3.2 SVM is a Convex OP

Machine Learning 1: Foundations

Marius Kloft *(TUK)*

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

Making the SVM Convex

Theorem

The linear hard-margin SVM from last week, that is,

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

can be equivalently rewritten in convex form as given below:

Linear hard-margin SVM in convex form

$$\min_{b \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^d} \quad \frac{1}{2} \|\mathbf{w}\|^2$$

convex, quadr.

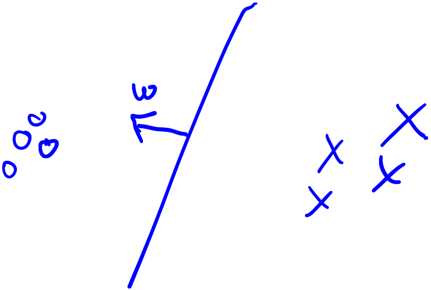
$$\text{s.t.} \qquad 1 - y_i(\mathbf{w}^\top \mathbf{x}_i + b) \leq 0 \quad \forall i = 1, \ldots, n$$

lin: ineqn

Core Idea of the Proof

The SVM results in a linear classifier $f(\mathbf{x}) = \mathrm{sign}(\mathbf{w}^\top \mathbf{x} + b)$, the parameters $(\mathbf{w}, b)$ of which are not unique:

$$\forall \lambda > 0 : \quad \mathrm{sign}(\mathbf{w}^\top \mathbf{x} + b) = \mathrm{sign}\big(\underbrace{\lambda \mathbf{w}}_{=:\mathbf{w}_\lambda}{}^\top \mathbf{x} + \underbrace{\lambda b}_{=:b_\lambda}\big). \quad (\star)$$



Core Idea of the Proof

The SVM results in a linear classifier $f(\mathbf{x}) = \text{sign}(\mathbf{w}^\top \mathbf{x} + b)$, the parameters $(\mathbf{w}, b)$ of which are not unique:

$$\forall \lambda > 0 : \quad \text{sign}(\mathbf{w}^\top \mathbf{x} + b) = \text{sign}\big(\underbrace{\lambda \mathbf{w}}_{=:\mathbf{w}_\lambda}{}^\top \mathbf{x} + \underbrace{\lambda b}_{=:b_\lambda}\big). \quad (\star)$$

Idea: if it helps, we could restrict in the SVM OP our search for **w** to **w**s that have some nice norm, and we would—by $(\star)$—still search the space of all linear classifiers.

Core Idea of the Proof

The SVM results in a linear classifier $f(\mathbf{x}) = \text{sign}(\mathbf{w}^\top \mathbf{x} + b)$, the parameters $(\mathbf{w}, b)$ of which are not unique:

$$\forall \lambda > 0: \quad \text{sign}(\mathbf{w}^\top \mathbf{x} + b) = \text{sign}\big(\underbrace{\lambda \mathbf{w}}_{=:\mathbf{w}_\lambda}{}^\top \mathbf{x} + \underbrace{\lambda b}_{=:b_\lambda}\big). \quad (\star)$$

Idea: if it helps, we could restrict in the SVM OP our search for **w** to **w**s that have some nice norm, and we would—by $(\star)$—still search the space of all linear classifiers.

For reasons that will become clear below, we choose the restriction $\|\mathbf{w}\| = 1/\gamma$, by setting $\lambda := \frac{1}{\|\mathbf{w}\|\gamma}$.

Proof (*) $\|\tilde{w}\| = \left\|\frac{w}{\|w\|\gamma}\right\| = \frac{\|w\|}{\|w\|\cdot\gamma} = \frac{1}{\gamma}$ (1/2)

$$\max_{\gamma, w, b} \gamma \quad \text{s.t.} \quad y_i(w^T x_i + b) \geq \|w\| \cdot \gamma$$

$$0 \quad 0 \quad 0 \quad 0 \qquad\qquad 0 + 0 \geq 0 \cdot 0 \checkmark$$

$$\Leftrightarrow \min_{\gamma \geq 0, w, b} \frac{1}{\gamma} \quad \text{s.t.} \quad y_i(w^T x_i + b) \geq \|w\| \cdot \gamma \quad | \cdot \lambda$$

$\forall \lambda > 0$

$$\Leftrightarrow \min_{\gamma > 0, w, b} \frac{1}{2\gamma^2} \quad \text{s.t.} \quad y_i(\lambda w^T x_i + \underbrace{\lambda b}_{=:\tilde{b}}) \geq \lambda \|w\| \cdot \gamma$$

$$\Leftrightarrow \min_{\gamma > 0, \tilde{b}, w} \frac{1}{2\gamma^2} \quad \text{s.t.} \quad y_i(\lambda w x_i + \tilde{b}) \geq \lambda \|w\| \gamma$$

$\lambda = \frac{1}{\|w\| \cdot \gamma}$

$$\Leftrightarrow \min_{\gamma > 0, \tilde{b}, w} \frac{1}{2}\|\tilde{w}\|^2 \quad \text{s.t.} \quad y_i\left(\underbrace{\frac{w^T}{\|w\|\gamma}}_{=:\tilde{w}} x_i + \tilde{b}\right) \geq 1$$

$$\Leftrightarrow \min_{\gamma > 0, \tilde{b}, \tilde{w}} \frac{1}{2}\|\tilde{w}\|^2 \quad \text{s.t.} \quad y_i(\tilde{w}^T x_i + b) \geq 1 \quad \square$$

Proof (2/2)

Soft-margin SVM

Similar, we can formulate a soft-margin SVM as convex optimization problem:

Linear soft-margin SVM in convex form

$$
\begin{aligned}
\min_{b\in\mathbb{R},\mathbf{w}\in\mathbb{R}^d,\boldsymbol{\xi}\in\mathbb{R}^n} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i \\
\text{s.t.} \quad & 1 - \xi_i - y_i(\mathbf{w}^\top \mathbf{x}_i + b) \leq 0, \quad -\xi_i \leq 0 \quad \forall i = 1, \ldots n
\end{aligned}
$$

Soft-margin SVM

Similar, we can formulate a soft-margin SVM as convex optimization problem:

Linear soft-margin SVM in convex form

$$\min_{b\in\mathbb{R},\mathbf{w}\in\mathbb{R}^d,\boldsymbol{\xi}\in\mathbb{R}^n} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i$$

$$\text{s.t.} \quad 1-\xi_i - y_i(\mathbf{w}^\top\mathbf{x}_i + b) \leq 0, \quad -\xi_i \leq 0 \quad \forall i = 1,\dots n$$

How to solve?

How to Solve SVM?

How to Solve SVM?

We could just put our (convex) SVM OP into one of the many solvers out there for convex optimization problems.

How to Solve SVM?

We could just put our (convex) SVM OP into one of the many solvers out there for convex optimization problems.

- In Python: library CVXOPT is a standard

How to Solve SVM?

We could just put our (convex) SVM OP into one of the many solvers out there for convex optimization problems.

- In Python: library CVXOPT is a standard

Problem: CVXOPT is very powerful, but rather slow
$\Rightarrow$ For big data we need a **faster** solution