \text{if } \Delta c_j >= 0 \\ \frac{-\Delta c_j}{D_j} & \text{if } \Delta c_j < 0 \end{cases}$$

where:

- Δc_j: the change made to the coefficient c_j
- I_j: the possible range of increase
- D_j: the possible range of decrease

If the sum of r_j is ≤ 1 the solution remain optimal, otherwise the solution might remain optimal, but is not guaranteed.

- What does it means that a LP problem is degenerate, what it are the consequences?
The solution to a LP problem is degenerate if the allowable increase or decrease of any constraint is zero. When it happen:

- Range is 0 cannot be used to detect if "alternate optimal solution" exist.
- Reduced costs for variables may not be unique. Also objective function coefficients for the variables must change buy at least as much as their respective deduced cost before the optimal solution would change.
- Optimality ranges for objective function coefficients still hold.
- The shadow prices and their ranges may still be interpreted in the usual way but may not be unique.

7 Formulation of integer linear programming problems

- How is characterized an ILP problem?
When one ore more variables in an LP problem must assume an integer value we have an ILP. Integer variables allow to build more accurate models. Integrity conditions are easy to state but make the problem much more difficult to solve.

Rounding the optimal solution for the relaxed problem dos not work because it could be infeasible or sub-optimal.

To solve ILP problem we can use the simplex method ore we can use the Branch-and-Bound technique.

- Which are the challenges faced when solving an Integer Linear Programming problem with respect to a linear programming one and how can be approached ?
The B&B algorithm is theoretically capable of solving any ILP, but it often requires a considerable amount of computational effort and time. Most ILP use a sub-optimality tolerance factor, which allowed to stop at once an integer solution is found that is within a certain percentage of the global optimal solution.

When using this strategy the bounds optioned form LP relaxation are very useful because optimal solution to an LP relaxation of an ILP problem provides us with an estimate of the optimal objective function.

- For maximization problem, the optimal relaxed objective function value is an upper bound of the optimal integer value.
- For minimization problems, the optimal relaxed objective function value is a lower bound on the optimal integer value.

8 The basic steps of Branch&Bound method

- Explain the B&B algorithm, naming the phases.
In the first step called "initialization" we relax all integrity conditions in ILP and solve the problem, the solution if not satisfy the integrity condition is used as a bound, and $Z_{best} = \pm\infty$.

In the second step named "Branching" we divided the problem optioned in the initialization step in two sub-problems, using a threshold on one of the variable that violate the integrity. The sub-problems are placed in a list of candidate problems to be solved. This step aimed at eliminating some unfeasible points from further consideration.

The third step is "Bounding" and it aimed at reducing the number of generated sub-problems. If the list is not empty, it solve a problem form the candidate ones. If it is infeasible the we go back at step 3. If it not satisfy integrity conditions we proceed with step 2. If it does and it is better then Z_{best} it become the new Z_{best}, and restart the step 3. If it does and is not better then Z_{best} go beck to step 3.

When the list is empty we pass to step 4, "Stop", where the current Z_{best} is the optimal solution.

- In the Branch and Bound algorithm, how are the LB and UB of each node determined?

When using this strategy the bounds optioned form LP relaxation are very useful because optimal solution to an LP relaxation of an ILP problem provides us with an estimate of the optimal objective function.

- For maximization problem, the optimal relaxed objective function value is an upper bound of the optimal integer value.
- For minimization problems, the optimal relaxed objective function value is a lower bound on the optimal integer value.

9 Network modeling

- What are the building blocks of a network model?

In many business scenarios it is helpful to represent the problem graphically as a network consisting of nodes and arcs.

In a network there are thee types of nodes:

- Supply: represent sources of goods or materials, represented by negative numbers.
- Demand: represent destinations or customers, represented by positive numbers
- Transshipment: are intermediate points where goods can be transferred from one route go another,

- Which are the characteristics of a Transshipment Problem?

