Preliminaries on convex sets and convex functions

Convex Sets:

- Subspace: A set C is a subspace if it contains all the linear combinations of any two points in C.
- Affine Combination: An affine combination of two vectors x and y is a point $\alpha x + \beta y$, where $\alpha + \beta = 1$.
- Affine Set: A set C is an affine set if it contains all the affine combinations of any two points in C.
- Convex Combination: A convex combination of two vectors x and y is a point $\alpha x + \beta y$, where $\alpha + \beta = 1$ and $0 \leq \alpha, \beta \leq 1$.
- Convex Set: A set C is a convex set if it contains all the convex combinations of any two points in C.
- Convex Hull: The convex hull of a set C is the smallest convex set that contains all the convex combinations of points in C.

Operations Preserving Convexity:
- Sum, difference, and intersection of convex sets result in convex sets.
- Union of convex sets does not necessarily preserve convexity.
- Interior and closure of a convex set are convex sets.

Affine Functions:
- An affine function is the sum of a linear combination and a constant.
- An affine function applied to each element of a convex set results in another convex set.

Cones:
- A cone is a set that contains all vectors scaled by non-negative factors.
- A polyhedral cone is a cone that passes through the origin and can be represented as a system of linear inequalities.

Convex Functions:

- A function f is convex if the graph of f lies below the line segment connecting any two points in its domain.
- Strict convexity and strong convexity are stricter conditions than convexity.

First Order Conditions:
- For a convex function, the function value at any point is greater than or equal to the first-order Taylor approximation of the function at that point.

Second Order Conditions:
- For a convex function, the Hessian matrix is positive semi-definite.
- For a strictly convex function, the Hessian matrix is positive definite.

Existence of Optimal Solutions

Convexity is an important property as it ensures the existence of global optima. However, for non-convex problems, the existence of optimal solutions may not be guaranteed.

Existence of Global Optima for Quadratic Programming Problems

This subsection focuses on quadratic programming problems, which are a specific type of optimization problem with a quadratic objective function and linear constraints. It explains that for convex quadratic programming problems, global optima always exist.

First Order Optimality Condition

Introduces the first-order optimality condition, which is a necessary condition for a point to be optimal. It explains that in unconstrained problems, the first-order optimality condition is satisfied when the gradient of the objective function is zero. In constrained problems, it is satisfied when the gradient of the Lagrangian function is zero. The first-order optimality condition is a necessary condition for a point to be a local minimum or maximum of the objective function in unconstrained optimization problems. It is based on the gradient (or subgradient) of the objective function.

Second Order Optimality Condition

This section introduces the second-order optimality condition, which is a stronger condition than the first-order optimality condition. It states that for a point to be optimal, the Hessian matrix of the objective function or the Hessian matrix of the Lagrangian function must be positive definite (or positive semi definite in some cases).

Lagrangian Duality

Provides a method to convert constrained optimization problems into unconstrained ones. It involves introducing Lagrange multipliers, which are variables that represent the constraints of the original problem. The Lagrangian function is formed by combining the objective function with the constraints, and the duality problem is derived by optimizing the Lagrangian function with respect to both the original variables and the Lagrange multipliers. The duality problem provides a lower bound on the optimal value of the original problem, and under certain conditions, strong duality holds, meaning that the optimal values of the primal and dual problems are equal.

Support Vector Machines (SVM)

Linear SVM

Is a supervised learning algorithm used for binary classification. It aims to find the hyperplane that separates the data points of different classes with the maximum margin. The linear SVM formulation involves finding the optimal hyperplane by solving a convex optimization problem. It seeks to minimize the classification errors while maximizing the margin between the hyperplane and the closest data points.

SVM with Soft Margin

In cases where the data is not linearly separable, the soft margin SVM allows for some miss classification errors by introducing slack variables. The objective function is modified to minimize the classification errors and the slack variables, while still aiming to maximize the margin and achieve a trade-off between margin maximization and error minimization.

