Convex Optimization Problems

SIE 449/549: Optimization for Machine Learning

Afrooz Jalilzadeh

Department of Systems and Industrial Engineering
University of Arizona

Convex Problem

- A **convex optimization problem** (or just a convex problem) is a problem consisting of minimizing a convex function over a convex set:

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & x \in C \end{aligned}$$

  - C is a convex set and f is a convex function over C

- A **functional form** of a convex problem can written as

$$\begin{aligned} \min_x \quad & f(x) \\ \text{s.t.} \quad & g_i(x) \leq 0, \qquad i = 1, \ldots, m \\ & h_j(x) = 0, \qquad j = 1, \ldots, p \end{aligned}$$

  - $f, g_1, \ldots, g_m : \mathbb{R}^n \to \mathbb{R}$ are convex functions
  - $h_1, h_2, \ldots, h_p : \mathbb{R}^m \to \mathbb{R}$ are affine functions

Convex Problem

- Convex problems are easy:
  - If x^* be a local minimum of f over C, then x^* is a global min
  - Set of optimal solutions of the problem

    $$\min\{f(x) : x \in C\}$$

    is convex. If, in addition, f is strictly convex over C, then there exists at most one optimal solution of the problem

Projection

set (C)	$P_C(x)$	Assumptions
$\mathbb{R}^n_+$	$\max(x, 0)$	–
$\ell \le x \le u$	$\min\{\max\{x, \ell\}, u\}$	$\ell \le u$
$\mathrm{B}(0, r)$	if $\|x\| > r$, then $x = \frac{x}{\|x\|} r$	$r > 0$
$\{x \mid Ax = b\}$	$x - A^T(AA^T)^{-1}(Ax - b)$	$A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$, A full row rank
$\{x \mid a^T x \le b\}$	$x - \frac{max\{a^T x - b, 0\}}{\|a\|^2} a$	$0 \neq a \in \mathbb{R}^n, b \in \mathbb{R}$

Table 1: Orthogonal Projections

Linear Programming

- Constraints and objective function are linear/affine and hence convex:

$$\begin{aligned} \min_{x} \quad & c^T x \\ s.t. \quad & Ax \leq b \\ & Bx = d \end{aligned} \tag{LP}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $B \in \mathbb{R}^{p \times n}$, $d \in \mathbb{R}^d$ and $c \in \mathbb{R}^n$

Convex Quadratic Problems

- Minimizing a convex quadratic function subject to affine constraints:

$$\begin{aligned} \min_{x} \quad & x^T Q x + 2b^T x \\ s.t. \quad & Ax \leq c \end{aligned} \tag{QP}$$

where $Q \in \mathbb{R}^{n\times n}$ is positive semidefinite, $b \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$ and $c \in \mathbb{R}^m$

Remark 1

If function f is twice differentiable such that $\|\nabla^2 f(x)\| \leq L$ for all x and some $L > 0$, then f has a Lipschitz continuous gradient with constant L.

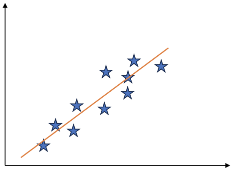
Quadratically Constrained Quadratic Problems (QCQP)

- Minimizing a convex quadratic function subject to quadratic constraints:

$$\begin{aligned} \min_{x} \quad & x^T Q x + 2b^T x + c \\ s.t. \quad & x^T A_i x + 2q_i^T x + r_i \le 0, \quad i = 1, 2, \ldots, m \end{aligned} \tag{QCQP}$$

where all matrices are positive semidefinite $Q \succeq 0$ and $A_i \succeq 0$ for all $i = 1, 2, \ldots, m$

Data Fitting – Linar Fitting



- Data: (s_i, t_i), $i = 1, \dots, m$, where $s_i \in \mathbb{R}^n$ and $t_i \in \mathbb{R}$:

$$t_i = s_i^T a + b$$

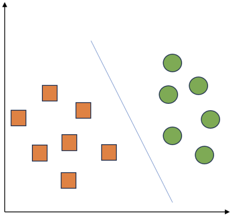
- Define $S = \begin{bmatrix} s_1^T & 1 \\ s_2^T & 1 \\ \vdots & \\ s_m^T & 1 \end{bmatrix}$, $t = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_m \end{bmatrix}$, and $x = \begin{bmatrix} a \\ b \end{bmatrix}$, hence the least squares problem is

$$\min_{x \in \mathbb{R}^{n+1}} \frac{1}{2} \|Sx - t\|^2$$

where $a = x(1 : n)$ and $b = x(n + 1)$.

Linear Classification

- Suppose that we are given two types of points in $\mathbb{R}^n$: type A and type B points
- $x_1, \ldots, x_m \in \mathbb{R}^n$ - Type A
- $x_{m+1}, \ldots, x_{m+p} \in \mathbb{R}^n$ - Type B



- Objective is to find a linear separator, which is a hyperplane of the form

$$H(w, \beta) = \{x \in \mathbb{R}^n : w^T x + \beta = 0\}$$

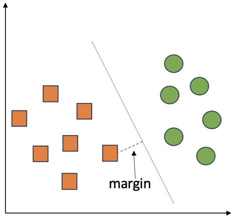
for which the type A and type B points are in its opposite sides:

$$\begin{aligned} w^T x_i + \beta < 0, &\quad i = 1, 2, \ldots, m, \\ w^T x_i + \beta > 0, &\quad i = m+1, m+2, \ldots, m+p \end{aligned}$$

- **Underlying Assumption**: the two sets of points are linearly separable, meaning that the set of inequalities has a solution

