TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

3.1 Convex Optimization Problems

Machine Learning 1: Foundations

Marius Kloft *(TUK)*

Recap

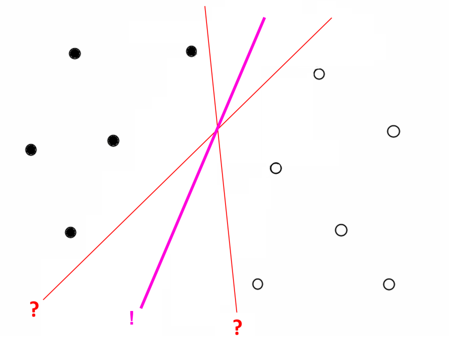
Linear classifiers

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

Recap

Linear classifiers

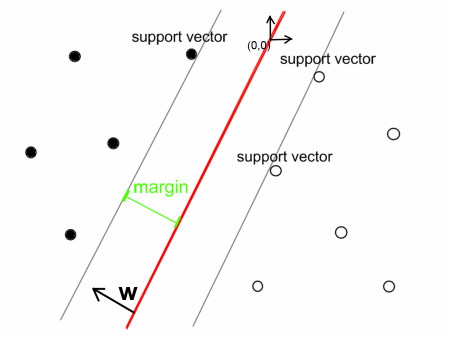
- Most important linear classifier: **support vector machine**
  - Which hyperplane to take?



Recap

Linear classifiers

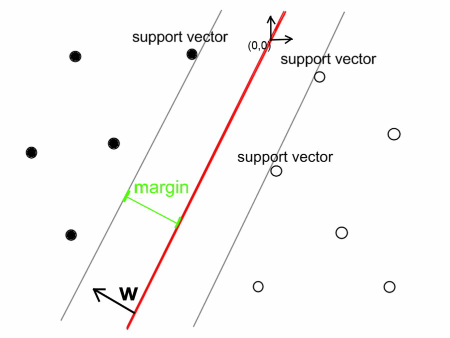
- Most important linear classifier: **support vector machine**
  - Which hyperplane to take?
  - The one that separates the data with the largest margin γ.



Recap

Linear classifiers

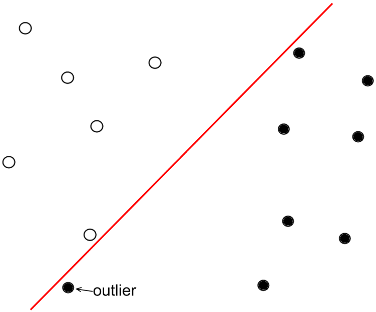
- Most important linear classifier: **support vector machine**
  - Which hyperplane to take?
  - The one that separates the data with the largest margin γ.



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

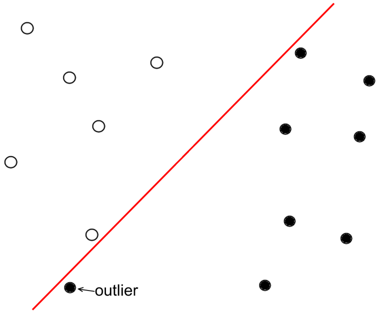
Recap (2)

Hard-Margin SVM not robust to outliers:



Recap (2)

Hard-Margin SVM not robust to outliers:

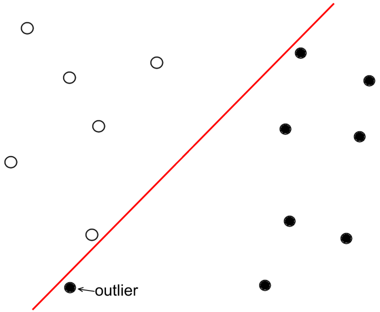


Remedy: **Linear Soft-Margin SVM**

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

Recap (2)

Hard-Margin SVM not robust to outliers:



Remedy: **Linear Soft-Margin SVM**

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

How to solve?

Contents of This Video

Convex Optimization ~~Problems~~ of SVM

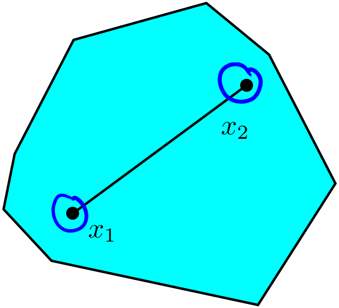
1. Convex Optimization Problems (OPs)
2. SVM is a Convex OP
3. How to Solve Convex OPs and SVM

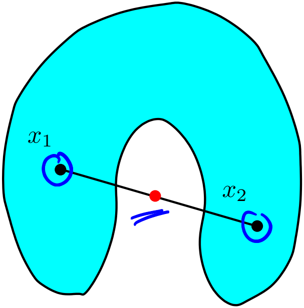
1. Convex Optimization Problems (OPs)
2. SVM is a Convex OP
3. How to Solve Convex OPs and SVM

