Machine Learning: Exercises for Block III
Kernel Methods
– partial solutions –

ML course
winter semester 2021
University of Saarland
Prof. Isabel Valera

Exercise 2

We want to show: $k(x_i, x_j) = \langle x_i, x_j \rangle$ is a positive definite (PD) kernel.

Such that by definition of PD: $\sum_{i,j} c_i c_j \langle x_i, x_j \rangle \geq 0$

Let $x \in \mathcal{H}$ st. $x = \sum_i c_i x_i$, $c_i \in \mathbb{R}$. Note that x can be zero.

$$
\begin{aligned}
\sum_{i,j} c_i c_j \langle x_i, x_j \rangle &= \sum_i \sum_j c_i c_j \langle x_i, x_j \rangle \\
&= \sum_i \sum_j \langle c_i x_i, c_j x_j \rangle \\
&= \left\langle \sum_i c_i x_i, \sum_j c_j x_j \right\rangle \\
&= \langle x, x \rangle \\
&\geq 0
\end{aligned}
$$

by linearity $\alpha \langle x_i, x_j \rangle = \langle \alpha x_i, x_j \rangle$ and conjugate symmetry with $c_i \in \mathbb{R}$

by linearity $\langle x+y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$ and conjugate symmetry with $c_i \in \mathbb{R}$

as defined above

by definition of dot product

Exercise 5

- There are ℓ groups, each having n_i samples with $i \in [\ell]$, where $[\ell] = 1, 2, \dots, \ell$.
- The penalty of a group equals the slack of the worst point in that group. We thus have one slack variable per group : $z = \{z_i\}_{i=1}^{\ell}$.
- The primal problem is given by:

$$\min_{\substack{w \in \mathbb{R}^d, b \in \mathbb{R} \\ z_i \in \mathbb{R}^n}} \frac{1}{2}\|w\|^2 + \sum_i c_i z_i$$

$$\text{s.t.} \quad y_i^j(\langle w, x_i^j \rangle + b) \geq 1 - z_i,$$

$$z_i \geq 0 \qquad \forall i \in [\ell], j \in [n_i]$$

where each group i is assigned a budget c_i.

Lagrangian

with Lagrange multipliers $\alpha = \{\alpha_i^j \mid i \in [d], j \in [n_i]\}$, $\beta = \{\beta_i\}_{i=1}^{n_i}$

$$L(w, b, z, \alpha, \beta) = \frac{1}{2}\|w\|^2 + \sum_i^{\ell} c_i z_i + \sum_i^{\ell}\sum_j^{n_i} \alpha_i^j \left[1 - z_i - y_i^j(\langle w, x_i^j \rangle + b)\right] - \sum_i^{\ell} \beta_i z_i$$

where $\alpha_i^j \geq 0$, $\beta_i \geq 0$, $\forall i \in [\ell], j \in [n_i]$.

Stationary point w.r.t. parameters

$$\frac{\partial L}{\partial w} = w - \sum_i^{l} \sum_j^{n_i} \alpha_i^j y_i^j x_i^j \overset{!}{=} 0$$

$$\Rightarrow \quad \boxed{w = \sum_i^{l} \sum_j^{n_i} \alpha_i^j y_i^j x_i^j} \qquad (1)$$

$$\frac{\partial L}{\partial b} = \boxed{\sum_i^{l} \sum_j^{n_i} \alpha_i^j y_i^j \overset{!}{=} 0} \qquad (2)$$

$$\frac{\partial L}{\partial z_i} = \boxed{C_i - \sum_j^{n_i} \alpha_i^j - \beta_i \overset{!}{=} 0} \qquad (3)$$

$$\Rightarrow \quad \beta_i = C_i - \sum_j^{n_i} \alpha_i^j \overset{!}{\geq} 0$$

$$\Rightarrow \quad \boxed{C_i \geq \sum_j^{n_i} \alpha_i^j} \qquad (4)$$

- Like in the original soft-margin problem the primal problem is convex in w, b
- Slaters condition holds $\Rightarrow$ strong duality
  $\Rightarrow$ complementary slackness

$$\alpha_i^j (1 - z_i - y_i^j (\langle w, x_i^j \rangle + b)) = 0 \text{ and}$$
$$\beta_i z_i = 0 \qquad \forall i \in [l],\ j \in [n_i]$$