Non-Linear SVM

Non-linear SVM extends the SVM approach to handle data that is not linearly separable in the original feature space. It employs the kernel trick, which maps the data into a higher-dimensional feature space, where linear separation becomes possible. The optimization problem for non-linear SVM involves solving for the support vectors and the kernel parameters to find the hyperplane that maximizes the margin.

Regression Problems

Polynomial Regression

It is a type of regression analysis where the relationship between the independent variable(s) and the dependent variable is modeled as an nth-degree polynomial. The polynomial regression problem involves finding the coefficients of the polynomial that best fits the data. The objective is typically to minimize the sum of squared errors between the predicted values and the actual values.

Polynomial Regression with $\| \cdot \|2$

Refers to using the Euclidean norm as a regularization term to prevent overfitting in polynomial regression. By adding a regularization term to the objective function, the model complexity is controlled, and the coefficients are penalized for having larger magnitudes.

Polynomial Regression with $\|\cdot\|_1$

Utilizes the L1 norm as a regularization term in polynomial regression. The L1 norm encourages sparsity in the coefficients, promoting feature selection and producing a simpler model by driving some coefficients to zero.

Polynomial Regression with $\|\cdot\|_\infty$

Incorporates the L∞ norm as a regularization term. The L∞ norm promotes robustness to outliers by restricting the maximum absolute value of the coefficients. It results in a more robust model that is less influenced by extreme data points.

ε-SV Regression

Is a variation of support vector regression (SVR) that allows for a tolerance ε around the predicted values. The objective of ε-SV regression is to find a function that fits the training data within the tolerance ε while maintaining the flatness of the function outside the tolerance region. It involves solving a constrained optimization problem to find the support vectors and determine the function that best fits the data.

Non-linear ε-SV Regression

Similar to non-linear SVM, non-linear ε-SV regression utilizes the kernel trick to handle non-linear relationships between the independent and dependent variables. The data is mapped into a higher-dimensional feature space using a kernel function, and the optimization problem is solved to find the support vectors and determine the non-linear function that satisfies the ε tolerance.

Clustering Problems

Optimization Model using the $\|\cdot\|_2$ as Distance

The objective is to minimize the sum of squared distances between data points and their assigned cluster centroids. Euclidean Distance model is an optimization model for clustering problems using the Euclidean distance ($\|\cdot\|_2$) as the measure of dissimilarity between data points.

k-means Algorithm

The k-means algorithm is a popular clustering algorithm that aims to partition data into k clusters. It starts by randomly initializing k cluster centroids and iteratively assigns data points to the nearest centroid, and then updates the centroids based on the newly assigned points. This process continues until convergence, minimizing the within-cluster sum of squared distances.

Optimization Model using the $\|\cdot\|_1$ as Distance

It is an optimization model for clustering problems using the Manhattan distance ($\|\cdot\|_1$) as the distance metric. The objective is to minimize the sum of absolute distances between data points and their assigned cluster centroids.

k-median Algorithm

The k-median algorithm is another clustering algorithm that aims to partition data into k clusters. Instead of using the mean of the data points as the centroid, it uses the median. The algorithm iteratively assigns data points to the nearest median and updates the medians based on the newly assigned points. The objective is to minimize the sum of distances between data points and their assigned cluster medians.

Solution Methods for Unconstrained Optimization Problems

Gradient Method

It is an iterative optimization algorithm that utilizes the gradient of the objective function to find the optimal solution. It starts with an initial point and updates the solution iteratively in the direction of the negative gradient, taking steps proportional to the learning rate. This process continues until convergence to a local minimum.

Conjugate Gradient Method

It is an iterative optimization algorithm suitable for quadratic objective functions. It combines the gradient information and the conjugate directions to efficiently search for the minimum. It iteratively updates the solution by finding the optimal step length along the conjugate directions.