In a network each arc have a direction. The objective of a network problem is to determine the number of objects that flow through the arcs by transforming it into a LP problem. The arcs become the decision variables, the objective is to minimize the total cost, each node define a constraint. Based on the correct inequality between supply and demand the correct rule must be applied for each node.

- Which are the characteristics of a Shortest Path Problem, (talk also about Equipment Replacement Problem)?

The shortest path problem want to fine the path that minimize the travel time (on the contrary the most scenic route want to maximize the scenic rating points). In this type of problem we have only one supply note with a supply of -1, only one demand node with a demand of $+1$, and all other nodes have supply/demand of 0. In this type of problems the coefficients are integer, then also the optimal solution is integer.

The problem of determining when to replace equipment can be modeled as a shortest path problem. Where each arc represents a choice, and once the problem is modeled with a graph it can be solved by applying the shortest path algorithm.

- Which are the characteristics of a Transportation and Assignment Problem?

In transportation and assignment problem we do not have Transshipment nodes, only Supply and Demand nodes. It is not a rule that the supply nodes are connected to all the demand nodes, but the problem can be set it up in this way by inserting the missing arcs with a very high cost, so it will never be chosen for the final solution.

- Which are the characteristics of a Generalized Network Flow Problem?
In some problems, a gain or loss occurs in flows over arcs, it is then required some modeling changes. In generalized problem, each arc is represented by the cost and the reduction of flow (reduction factor).

- Which are the characteristics of a Maximal Flow Problem, which is its dual problem and how you can solve both the primal and the dual problem?
In some network problems, the objective is to determine the maximum amount of flow that can occur through a network. The arcs in these problems have two limits, an upper and a lower flow limit. This type of problems can be solved as a simple Transshipment problem.

The network can be transformed in to a Transshipment network by adding an arc form the final node to the initial node and assigning a demand of 0 to all nodes in the network, and finally maximizing the flow on the arcs.

The maximum flow problem is a dual problem, corresponding to the search for the minimum cut (bottleneck), that is the cut with the minimum cost in terms of the cumulative capacity among the various capacities assigned to the edges. It is always possible to divide the nodes in two subset where the arcs going form a subset to another are all saturated (at maximum capacity).

- Which are the characteristics of a Minimal Spanning Tree Problem?
For a network with n nodes, a "spanning tree" is a set of $n-1$ arcs tat connect all the nodes and contain no loops. The minimal spanning tree problem involves determining the set of arcs that connects all the nodes at minimum cost. This problem can be solved with LP but also with greedy algorithms (algorithm that finds a solution by accepting at each step a current sub-optimal solution, and iteratively improving it).

In this network there are no directed edges, and the objective is to determine which edge to activate and which to deactivate.

In the first step we selected a node and identify it as the current sub-network. In the second step we ad to the sub-network the cheapest arc (tie can be broken arbitrary). In the thirds step if the nodes are all in the network we stop, otherwise we go back to step 2.

10 Single variable Nonlinear optimization: Bisection method, Newton method

- Explain the difference that the optimal of an NLP problem has form an LP one.
An NLP problem as a non linear objective function and/or one or more non linear constants. In NLP the optimal solution is not necessarily at a corner of the feasible region, it also may lie within the feasible region. Simplex method cannot be applied to NLP problems.

Strategies applied to NLP problems are iterative (algorithms that iteratively improve the current solution until finding the optimal, moving form point to point within the feasible region, stopping when no improvement is possible in any direction).

The problem is that the search of an optimal solution depend on the starting point. The algorithm may stop at a local optimal solution, rather then at the global one. It is not guaranteed to fined the global optimal. We should avoid null starting points, and chose values that have the same magnitude as the expected optimal.

- Explain the steps for bisection method.
The bisection method is used to find the roots of a function. It is used in optimization problem applied to the first derivative of a function to find stationary points. If a function have unimodality, and we have two points one for which the function is always positive and one for which the function is always negative, a

root exist in between the two points (could exist more then one, also if both positive or negative a root could exist).

