# Kernel methods in machine learning - Penn and paper 3

Christian Segercrantz 481056

May 3, 2022

# Task 1

$$
\begin{aligned}
f(\theta x + (\theta - 1)y) \leq & \theta f(x) + (1-\theta) f(y) && (1)\\
f(z) := & ||z|| && (2)\\
f(\theta x + (\theta - 1)y) = & ||\theta x + (\theta - 1)y|| && (3)\\
\leq & ||\theta x|| + ||(\theta - 1)y||, \qquad |\text{Triangel inequality} && (4)\\
= & \theta ||x|| + (\theta - 1)||y||, \qquad |\text{Absolutely scaleable} && (5)\\
= & \theta f(x) + (1-\theta) f(y) && (6)\\
\implies f(\theta x + (\theta - 1)y) \leq & \theta f(x) + (1-\theta) f(y), \qquad f(z) := ||z|| && (7)
\end{aligned}
$$

# Task 2

$$
\begin{aligned}
\theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 = & (\theta_1 + \theta_2)\left(\frac{\theta_1}{\theta_1 + \theta_2} x_1 + \frac{\theta_2}{\theta_1 + \theta_2} x_2\right) + \theta_3 x_3 && (8)\\
= & (1 - \theta_3)\left(\frac{\theta_1}{\theta_1 + \theta_2} x_1 + \frac{\theta_2}{\theta_1 + \theta_2} x_2\right) + \theta_3 x_3, \qquad |\theta_1 + \theta_2 = 1 - \theta_3 && (9)\\
x' := & \frac{\theta_1}{\theta_1 + \theta_2} x_1 + \frac{\theta_2}{\theta_1 + \theta_2} x_2 && (10)\\
= & (1 - \theta_3) x' + \theta_3 x_3 && (11)\\
& && (12)
\end{aligned}
$$

We now show that $x'$ is part of the convex set

$$
\begin{aligned}
\theta_1' := & \frac{\theta_1}{\theta_1 + \theta_2} && (13)\\
\theta_2' := & \frac{\theta_2}{\theta_1 + \theta_2} && (14)\\
\theta_1' + \theta_2' = & 1, \quad \theta_1', \theta_2' \geq 0 && (15)\\
\implies \theta_1' x_1 + \theta_2' x_2 = & \theta_1' x_1 + (1 - \theta_1') x_2 && (16)
\end{aligned}
$$

As we can see, $x'$ is part of the covex set $C$ and thus, we can see that the points $(1-\theta_3)x' + \theta_3 x_3$ are also part of the convex set.

# Task 3

$$
\begin{aligned}
\min_{w,\xi,b} \quad & \frac{1}{2}||w||^2 + C\sum_{i=1}^{m} \xi_i && (17)\\
\text{subject to:} \quad & y_i(w^\top \phi(x_i) + b) \geq 1 - \xi_i && (18)\\
& \xi_i \geq 0, i = 1, ..., m && (19)\\
& && (20)
\end{aligned}
$$

## Task 3a

The Lagrangiang function becomes

$$L(w, \xi, b, \lambda, \nu) = \frac{1}{2}||w||^2 + C\sum_{i=1}^{m} \xi_i + \sum_{i=1}^{m} \lambda_i \left(1 - \xi_i - y_i(w^\top \phi(x_i) + b)\right) + \sum_{i=1}^{m} -\nu_i \xi_i \tag{21}$$

$$\tag{22}$$

2

## Task 3b

The KKT conditions state, for the part that is relevant for our problem, that

1. All inequality constraints $f_i(w^*, \xi^*, b^*) \leq 0, i = 1, ..., m$ which gives us

$$1 - \xi_i - (y_i(w^\top \phi(x_i) + b)) \leq 0 \tag{23}$$

$$-\xi_i \leq 0 \tag{24}$$

2. Non-negativity of dual variables of the inequality constraints $\lambda_i^* \geq 0, \nu_i^* \geq 0, i = 1, ..., m$
3. Complementary slackness $\lambda_i^* f_i^*(x^*, \xi^*, b^*) = 0, \nu_i^* f_i^*(x^*, \xi^*, b^*) = 0, i = 1, ..., m$ which gives us