Linear Classification

- The margin of the separator is the distance of the hyperplane to the closest point
- Goal: Finding the separator with the largest margin
- We need a formula for the distance between a point and a hyperplane



Lemma 1

Let $H(a,b) = \{x \in \mathbb{R}^n : a^T x + b = 0\}$, where $\mathbf{0} \neq a \in \mathbb{R}^n$ and $b \in \mathbb{R}$. Let $y \in \mathbb{R}^n$, then the distance between y and the set H is given by

$$d(y, H(a,b)) = \frac{|a^T y + b|}{\|a\|}.$$

- The margin corresponding to a hyperplane H is

$$\min_{i=1,2,\ldots,m+p} \frac{|w^T x_i + \beta|}{\|w\|}$$

Linear Classification

- So far, the problem can be formulated as

$$\max_{w,\beta} \left\{ \min_{i=1,2,\ldots,m+p} \frac{|w^T x_i + \beta|}{\|w\|} \right\}$$
$$s.t. \quad w^T x_i + \beta < 0, \quad i = 1, 2, \ldots, m,$$
$$w^T x_i + \beta > 0, \quad i = m+1, m+2, \ldots, m+p$$

Challenge: This formulation is NOT convex and cannot be easily handled

- If (w, β) is an optimal solution, then $(\alpha w, \alpha\beta)$ for any $\alpha \neq 0$ is an optimal solution
- We can therefore decide that

$$\min_{i=1,2,\ldots,m+p} |w^T x_i + \beta| = 1$$

so, the problem can be rewritten as:

Linear Classification/ Support Vector Machine (SVM)

- Combination of the first equality and other inequalities implies that:

$$\begin{aligned}
\min_{w,\beta} \quad & \frac{1}{2}\|w\|^2 \\
s.t. \quad & w^T x_i + \beta \leq -1, \quad i = 1, 2, \ldots, m, \\
& w^T x_i + \beta \geq 1, \quad i = m+1, m+2, \ldots, m+p
\end{aligned}$$

- Define $y \in \mathbb{R}^{m+p}$: $y_i = -1$ for $i = 1, \ldots, m$ and $y_i = 1$ for $i = m+1, \ldots, m+p$

$$\begin{aligned}
\min_{w,\beta} \quad & \frac{1}{2}\|w\|^2 \\
s.t. \quad & (w^T x_i + \beta) y_i \geq 1, \quad i = 1, 2, \ldots, m+p
\end{aligned}$$

- Computing projection onto the constraints is hard!

Dual SVM

- Let $\alpha \in \mathbb{R}^{m+p}$ be the dual variable, then the dual SVM is:

$$\begin{aligned}
\max_{\alpha} \quad & \sum_{i=1}^{m+p} \alpha_i - \frac{1}{2} \sum_{i,j} y_i y_j \alpha_i \alpha_j (x_i x_j^T) \\
s.t. \quad & \sum_{i=1}^{m+p} \alpha_i y_i = 0, \\
& \alpha \geq 0.
\end{aligned}$$

- In dual formulation we will solve for α directly
- w and β can be computed from α, if needed

$$w = \sum_i \alpha_i y_i x_i$$

$$\beta = y_k - w^T x_k, \text{ for any } k \text{ where } \alpha_k > 0$$

Dual SVM

- Define $X = \begin{bmatrix} x_1^T \\ \vdots \\ x_{m+p}^T \end{bmatrix}$, $Y = diag(y)$ and $Q = YXX^TY$, then we have

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^{m+p} \alpha_i - \frac{1}{2} \sum_{i,j} y_i y_j \alpha_i \alpha_j (x_i x_j^T) \\ s.t. \quad & \sum_{i=1}^{m+p} \alpha_i y_i = 0, \\ & \alpha \geq 0. \end{aligned}$$

$$\begin{aligned} \max_{\alpha} \quad & e^T \alpha - \frac{1}{2} \alpha^T Q \alpha \\ s.t. \quad & \sum_{i=1}^{m+p} \alpha_i y_i = 0, \\ & \alpha \geq 0. \end{aligned}$$

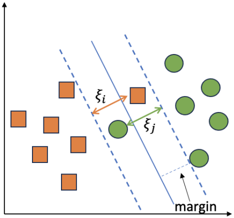
Soft Margin SVM

- What if the two sets of points are not linearly separable?

- We allow "error" ξ_i in classification

$$\min_{w,\beta,\xi_i} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m+p} \xi_i$$

$$\begin{aligned} s.t. \quad (w^T x_i + \beta)y_i &\geq 1 - \xi_i, & i = 1, 2, \ldots, m+p \\ \xi_i &\geq 0, & i = 1, 2, \ldots, m+p \end{aligned}$$



- $\xi_i = 0$ if there is no error

- C : tradeoff parameter between error and margin

Dual Soft Margin SVM

- Define $X = \begin{bmatrix} x_1^T \\ \vdots \\ x_{m+p}^T \end{bmatrix}$, $Y = diag(y)$ and $Q = YXX^TY$, then we have

$$\begin{aligned} \max_{\alpha} \quad & e^T\alpha - \frac{1}{2}\alpha^T Q\alpha \\ s.t. \quad & \sum_{i=1}^{m+p} \alpha_i y_i = 0, \\ & 0 \le \alpha_i \le C, \ \forall i \end{aligned}$$

$$w = \sum_i \alpha_i y_i x_i$$

$$\beta = y_k - w^T x_k, \text{ for any } k \text{ where } 0 < \alpha_k < C$$

What changed?

- Added upper bound of C on α_i
- Intuitive explanation: $\alpha_i \to \infty$ when constraints are violated (points misclassified)
- Upper bound of C limits the α_i, so misclassifications are allowed