Dual problem

- with $\|w\|^2 = w^T w$ (1)

$$= \left(\sum_i^{\ell}\sum_j^{n_i} \alpha_i^j y_i^j x_i^j\right)^T \left(\sum_i^{\ell}\sum_j^{n_i} \alpha_i^j y_i^j x_i^j\right)$$

$$= \sum_i^{\ell}\sum_{i'}^{\ell}\sum_j^{n_i}\sum_{j'}^{n_{i'}} \alpha_i^j \alpha_{i'}^{j'} y_i^j y_{i'}^{j'} \langle x_i^j, x_{i'}^{j'} \rangle$$

- and $\sum_i^{\ell}\sum_j^{n_i} \alpha_i^j y_i^j \langle w, x_i^j \rangle = w^T \sum_i^{\ell}\sum_j^{n_i} \alpha_i^j y_i^j x_i^j = \|w\|^2$ (5)

- We replace w, b, β in the Lagrangian and formulate the dual problem using (1) – (6)

$$L(w,b,z,\alpha,\beta) = \frac{1}{2}\|w\|^2 + \sum_i^{\ell} C_i z_i$$

$$+ \sum_i^{\ell}\sum_j^{n_i} \alpha_i^j \left[1 - z_i - y_i^j(\langle w, x_i^j\rangle + b)\right]$$

$$- \sum_i^{\ell} \beta_i z_i$$

$$= \frac{1}{2}\|w\|^2 - \sum_i^{\ell}\sum_j^{n_i} \alpha_i y_i^j \langle w, x_i^j \rangle - \sum_i^{\ell}\sum_j^{n_i} \alpha_i y_i^j b$$

$$+ \sum_i^{\ell}\sum_j^{n_i} \alpha_i^j + \sum_i^{\ell} z_i\left(C_i - \sum_j \alpha_i^j - \beta\right)$$

(5)

$$= -\frac{1}{2}\|w\|^2 - \sum_i^{\ell}\sum_j^{n_i} \underbrace{\cancel{\alpha_i y_i^j}}_{=0} b \quad (2)$$

$$+ \sum_i^{\ell}\sum_j^{n_i} \alpha_i^j + \sum_i^{\ell} z_i \underbrace{\cancel{\left(C_i - \sum_j^{n_i} \alpha_i^j - \beta\right)}}_{=0} \quad (3)$$

(1)

$$\mathcal{L}(\alpha) = -\frac{1}{2}\left(\sum_i^{\ell}\sum_{i'}^{\ell}\sum_j^{n_i}\sum_{j'}^{n_{i'}} \alpha_i^j \alpha_{i'}^{j'} y_i^j y_{i'}^{j'} \langle x_i^j, x_{i'}^{j'}\rangle\right) + \sum_i^{\ell}\sum_j^{n_i} \alpha_i^j$$

constraints: $\alpha_i^j \geq 0$, $\underbrace{\sum_i^{\ell}\sum_j^{n_i} \alpha_i^j y_i^j = 0}_{(2)}$, $\underbrace{\sum_j^{n_i} \alpha_i^j \leq C_i}_{(4)}$ $\forall i \in [d], j \in [n_i]$

Exercise 8

$$L(w,b,z,\hat{z},\alpha,\hat{\alpha}) = C\sum_{i=1}^{n}\left(z_i+\hat{z}_i\right) + \frac{1}{2}\|w\|^2 - \sum_{i=1}\left(\beta_i z_i + \hat{\beta}_i\hat{z}_i\right)$$

$$- \sum_{i=1}^{n}\alpha_i\left(\varepsilon + z_i + y(x_i) - y_i\right) - \sum_{i=1}^{n}\hat{\alpha}_i\left(\varepsilon+\hat{z}_i - y(x_i) + y_i\right)$$

with $z_i \geq 0$, $z=\{z_i\}_{i=1}^{n}$, $\hat{z}=\{\hat{z}_i\}_{i=1}^{n}$, $\alpha=\{\alpha_i\}_{i=1}^{n}$, $\hat{\alpha}=\{\hat{\alpha}_i\}_{i=1}^{n}$.