$$\lambda_i(1 - \xi_i - (y_i(w^\top \phi(x_i) + b))) = 0, \quad i = 1, ..., m \tag{25}$$

$$\implies 1 - \xi_i - (y_i(w^\top \phi(x_i) + b)) = 0 \quad \vee \quad \lambda_i = 0 \tag{26}$$

$$-\nu_i \xi_i = 0, \quad i = 1, ..., m \tag{27}$$

$$\implies \nu_i = 0 \quad \vee \quad \xi_i = 0 \tag{28}$$

4. The derivative of the Lagrangian vanishes $\nabla_{w,\xi,b} L(w^*, \xi^*, b^*, \lambda^*, \nu^*) = 0$

Let us calculate the fourth point above:

$$\frac{\partial}{\partial \xi} L(w, \xi, b, \lambda, \nu) = \frac{\partial}{\partial \xi} \left( \frac{1}{2} ||w||^2 + C \sum_{i=1}^{m} \xi_i + \sum_{i=1}^{m} \lambda_i \left(1 - \xi_i - y_i(w^\top \phi(x_i) + b)\right) + \sum_{i=1}^{m} -\nu_i \xi_i \right) \tag{29}$$

$$= Cm - \sum_{i=1}^{m} (\lambda_i^* + \nu_i^*) \tag{30}$$

$$\frac{\partial}{\partial \xi} L(w, \xi, b, \lambda, \nu) = 0 \implies Cm = \sum_{i=1}^{m} \lambda_i + \nu_i \tag{31}$$

$$\frac{\partial}{\partial w} L(w, \xi, b, \lambda, \nu) = \frac{\partial}{\partial w} \left( \frac{1}{2} ||w||^2 + C \sum_{i=1}^{m} \xi_i + \sum_{i=1}^{m} \lambda_i \left(1 - \xi_i - y_i(w^\top \phi(x_i) + b)\right) + \sum_{i=1}^{m} -\nu_i \xi_i \right) \tag{32}$$

$$= w - \sum_{i=1}^{m} \lambda_i y_i \phi(x_i) \tag{33}$$

$$\frac{\partial}{\partial w} L(w, \xi, b, \lambda, \nu) = 0 \implies w^* = \sum_{i=1}^{m} \lambda_i y_i \phi(x_i) \tag{34}$$

$$\frac{\partial}{\partial b} L(w, \xi, b, \lambda, \nu) = \frac{\partial}{\partial b} \left( \frac{1}{2} ||w||^2 + C \sum_{i=1}^{m} \xi_i + \sum_{i=1}^{m} \lambda_i \left(1 - \xi_i - y_i(w^\top \phi(x_i) + b)\right) + \sum_{i=1}^{m} -\nu_i \xi_i \right) \tag{35}$$

$$= \sum_{i=1}^{m} \lambda_i y_i \tag{36}$$

$$\frac{\partial}{\partial b} L(w, \xi, b, \lambda, \nu) = 0 \wedge \lambda_i \geq 0 \implies \lambda_i^* y_i = 0 \tag{37}$$

$$\tag{38}$$

To state clearly:

$$\sum_{i=1}^{m} \lambda_i^* y_i = 0, \quad \forall i \tag{39}$$

$$w^* = \sum_{i=1}^{m} \lambda_i^* y_i \phi(x_i) \tag{40}$$

$$\lambda_i^* + \nu_i^* = C \implies 0 \leq \lambda_i^* \leq C, \quad \forall i \tag{41}$$

