Quiz: SVMs

Lecture series „Machine Learning“

Niels Landwehr

Research Group „Data Science“
Institute of Computer Science
University of Hildesheim

Quiz: Primal Hard-Margin SVM

- Assume a two-class classification problem with one-dimensional input space $\mathcal{X} = \mathbb{R}$ and classes {-1,1}.
- Let $\mathcal{D} = \{(x_1, y_1), (x_2, y_2)\}$ with $(x_1, y_1) = (0, -1)$ and $(x_2, y_2) = (1, 1)$ denote the training data set
- We train a primal hard-margin SVM of the form $f_\theta(x) = \theta x + b$
- **Question**: What is the resulting solution for θ and b ?
  - $\theta = 0.5,\ b = 1$
  - $\theta = 4,\ b = -2$
  - $\theta = 1,\ b = -1$
  - $\theta = 2,\ b = -1$
  - $\theta = -1,\ b = 0$

Solution: Primal Hard-Margin SVM

- **Solution**: the optimization problem is

$$\begin{aligned}&\text{Find } \boldsymbol{\theta}, b\\&\text{such that } y_n(\mathbf{x}_n{}^{\mathrm{T}}\boldsymbol{\theta}+b)\geq 1 \quad \text{for } n\in\{1,\ldots,N\}\\&\text{and } \frac{1}{2}\|\boldsymbol{\theta}\|^2 \text{ minimal}\end{aligned}$$

- Write down the constraints and plug in the values for the data points:

$$y_1(x_1\theta+b)\geq 1 \quad\Rightarrow\quad (-1)\cdot(0\cdot\theta+b)\geq 1 \quad\Rightarrow\quad -b\geq 1$$

$$y_2(x_2\theta+b)\geq 1 \quad\Rightarrow\quad 1\cdot(1\cdot\theta+b)\geq 1 \quad\Rightarrow\quad \theta+b\geq 1$$

- Adding up the two final inqualities implies $\theta\geq 2$
- We see that for $\theta=2,\ b=-1$ the constraints are satisfied. This is also the solution with minimum $\|\theta\|$ (because of the condition $\theta\geq 2$)

Quiz: Dual Hard-Margin SVM

- Assume still the same problem setting and the same data set
- We now train a dual hard-margin SVM
- **Question**: what is the resulting dual parameter vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$?
  - $\alpha_1 = 1,\ \alpha_2 = 0$
  - $\alpha_1 = 2,\ \alpha_2 = -2$
  - $\alpha_1 = 2,\ \alpha_2 = 2$
  - $\alpha_1 = -1,\ \alpha_2 = 1$
  - $\alpha_1 = 1,\ \alpha_2 = 1$

Solution: Dual Hard-Margin SVM

- **Solution**: the dual optimization problem is

$$\boldsymbol{\alpha}^* = \arg\max_{\boldsymbol{\alpha}} \tilde{L}(\boldsymbol{\alpha})$$

$$\text{subject to } \sum_{n=1}^{N} \alpha_n y_n = 0 \text{ and } \alpha_n \geq 0 \text{ for all n} \in \{1,\ldots,N\}$$

$$\text{where } \tilde{L}(\boldsymbol{\alpha}) = -\frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n\alpha_m y_n y_m \mathbf{x}_n^{\mathrm{T}}\mathbf{x}_m + \sum_{n=1}^{N}\alpha_n$$

$$\Rightarrow \quad \alpha_1 y_1 + \alpha_2 y_2 = 0$$

$$\Rightarrow \quad -\alpha_1 + \alpha_2 = 0$$

$$\Rightarrow \quad \alpha_1 = \alpha_2$$

- We also know from the lecture the primal solution can be recovered by $\boldsymbol{\theta}^* = \sum_{n=1}^{N} \alpha_n^* y_n \mathbf{x}_n$

$$\Rightarrow \quad \theta = \alpha_1 y_1 x_1 + \alpha_2 y_2 x_2$$

$$\Rightarrow \quad 2 = \alpha_1 \cdot (-1) \cdot 0 + \alpha_2 \cdot 1 \cdot 1$$

$$\Rightarrow \quad \alpha_2 = 2$$

- So the correct solution is $\alpha_1 = 2,\ \alpha_2 = 2$

Quiz: Primal Soft-Margin SVM

- Now assume any binary classification problem with instance space $\mathcal{X} = \mathbb{R}^M$
- When training a primal soft-margin SVM, we solve the optimization problem

maximize margin

minimize margin violations

$$\begin{aligned}
&\text{Find } \boldsymbol{\theta}, b, \boldsymbol{\xi} \\
&\text{such that } \frac{1}{2}\|\boldsymbol{\theta}\|^2 + \gamma \sum_{n=1}^{N} \xi_n \text{ minimal} \\
&\text{with } y_n(\mathbf{x}_n{}^{\mathrm{T}}\boldsymbol{\theta} + b) \geq 1 - \xi_n \text{ and } \xi_n \geq 0
\end{aligned}$$

- **Question**: Given a data set $\mathcal{D}$, if we train a model with small γ and a model with larger γ, which of the two models will be more susceptible to overfitting?
  - The model with larger γ
  - The model with smaller γ
  - They are equally susceptible to overfitting
  - Cannot say anything based on the information provided

Solution: Primal Soft-Margin SVM

- **Solution**: the model with larger γ will be more susceptible to overfitting

- The easiest way to see this is by reformulating the problem as regularized loss minimization (Slide 21 in lecture):

$$\arg\min_{\boldsymbol{\theta},b} \frac{1}{2} \|\boldsymbol{\theta}\|^2 + \gamma \sum_{n=1}^{N} \max(0, 1 - y_n f_{\boldsymbol{\theta}}(\mathbf{x}_n))$$

$$= \arg\min_{\boldsymbol{\theta},b} \frac{1}{N} \sum_{n=1}^{N} \ell(f_{\boldsymbol{\theta}}(\mathbf{x}_n), y_n) + \underbrace{\frac{1}{2N\gamma}}_{\text{regularization weight}} \|\boldsymbol{\theta}\|^2$$

$$\ell(f_{\boldsymbol{\theta}}(\mathbf{x}_n), y_n) = \max(0, 1 - y_n f_{\boldsymbol{\theta}}(\mathbf{x}_n))$$

- The larger the γ, the smaller the regularization weight