# Kernel methods in machine learning - Penn and paper 2

Christian Segercrantz 481056

April 5, 2022

# Task 1

$$
\begin{aligned}
k_c(\mathbf{x}_i, \mathbf{x}_j) &= \langle \phi_c(\mathbf{x}_i), \phi_c(\mathbf{x}_j) \rangle && (1)\\
&= \left\langle \phi(\mathbf{x}_i) + \frac{1}{N}\sum_{p=1}^{N} \mathbf{x}_p, \phi(\mathbf{x}_j) + \frac{1}{N}\sum_{q=1}^{N} \mathbf{x}_q \right\rangle && (2)\\
&= \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle + \left\langle \phi(\mathbf{x}_i), \frac{1}{N}\sum_{q=1}^{N} \phi(\mathbf{x}_q) \right\rangle + \left\langle \phi(\mathbf{x}_j), \frac{1}{N}\sum_{p=1}^{N} \phi(\mathbf{x}_p) \right\rangle + \left\langle \frac{1}{N}\sum_{q=1}^{N} \phi(\mathbf{x}_q), \frac{1}{N}\sum_{p=1}^{N} \phi(\mathbf{x}_p) \right\rangle && (3)\\
&= \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle + \frac{1}{N}\sum_{q=1}^{N} \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_q) \rangle + \frac{1}{N}\sum_{p=1}^{N} \langle \phi(\mathbf{x}_j), \phi(\mathbf{x}_p) \rangle + \frac{1}{N^2}\sum_{q=1}^{N}\sum_{p=1}^{N} \langle \phi(\mathbf{x}_q), \phi(\mathbf{x}_p) \rangle && (4)\\
&= k\left(\phi(\mathbf{x}_i), \phi(\mathbf{x}_j)\right) + \frac{1}{N}\sum_{q=1}^{N} k\left(\phi(\mathbf{x}_i), \phi(\mathbf{x}_q)\right) + \frac{1}{N}\sum_{p=1}^{N} k\left(\phi(\mathbf{x}_j), \phi(\mathbf{x}_p)\right) + \frac{1}{N^2}\sum_{q=1}^{N}\sum_{p=1}^{N} k\left(\phi(\mathbf{x}_q), \phi(\mathbf{x}_p)\right) && (5)\\
& && (6)
\end{aligned}
$$

# Task 2

## 2.1

$$
\begin{aligned}
P(y = C_1 \mid X = \hat{x}) &= \frac{P(y = C_1, X = \hat{x})}{P(X = \hat{x})}, \qquad \text{|Law of conditional probability} && (7)\\
P(X = \hat{x}) &= P(X = \hat{x} \mid y = C_1)P(y = C_1) + P(X = \hat{x} \mid y = C_2)P(y = C_2), \qquad \text{|Law of total probability} && (8)\\
&= P(X = \hat{x}, y = C_1) + P(X = \hat{x}, y = C_2) && (9)\\
P(y = C_1 \mid X = \hat{x}) &= \frac{P(y = C_1, X = \hat{x})}{P(X = \hat{x}, y = C_1) + P(X = \hat{x}, y = C_2)} && (10)\\
P(y = C, X = x) &= p(y = C, X = x)dx && (11)\\
P(y = C_1 \mid X = \hat{x}) &= \frac{p(y = C_1, X = \hat{x})dx}{p(X = \hat{x}, y = C_1)dx + p(X = \hat{x}, y = C_2)dx} && (12)\\
&= \frac{p(y = C_1, X = \hat{x})}{p(X = \hat{x}, y = C_1) + p(X = \hat{x}, y = C_2)} && (13)\\
& && (14)
\end{aligned}
$$

## 2.2

We can calculate the the minimum classification error as the intersection of the density functions of the classes:

$$
\begin{aligned}
P(\text{minimum misclassification error}) &= \int_{x \in X} \min(p(x, C_1), p(x, C_2))dx && (15)\\
&\leq \int_{x \in X} (p(x, C_1)p(x, C_2))^{1/2}dx && (16)
\end{aligned}
$$

According to the given inequality $min(a, b) \leq (ab)^{1/2}$.

## Task 3

The decision function is given by a similar function to the binary classifier, but over all classes. I.e. the decision function is such a function that output's the optimal $k$ based on the probability of each possibility.

The normalization is given by the sum of all possibilities in order to make the function output a value in the range $[0, 1]$

$$\arg\max_k \frac{1}{Z} \exp \langle \mathbf{w}_k, \mathbf{x}_i \rangle, k = 1, 2, ...N \tag{17}$$

$$Z = \sum_{k=1}^{N} \exp \langle \mathbf{w}_k, \mathbf{x}_i \rangle, \tag{18}$$

$$\implies \arg\max_k \frac{\exp \langle \mathbf{w}_k, \mathbf{x}_i \rangle}{\sum_{k=1}^{N} \exp \langle \mathbf{w}_k, \mathbf{x}_i \rangle}, k = 1, 2, ...N \tag{19}$$

$$\tag{20}$$

which is the softmax function.

## Task 4

We start by calculating the expecation:

$$\mathbb{E}[e^{\lambda\epsilon}] = \sum_{\{-1,1\}} \frac{1}{2} e^{\lambda\epsilon} d\epsilon \tag{21}$$

$$= \frac{e^{-\lambda} + e^{\lambda}}{2} \tag{22}$$

and power expand the expecation to

$$e^{\lambda} = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} \tag{23}$$

$$e^{-\lambda} = \sum_{n=0}^{\infty} \frac{(-\lambda)^n}{n!} \tag{24}$$

$$\frac{e^{-\lambda} + e^{\lambda}}{2} = \frac{1}{2} \left( \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} + \sum_{n=0}^{\infty} \frac{(-\lambda)^n}{n!} \right) \tag{25}$$

$$= \sum_{n=0}^{\infty} \frac{\lambda^n + (-\lambda)^n}{2n!} \tag{26}$$

$$= \sum_{n=0}^{\infty} \frac{\lambda^{2n}}{(2n)!} \tag{27}$$

Since $\lambda^n + (-\lambda)^n = 0$ when $n$ is odd, we could rewrite the sum as as the one on the last row.

Now we power series expand the other term:

$$e^{\frac{\lambda^2}{2}} = \sum_{n=0}^{\infty} \frac{\left(\frac{\lambda^2}{2}\right)^n}{n!} = \sum_{n=0}^{\infty} \frac{\lambda^{2n}}{2^n n!} \tag{28}$$

Since 2! grows faster than $2^n$ we can conclude that $\mathbb{E}[e^{\lambda\epsilon}] \leq e^{\frac{\lambda^2}{2}}$.