3.1 Convex Optimization Problems

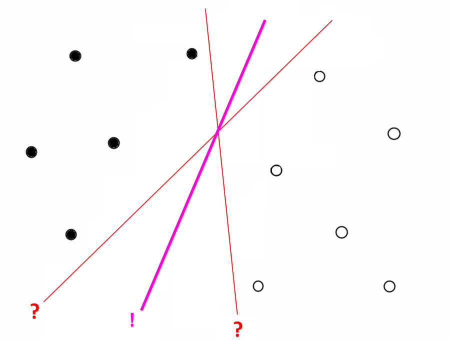
Machine Learning 1: Foundations

Marius Kloft *(TUK)*

Recap

Linear classifiers

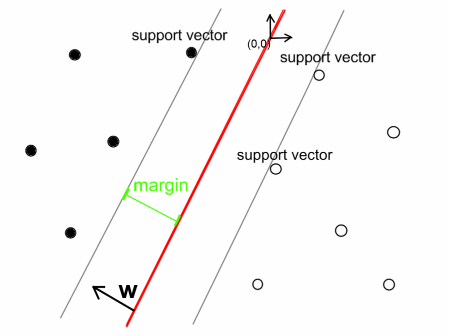
- Most important linear classifier: **support vector machine**
  - Which hyperplane to take?



Recap

Linear classifiers

- Most important linear classifier: **support vector machine**
  - Which hyperplane to take?
  - The one that separates the data with the largest margin γ.

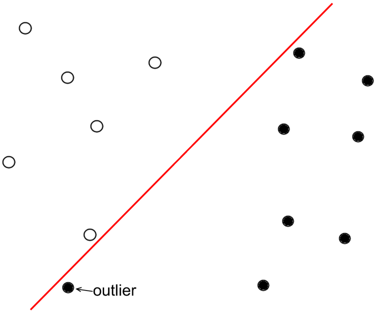


-

$$\max_{\gamma, b \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^d \setminus \{0\}} \quad \gamma \qquad \text{s.t.} \quad y_i \underbrace{\frac{\mathbf{w}^\top \mathbf{x}_i + b}{\|\mathbf{w}\|}}_{= d(x_i, H)} \geq \gamma \quad \forall i = 1, \ldots, n.$$

Recap (2)

Hard-Margin SVM not robust to outliers:



Remedy: **Linear Soft-Margin SVM**

$$\max_{\gamma, b\in\mathbb{R}, \mathbf{w}\in\mathbb{R}^d\setminus\{0\}, \boldsymbol{\xi}\in\mathbb{R}^n} \gamma - C\sum_{i=1}^{n}\xi_i \quad \text{s.t.} \quad y_i\frac{\mathbf{w}^\top\mathbf{x}_i + b}{\|w\|} \geq \gamma - \xi_i, \;\; \xi_i \geq 0$$

How to solve?

Contents of This Class

Convex Optimization Problems

1. Convex Optimization Problems (OPs)
2. SVM is a Convex OP
3. How to Solve Convex OPs and SVM

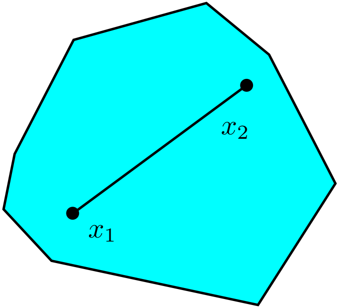
1. Convex Optimization Problems (OPs)
2. SVM is a Convex OP
3. How to Solve Convex OPs and SVM

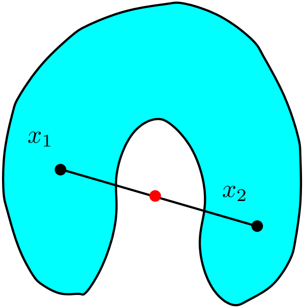
Convex sets

Definition

A set $\mathcal{X} \subset \mathbb{R}^d$ is called **convex** if and only if the line segment connecting any two points in $\mathcal{X}$ entirely lies within $\mathcal{X}$, that is,

$$\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}, \ \forall \theta \in [0,1]: \quad (1-\theta)\mathbf{x}_1 + \theta \mathbf{x}_2 \in \mathcal{X}.$$