The Bisection method start with two points such that the product of the function applied to them is negative, we dived the range in two by finding a mid point, if the midpoint evaluates to zero then is a root, otherwise the root will be in one of the two divided parts. By recursively subdividing the range, we will find the points of minimum/maximum.

We can decide to make the algorithm stop when the current mid point is not changing much between iterations. To do so we can specify a threshold of "error tolerance", that when is meet the algorithm stops. Another method is to chose in advance the number of iterations.

- Which are the advantages and the drawbacks of bisection method?
The bisection method advantages are that it always converge and that is guaranteed that the root bracket get halved with each iteration.

The drawbacks are that it converges slowly, especially when one of the initial guesses is close to the root. In addition if a function just touches the x-axis the bisection method is unable to find the lower and upper guesses. Some function may even change the sign and do not having a root.

- Which are the two possible type of optimization algorithms, what is a termination criteria?
There are two types of optimization algorithms:

1. Dicotomous: find the root of the equation of the derivative equal to zero and each iteration reduce such interval.

2. Approximation: use local approximation of the function to be optimized, it need a termination criteria

A termination criteria could be:

- the solution is accurate with a given level of tolerance
- very small improvement form one iteration to the next
- the maximum number of iteration has been reached
- the solution diverges
- the solutions are in a loop

- Explain the steps for Univariate Newton method.
Newton method is an algorithm to have a numerical solution of the problem.

In the first step, it fit a quadratic approximation of a function $f(x)$ using both gradient and curvature information at x. In the second step it use the taylor approximation $f(x+h)$ and then foot fining of f' to find $h = -\frac{f'(x)}{f''(x)}$.

The idea is to determine the type of approximation and then optimize the approximation of f at a given point. We find that the new guess $x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$.

At each iteration the next point is defined by moving the x to the point of intersection between the x-axis and the tangent of approximation at point x. The algorithm converges quadratic.

- Which are drawbacks of Newton method?
Newton's method converges quickly but doesn't guarantee convergences. The global convergence is poor, often fail if the starting point is to far form the minimum. It must be used with a globalization strategy which reduces the step length until function decrease is assured. In this method we have no control over the direction in which we are going to search, it depend on the curvature.

11 Multivariable Nonlinear optimization: Gradient method, Newton method

- Explain the steps for the Gradient method.
In multivariate case, the gradient vector is perpendicular to the hyperplane tangent to the contour surfaces of f.

All non linear optimization algorithms carry out two fundamental steps:

1. Stating from x_k chose a search direction d_k
2. Minimize or maximize along that direction to fund a new point: $x_{k+1} = x_k + \alpha_k d_k$ where α_k is a positive scalar called step size.

This strategy is called "Steepest Descent"

It use gradient for maximization and the negative gradient for minimization. The steepest descent method in the first step chose an initial point x_0, in the second step calculate the gradient, in the third step calculate the direction, in the fourth step it calculate the next x_{k+1} and finally it use a single variable optimization method to determine α_k.

To determinate the convergence is possible to use some tolerance parameter, like the absolute difference of the function form two iteration or e the norm of the gradient at the following step.

- Explain the steps for the Multivariate Newton method.
Newton's method approximate $f(x)$ with a quadratic function in the neighborhood of the current point using the Taylor-series expansion of f then optimizes the approximated quadratic function to obtain the new iterate point.

Newton step, it moves to a stationary point of the second order approximation derived form the Taylor-series expansion. If $H(x_k)$ is definite positive than only one iteration is required for a quadratic function to reach the optimum point, form any starting point.

In the first step we set the number of iteration to 0, in the second we chose a starting point x_k, then in the third we calculate the gradient and the Hessian matrix, then in the fourth step we calculate the next $x_{k+1} = x_k - H(x_k)^{-1} - \nabla f(x_k)$. Finally in the last step using a convergence criteria we determine the convergence, if it hasn't converged we return to step two.