Newton Methods

This method are iterative optimization algorithms that utilize both the gradient and the Hessian matrix of the objective function. They provide faster convergence rates compared to gradient-based methods. The Newton method with line search finds the optimal step length using line search techniques. The quasi-Newton method approximates the Hessian matrix to avoid the computational burden of computing the exact Hessian.

Solution Methods for Constrained Optimization Problems

Active-Set Method

The active-set method is an iterative optimization algorithm for constrained optimization problems. It starts with an initial feasible solution and iteratively updates the solution by solving a subproblem involving the active constraints. The active set represents the subset of constraints that are active at the solution.

Penalty Method

The penalty method is an approach to handle constraints in optimization problems by penalizing the objective function for violating the constraints. It introduces a penalty term that increases as the constraints are violated, and the objective is modified to minimize the penalized function. The exact penalty method uses an exact penalty function, whereas the barrier method approximates the penalty function using barrier functions.

Barrier Method

The barrier method is a specific type of penalty method that replaces the constraints with a sequence of barrier functions. These barrier functions approach infinity as the solution approaches the boundaries of the feasible region, effectively enforcing the constraints. The optimization problem is solved iteratively by minimizing the penalized objective function.

Multi-objective optimization problems

Involve optimizing multiple conflicting objectives simultaneously. The goal is to find a set of solutions that represents a trade-off between the different objectives, as there is typically no single optimal solution.

Existence and Optimality Conditions

Existence results in multi-objective optimization problems focus on establishing the existence of feasible solutions that satisfy certain constraints. Various mathematical conditions and assumptions may be used to prove the existence of feasible solutions in different problem settings.

Optimality Conditions

It aim to define the notion of optimality for the problem. Different optimality concepts are used, such as Pareto optimality and epsilon-efficiency. Pareto optimality states that a solution is optimal if there is no other solution that improves at least one objective without deteriorating any other objective. Epsilon-efficiency relaxes the strict criteria of Pareto optimality by allowing for a small degree of objective improvement.

First-Order Optimality Conditions for Unconstrained Problems

First-order optimality conditions provide necessary conditions for optimality in unconstrained multi-objective optimization problems. These conditions involve the gradient or sub gradient vectors of the objective functions and define critical points where no descent direction exists.

First-Order Optimality Conditions for Constrained Problems

First-order optimality conditions for constrained multi-objective optimization problems incorporate the constraints into the optimality conditions. These conditions involve both the gradient or sub gradient vectors of the objective functions and the constraint functions, taking into account feasible directions satisfying the constraints.

Scalarization Method

This method converts a multi-objective optimization problem into a single-objective optimization problem by combining the multiple objectives into a scalar function. This scalar function represents a weighted sum or a combination of the objectives. By solving the scalarized problem, different trade-offs between the objectives can be explored.

Goal Method

The goal method is an approach for solving multi-objective optimization problems by defining a set of goals or reference points in the objective space. The optimization problem then seeks to find solutions that achieve or approximate these goals. The goal method provides a flexible way to handle multi-objective problems and allows decision-makers to specify their desired outcomes.

Non-cooperative Game Theory

Matrix Games

Matrix games, also known as two-player zero-sum games, involve two players with conflicting interests. The game is represented by a matrix, where each player chooses a strategy, and their payoffs depend on the chosen strategies. Mixed strategies refer to when players use probability distributions to select their strategies, rather than deterministically choosing a single strategy.

Bimatrix Games

Bimatrix games generalize matrix games to include more than two players. In bimatrix games, each player has a set of strategies, and the payoffs are determined by the combination of strategies chosen by all players. Finding the equilibrium in bimatrix games is more complex compared to matrix games.

Convex Games

Convex games are a class of games where the payoff functions are convex with respect to the players' strategies. These games have desirable properties, such as the existence of Nash equilibria. Merit functions are used to measure the quality of a strategy profile and provide information on the optimality of a given strategy.