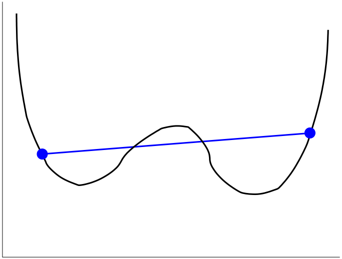


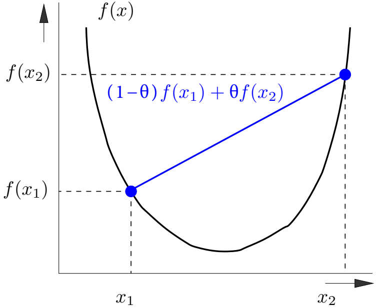
Convex Functions

Definition

A function $f : \mathbb{R}^d \to \mathbb{R}$ is **convex** if and only if the the set above the graph is convex, that is, $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^d, \ \forall \theta \in [0,1]$:

$$f((1-\theta)\mathbf{x}_1 + \theta\mathbf{x}_2) \leq (1-\theta)f(\mathbf{x}_1) + \theta f(\mathbf{x}_2).$$





Definition

A function f is **concave** if and only if $-f$ is convex.

How can I check that f is convex?

Is the easiest for (twice) **differentiable** functions f:

Theorem (second-order condition)

f is convex if and only if the Hessian matrix $H_f(\mathbf{x})$ is positive semi-definite for all $\mathbf{x} \in \mathbb{R}^d$.

Recall from matrix algebra:

Lemma

The following statements regarding a symmetric matrix $M \in \mathbb{R}^{d \times d}$ are equivalent:

- M is positive semi-definite (we write $M \succeq 0$)
- $\mathbf{x}^\top M \mathbf{x} \geq 0$ for all $x \in \mathbb{R}^d$
- All eigenvalues of M are non-negative
- All principal minors of M are positive.

We use the notation $A \succeq 0$ to denote a positive semi-definite matrix.

Examples of convex and non-convex functions

Example:

- $f : \mathbb{R} \to \mathbb{R}$ $\quad\quad$ $g : \mathbb{R}^2 \to \mathbb{R}$
  $x \mapsto 2x - 3x^2 + 7$ $\quad\quad$ $(x, y) \mapsto x^2 + y^2$

$f'(x) = 2 - 6x$
$f''(x) = -6 < 0. \implies f$ is not convex.

$$\nabla g(x) = \begin{pmatrix} 2x \\ 2y \end{pmatrix} \implies H = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \succcurlyeq 0.$$

Because the determinant of all the minors are positive.
$\det(2) = 2, \det(H) = 4.$
$\implies g$ is convex.

Convex Optimization Problems

Let $f_0, f_1, \ldots, f_n, g_1, \ldots, g_m : \mathcal{X} \to \mathbb{R}$.

Definition

1. An **optimization problem (OP) is**:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^d} \quad & f_0(\mathbf{x}) \\ \text{s.t.} \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, n \\ & g_j(\mathbf{x}) = 0, \quad j = 1, \ldots, m \end{aligned}$$

2. An OP is called **convex** if and only if the functions $f_0, f_1, \ldots, f_n$ are *convex* and $g_1, \ldots, g_m$ are *linear*.

- "s.t." means "subject to".
- $f_0 : \mathbb{R}^d \to \mathbb{R}$ is the **objective function**
- $f_i(\mathbf{x}) \leq 0$, $i = 1, \ldots, n$ are the **inequality constraints**
- $g_j(\mathbf{x}) = 0$, $j = 1, \ldots, m$ are the **equality constraints**

Example

Proposition

The linear hard-margin SVM,

$$\max_{\gamma, b\in\mathbb{R}, \mathbf{w}\in\mathbb{R}^d\setminus\{0\}} \gamma \quad \text{s.t.} \quad y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq \|\mathbf{w}\|\,\gamma,$$

is not convex.

Proof.

It suffices to show that the constraint function

$$f(\gamma, w, b) = \|\mathbf{w}\|\,\gamma - y_i(\mathbf{w}^\top \mathbf{x}_i + b)$$

is not convex. This is easiest to see for $d = 1$ and $w > 0$.

Then $\quad \|w\|\,\gamma = |w|\,\gamma = w\gamma \quad$ and $\quad H_f = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

Since $\det\left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right) = -1$, the Hessian H_f is not positive semi-definite, so f is not convex. □

Refs I

S. P. Boyd and L. Vandenberghe, Convex optimization. New York: Cambridge Univ. Press, 2004.