- Which of the two (Newton and Gradient) methods is preferred and why?
Newton's method uses both gradient and the Hessian matrix. This usually reduces the number of iteration needed, but increases the computation needed for each iteration. So, for very complex functions, a simpler method is usually faster.

12 Lagrangian duality in Nonlinear Optimization

- Explain KKT?
If in a NLP problem the f, g and h are continuously differentiable, and if x^* is a local optimum, then exist λ^* that satisfies the Karush-Kurner-Tucker system.

This system have two qualification:

- Affine constraints: If g_i and h_j are affine, then a constraint qualification hold at any feasible point x.

- Slater condition: if g_i are convex and h_j are affine, then exist x_i s.t. $g_i(x_i) < 0$ and $h_i(x_i) < 0$ then a constraint qualification hold for any feasible point x.

KKT Theorem gives necessary optimality conditions, but not sufficient ones. For convex problems (f and g are convex and h is affine) x^* that solve the KKT system is the global optimum.

- What is the Lagrangian relaxation? What is and why is needed the Lagrangian dual problem?
(P) is the "primal problem".

Given $\lambda > 0$ the problem that minimize the Lagrange function $L(x, \lambda, \mu)$ is called "Lagrangian realization" of the problem (P).

$\varphi(\lambda, \mu) = \inf_{x \in R^n} L(x, \lambda, \mu)$ is called the "Lagrangian dual function". A dual function is concave, may be equal to $-\infty$ at some pint, and may be not differentiable at some point.

The Lagrangian relaxation provide a lower bound for the optimal value of the problem (P).

The problem that maximize the Lagrangian dual function $\varphi(\lambda, \mu)$ is called "Lagrangian dual problem" of (P). It consists in fining the best lower bound for the optimal value of (P). It is always a convex problem.

- For "weak duality" the optimal for the dual problem is always inferior to the optimal of the primal problem.

- For "strong duality" if (P) is convex exist an optimal solution x^* for which constraint qualification holds. It means that the the optimal for the dual problem is equal to the optimal of the primal problem, and the optimal (λ^*, μ^*) for the dual problem is a KKT multipliers vector associated to x^*.

13 Nonlinear Support Vector Machines for classification problems

- What are the differences between linear SVM, linear SVM with soft margin, and non linear SVM?
The SVM work given a set of objects partitioned in several classes with known labels, the objective is to predict the class of any future objective with unknown label. In linear SVM there are many possible separating hyperplanes, and we look for the one with the maximum margin separation. The problem is equivalent to solve a quadratic programming problem.

Not always the two planes are linearly separable, for this reason we can introduce the slack variables to soft the margins.

We can also consider the problem of non-linear SVM, were we assume that the labels are linearly separable in another space called "feature space", the resulting is a non linear separation in R^n.

- Explain the steps to solve a non-linear SVM.
If two sets A and B are not linearly separable, we can assume that they are linearly separable in other spaces. So we use a map function form the current space to a space H named the "feature space". Te try to linearly separate the images of x_i in H.

To do so we define a primal problem and then derive the dual problem.

$$\begin{cases} \max_\lambda -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j \phi(x_i)^T \phi(x_j) \lambda_i \lambda_j + \sum_{i=1}^{l} \lambda_i \\ \sum_{i=1}^{l} \lambda_i y_i = 0 \\ 0 \le \lambda_i \le C & i = 1, \dots, l \end{cases}$$

In the first step we solve the dual problem λ^*, in the second step we compute $w^* = \sum_{i=1}^{l} \lambda_i y_i \phi(x_i)$, and then in the third one use any $0 < \lambda_i < C$ tor fining b^*: $b^* = \frac{1}{y_i} \sum_{j=1}^{l} \lambda_j^* y_j \phi(x_i)^T \phi(x_j)$. In the fourth step we find the decision function as $f(x) = sign((w^*)^T \phi(x) + b^*)$.