Taking derivatives (recall $y(x_i) = w^T\phi(x_i)+b$).

$$\frac{\partial L}{\partial w} = w - \sum_i \alpha_i\phi(x_i) + \sum_i\hat{\alpha}_i\phi(x_i) = 0$$

$$\Rightarrow \quad \boxed{w = \sum_i \phi(x_i)\left(\alpha_i-\hat{\alpha}_i\right)} \quad 1$$

$$\frac{\partial L}{\partial b} = \boxed{-\sum_i\alpha_i + \sum_i\hat{\alpha}_i = 0} \quad 2$$

$$\frac{\partial L}{\partial z_i} = \boxed{C-\beta_i-\alpha_i = 0} \quad 3$$

$$\frac{\partial L}{\partial \tilde{z}_i} = \boxed{C-\hat{\beta}_i-\hat{\alpha}_i = 0} \quad 4$$

Note: 5

$$w^T\phi(x_i)(\alpha_i-\hat{\alpha}_i)$$
$$= w^Tw = \|w\|^2$$

$$L(w,b,z,\hat{z},\alpha,\hat{\alpha}) = C\sum_{i=1}^{n}\left(z_i+\hat{z}_i\right) + \frac{1}{2}\|w\|^2 - \sum_{i=1}\left(\beta_i z_i + \hat{\beta}_i\hat{z}_i\right)$$

$$- \sum_{i=1}^{n}\alpha_i\left(\varepsilon + z_i + y(x_i) - y_i\right) - \sum_{i=1}^{n}\hat{\alpha}_i\left(\varepsilon+\hat{z}_i - y(x_i) + y_i\right)$$

$$= \frac{1}{2}\|w\|^2 + \sum_i \underbrace{\left(C-\cancel{\beta_i-\alpha_i}\right)}_{=0 \; 3} + \hat{z}_i\underbrace{\left(C-\cancel{\hat{\beta}_i-\hat{\alpha}_i}\right)}_{=0 \; 4}$$

$$5 \qquad - \sum_i \varepsilon(\alpha_i+\hat{\alpha}_i) + y_i(\hat{\alpha}_i-\alpha_i) + w^T\phi(x_i)(\alpha_i-\hat{\alpha}_i) + b\underbrace{\left(\cancel{\alpha_i-\hat{\alpha}_i}\right)}_{=0 \; 2}$$

$$= -\frac{1}{2}\|w\|^2 - \sum_i \varepsilon(\alpha_i+\hat{\alpha}_i) + y_i(\hat{\alpha}_i-\alpha_i)$$

$$= -\frac{1}{2}\left(\sum_i\phi(x_i)(\alpha_i-\hat{\alpha}_i)\right)^T\left(\sum_i\phi(x_i)(\alpha_i-\hat{\alpha}_i)\right)$$

$$- \sum_i \varepsilon(\alpha_i+\hat{\alpha}_i) + y_i(\hat{\alpha}_i-\alpha_i)$$

$$= -\frac{1}{2}\sum_i\sum_j(\alpha_i-\hat{\alpha}_i)(\alpha_j-\hat{\alpha}_j)\underbrace{\phi(x_i)^T\phi(x_j)}_{k(x_i,x_j)} - \sum_i\varepsilon(\alpha_i+\hat{\alpha}_i) + \sum_i y_i(\hat{\alpha}_i-\alpha_i)$$