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 $\mathcal{X}$





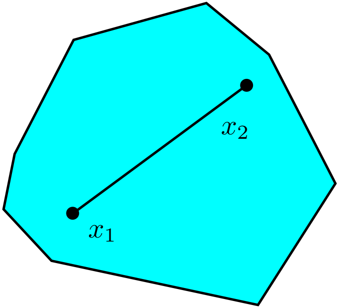
Convex sets

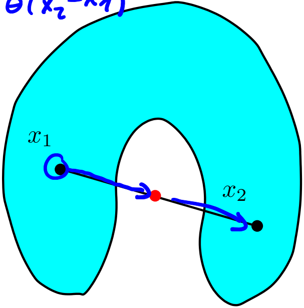
Definition

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,

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

$= x_1 + \theta(x_2 - x_1)$

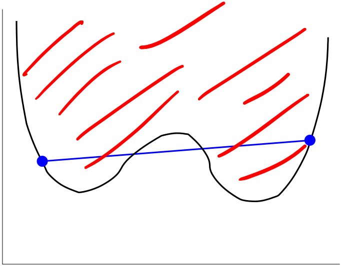




Convex Functions

Definition

A function $f : \mathbb{R}^d \to \mathbb{R}$ is **convex** if and only if the the set above the graph is convex

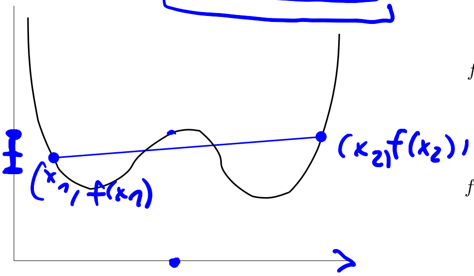


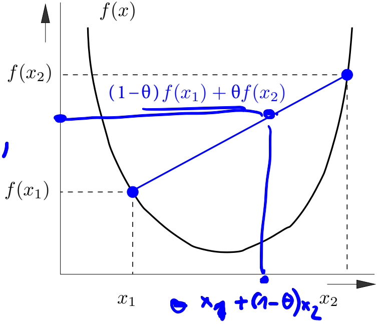
Convex Functions

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, $\forall \mathbf{x}_1, \mathbf{x}_2 \in \cancel{\mathcal{X}}$ $\mathbb{R}^d$, $\forall \theta \in [0, 1]$:

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



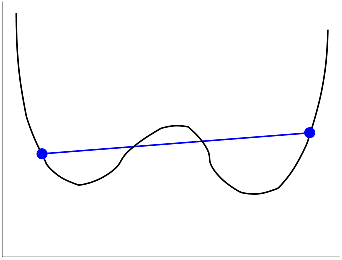


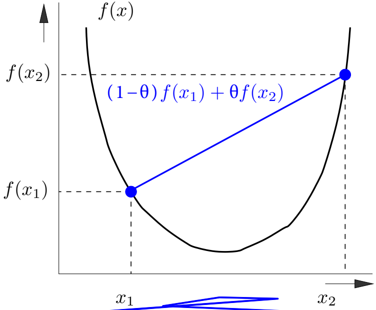
Convex Functions

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, $\forall \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}, \ \forall \theta \in [0, 1]$:

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





Definition

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

How can I check that f is convex?

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

How can I check that f is convex?

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

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

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 $\mathbf{x} \in \mathbb{R}^d$.

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

$f'(x) = -(2 - 6x)$

$f''(x) = +6 > 0$

$\Rightarrow -f$ is concave

Ex: $f: \mathbb{R}^2 \to \mathbb{R}$, $(x,y) \mapsto x^2 + y^2$

$\nabla f(x,y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$

$H_f(x,y) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \succeq 0$

$\det(2) = 2 > 0$

$\det \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = 4 > 0$

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

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**:

objective

inequality constraints

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

$x \in \mathbb{R}^d$

equality constraints

subject to

Ex.

$$
\begin{aligned}
\min_{x \in \mathbb{R}} \quad & x^2 - 2x + 7 \\
\text{s.t.} \quad & x \geq 0
\end{aligned}
$$

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**:

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

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*.

Example

Recall from last week:

Linear Hard-margin SVM~~Linear Hard-margin SVM~~

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

Example

Recall from last week:

Linear Hard-margin SVM Linear Hard-margin SVM

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

Is this SVM a convex OP?

to show: $f : (w, \gamma) \mapsto \|w\| \gamma - y_i (w^T x_i + b)$

Assume $w \in \mathbb{R}_+$:

$$D f(w, \gamma) = \begin{pmatrix} \gamma - y_i x_i \\ w \end{pmatrix}$$

$$H_f(w, \gamma) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\det(H_f(w, \gamma)) = 0 - 1 = -1 < 0$$

$\Rightarrow$ f is not convex Δ

Example

Recall from last week:

Linear Hard-margin SVMLinear Hard-margin SVM

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

Is this SVM a convex OP?

Recall:

1. Standard form of an OP is: $\begin{array}{ll} \min_{\mathbf{x}\in\mathbb{R}^d} & f_0(\mathbf{x}) \\ \text{s.t.} & f_i(\mathbf{x}) \leq 0, \quad i = 1, \ldots, n \end{array}$
2. Thm: f is convex if the second derivative is positive.

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

Refs I