- What is a kernel function?

In SVM is not need to explicitly know the images $\phi(x_i)$, but only $\phi(x_i)^T \phi(x_i)$, and to do so we use kernel functions. A function $k : R^n \times R^n \to R$ is called kernel if there exists a map $\phi : R^n \to H$ such that $k(x,y) = <\phi(x), \phi(y)>$ where $<\cdot,\cdot>$ is a scalar product in H.

Some examples are:

- $k(x,y) = x^T y$
- Polynomial: $k(x,y) = (x^T y + 1)^p$, whit $p \geq 1$
- Gaussian: $k(x,y) = e^{-\gamma ||x-y||^2}$
- $k(x,y) = \tanh(\beta x^T + \gamma)$, with suitable β and γ

The matrix K where $K_{ij} = kx_i, x_j$ is positive definite.

The dual problem use the kernel function to map in the feature space. So the real first step in solving non linear SVM is to chose the kernel.

Finally the Separating surface $f(x) = 0$ is linear in features space and non linear in input space.

14 e-Support Vector Regression

- Explain the purpose of e-SV linear regression.

Having a set of training data $\{(x_1,y_1),...,(x_l,y_l)\}$, in ε-SV regression we aim to find a function f that has at most ε deviation form the targets y_i for all training data, and is as flat as possible. We start with linear regression, considering an affine function $f(x) = w^t x + b$ and set the "tolerance parameter" ε. To achieve flatness we seek a small w, that is we aim to solve a complex quadratic optimization problem to find it.

- Why slack variables are insert in e-SV linear regression, what does the C parameter?

If the tolerance parameter ε is to small the model may be infeasible. For this reason can extend the model by introducing the slack variables to relax the constraints of the problem. In this case the parameter C gives the trade-off between the flatness of f and the tolerance to deviations larger than ε.

Doing this transformation we prefer to solve the relative dual problem, which is a convex quadratic programming problem.

- What is the difference in purpose form e-SV linear regression and the non linear one?

In order to generate a non linear regression function f we use the kernel to map in the feature space. The regression is not linear in the input space but it is in the feature space.

- Explain the steps for Nonlinear e-SV regression.

To solve we derive the dual problem form the primal one:

$$\begin{cases} \max_{\lambda^+,\lambda^-} -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\lambda_i^+ - \lambda_i^-)(\lambda_j^+ - \lambda_j^-)k(x_i,x_j) - \varepsilon\sum_{i=1}^{l}(\lambda_i^+ - \lambda_i^-) + \sum_{i=1}^{l} y_i(\lambda_i^+ - \lambda_i^-) \\ \sum_{i=1}^{l}(\lambda_i^+ - \lambda_i^-) = 0 \\ \lambda_i^+, \lambda_i^- \in [0,C] \end{cases}$$

To solve a non-linear ε-SV regression we start by choosing a kernel function, then we solve the dual problem to find (λ^+, λ^-), with that we can obtain $b = y_i \pm \varepsilon - \sum_{j=1}^{l}(\lambda_j^+ - \lambda_j^-)k(x_i, x_j)$ for some $0 < \lambda^\pm < C$, and finally we can calculate the recession functions: $f(x) = \sum_{i=1}^{l}(\lambda_i^+ - \lambda_i^-)k(x_i, x) + b$.

The recession function is linear in the feature space and non-linear in the input space.

15 Risk attitudes and utility functions Decision trees

- What are certainty and uncertainty, what is risk, and what are the lotteries?

"Certainty" is when all potential consequences of a given choice are clearly understood, allowing for precise and reliable decision-making. "Uncertainty" is when outcomes of various option are unknown, making it difficult to make a definitive choice. In uncertainty is not possible to make an informed decision.