$$\tag{42}$$

## Task 3c

$$\max_{\lambda,\nu} \quad q(\lambda,\nu) \tag{43}$$

$$\text{subject to:} \quad \sum_{i=1}^{m} \lambda_i y_i = 0 \tag{44}$$

$$0 \leq \lambda_i^* \leq C, \forall i \tag{45}$$

$$(\lambda \geq 0) \tag{46}$$

$$\nu \geq 0 \tag{47}$$

$$\tag{48}$$

where

$$q(\lambda,\nu) = \min_{w,\xi,b} L(w,\xi,b,\lambda,\nu) \tag{49}$$

$$= \frac{1}{2}||w^*||^2 + C\sum_{i=1}^{m} \xi_i^* + \sum_{i=1}^{m} \lambda_i \left(1 - \xi_i^* - y_i({w^*}^\top \phi(x_i) + b^*)\right) + \sum_{i=1}^{m} -\nu_i \xi_i^* \tag{50}$$

$$= \frac{1}{2}||w^*||^2 + C\sum_{i=1}^{m} \xi_i^* + \sum_{i=1}^{m} (\lambda_i - \lambda_i \xi_i^* - \lambda_i y_i {w^*}^\top \phi(x_i) - \lambda_i y_i b^*) + \sum_{i=1}^{m} -\nu_i \xi_i^* \tag{51}$$

$$= \frac{1}{2}||w^*||^2 + \sum_{i=1}^{m} C\xi_i^* + \lambda_i - \lambda_i \xi_i^* - \lambda_i y_i {w^*}^\top \phi(x_i) - \lambda_i y_i b^* - \nu_i \xi_i^* \tag{52}$$

$$= \frac{1}{2}||w^*||^2 + \sum_{i=1}^{m} \xi_i^* (\underbrace{C - \lambda_i - \nu_i}_{0}) + \lambda_i - \lambda_i y_i {w^*}^\top \phi(x_i) - b^* \underbrace{\sum_{i=1}^{m} \lambda_i y_i}_{0} \tag{53}$$

$$= \frac{1}{2}||w^*||^2 + \sum_{i=1}^{m} \lambda_i - \lambda_i y_i {w^*}^\top \phi(x_i) \tag{54}$$

$$= \frac{1}{2}{w^*}^\top w^* + \sum_{i=1}^{m} \lambda_i - \lambda_i y_i {w^*}^\top \phi(x_i) \tag{55}$$

$$= \frac{1}{2}\left(\sum_{i=1}^{m} \lambda_i y_i \phi(x_i)\right)^\top \left(\sum_{j=1}^{m} \lambda_j y_j \phi(x_j)\right) + \sum_{i=1}^{m} \lambda_i - \sum_{i=1}^{m} \lambda_i y_i \left(\sum_{j=1}^{m} \lambda_j y_j \phi(x_j)\right)^\top \phi(x_i) \tag{56}$$

$$= \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i \lambda_j y_i y_j \phi(x_i)^\top \phi(x_j) + \sum_{i=1}^{m} \lambda_i - \sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i \lambda_j y_i y_j \phi(x_j)^\top \phi(x_i), \quad |\kappa(x_i,x_j) = \phi(x_i)^\top \phi(x_j) \tag{57}$$

$$= \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i \lambda_j y_i y_j \kappa(x_i,x_j) - \sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i \lambda_j y_i y_j \kappa(x_i,x_j) + \sum_{i=1}^{m} \lambda_i \tag{58}$$

$$= -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \lambda_i \lambda_j y_i y_j \kappa(x_i,x_j) + \sum_{i=1}^{m} \lambda_i \tag{59}$$

To summarize

$$\max_{\lambda} \quad -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \kappa(x_i, x_j) + \sum_{i=1}^{m}\lambda_i \tag{60}$$

$$\text{subject to:} \quad \sum_{i=1}^{m}\lambda_i y_i = 0 \tag{61}$$

$$0 \le \lambda_i \le C, \forall i \tag{62}$$

$$\tag{63}$$