If there is some information available about the probability of different outcomes, we call it "risk". In this case, while the outcome cannot be known with absolute certainty, it is possible to estimate the likelihood of each possible outcome and make a decision based on that estimation.

In decision theory, we talk about "lotteries" by identifying uncertain outcomes and representing them with probability theory.

- Which are the axioms in decision theory?

The axioms are:

- Orderability: for every pair of primitive outcomes or $A \succ B$, or $B \succ A$, or $A = B$.
- Transitivity: if $A \succ B$ and $B \succ C$, then $A \succ C$.
- Continuity: if $A \succ B$ and $B \succ C$, then with some probability the two lotteries, one with choice A and C and the other with only B, are equivalent preferable.
- Substitutability: if $A \succ B$ then given two lotteries that are exactly the same, except that one has A in a particular position and the other has B, the one with A is better.
- Monotonicity: if $A \succ B$ a lottery that gives A over B with higher probability is better.
- Decomposbility: a two stage lottery, where in the first stage you get A with probability p, and in the second you get B with probability q and C with probability $(1-q)$; is equivalent to a single stage lottery with three possible outcomes: A with probability p, B with probability $(1-p)q$ and C with probability $(1-p)(1-q)$.

- What is a Utility Function?

If preferences satisfy the six axioms then exist an utility function U such that:

- if $A \succ B$ then $U(A) > U(B)$
- if A and B are equivalent then $U(A) = U(B)$

Utility of a lottery is equal to expected utility of the outcomes.

- Explain the three different type of risk attitude, and how they modify the utility function.
We can have dirrerent risk attitude:

- Risk Adverse: the shape of the interpolation curve between utilities is concave, meaning that people often prefer a certain small amount of money rather than a higher expected value.
- Risk Neutrality: it have a linear utility function, the expected utility value is proportional to the value that we could obtain.
- Risk Seeking: the utility function is convex, some people prefer to risk higher expected value rather then sure smaller amount of money.

- What is the purpose of a decision tree?
The analysis of complex decision involve significant uncertainty. Collecting additional in formation can be crucial tin reducing the level of uncertainty. The decision taken can be heavily influenced by manager's attitude toward risk.

For structuring and analyzing managerial decision problems in the face of uncertainty we can use decision analysis methods, in particular the model called decision tree.

- How is made a decision tree?
Decision tree is made of several key elements:

- Decision node: identify possible alternatives
- Chance node: provide possible development of outcomes
- End point: identify the problem completed
- Time: is the dimension on which the tree is developed

In order to decide which alternative to select we need a decision criteria. Expected value (EV) is often a good measure of the value of an alternative since over the long run is the average amount that you expect to make form selecting the alternative.

- What is Certainty Equivalent, and which is its relationship with Expected Value?
Using expected value as a decision criteria is reasonable as long as the stakes involved are small enough to rely on the long -term averages. We can use instead the "Certainty Equivalent" (or Selling Price), which is the the guaranteed amount that you would be equally happy to receive instead of a risky alternative.

Depending on the certainty equivalent compared to the expected value, it is possible to categories the risk attitude:

- Risk Averse: If your Certainty Equivalent $<$ Expected Value
- Risk Neutral: If your Certainty Equivalent $=$ Expected Value
- Risk Seeking: If your Certainty Equivalent $>$ Expected Value

It is possible to transforms the utility function in the exponential utility function, as $U(x) = 1 - e^{-x/R}$ where $R > 0$ is the "risk tolerance", and as it become larger the utility function display less risk aversion. For each utility function there is a specific "certainty equivalent", for the exponential one it is: $CE = -R \cdot ln(1 - E[U])$.

16 The value of information

- Explain the concept of perfect information.

The concept of the "value of perfect information" revolves around the idea that if we have all the necessary information to make a decision, we can eliminate any uncertainty surrounding the potential outcomes of the available alternatives. It follows that no amount of information that is less than perfect can be more valuable than perfect information itself. The value of perfect information serves a s the maximum value that any source of information can provide.