Machine Learning I: Foundations
Exercise Sheet 10

Prof. Marius Kloft TA:Billy Joe Franks

15.07.2021
Deadline: 13.07.2021

This will be the last exercise sheet of Machine Learning I: Foundations.

1) (MANDATORY) 10 Points

Prove the theorem on slide 4 of the lecture 10.3 slides:

Theorem 1 *The centered kernel matrix $\tilde{K}$ can be computed from the (uncentered) kernel matrix K by:*

$$\tilde{K} = \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right) K \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}^T\right)$$

First of all from the lecture we have that

$$\tilde{K} = \tilde{\Phi}(X)^T \tilde{\Phi}(X)$$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \Phi(\mathbf{x}_i)$$

$$\tilde{\Phi}(\mathbf{x}_i) = \Phi(\mathbf{x}_i) - \hat{\mu}.$$

Notice that

$$\tilde{K} = (\Phi(X) - \hat{\mu}\mathbf{1}^T)^T (\Phi(X) - \hat{\mu}\mathbf{1}^T).$$

The $\mathbf{1}^T$, basically copies $\hat{\mu}$ into each column of the resulting matrix. Now we notice that

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} \Phi(\mathbf{x}_i) = \frac{1}{n}\Phi(X)\mathbf{1}.$$

Then we can simply follow that

$$\begin{aligned}
\tilde{K} &= \left(\Phi(X) - \hat{\mu}\mathbf{1}^T\right)^T \left(\Phi(X) - \hat{\mu}\mathbf{1}^T\right) \\
&= \left(\Phi(X)\left(I - \frac{\mathbf{1}}{n}\mathbf{1}^T\right)\right)^T \left(\Phi(X)\left(I - \frac{\mathbf{1}}{n}\mathbf{1}^T\right)\right) \\
&= \left(I - \frac{\mathbf{1}\mathbf{1}^T}{n}\right)^T \Phi(X)^T \Phi(X) \left(I - \frac{\mathbf{1}\mathbf{1}^T}{n}\right) \\
&= \left(I - \frac{\mathbf{1}\mathbf{1}^T}{n}\right)^T K \left(I - \frac{\mathbf{1}\mathbf{1}^T}{n}\right)
\end{aligned}$$

2) Prove the theorem on slide 3 of the lecture 10.3 slides:

Theorem 2 *The kPCA solution $\alpha^* = (\alpha_1^*, \ldots, \alpha_n^*)$ is given by the k largest Eigenvectors of the (centered) kernel matrix K.*

The objective function for kPCA from the slides is

$$\max \alpha^T K^2 \alpha$$
$$\text{s.t. } \alpha^T K \alpha = 1$$

while enforcing an orthogonality constraint for following eigenvectors. Since K is positive-semi-definite, $\sqrt{K}$ exists. This can be done using the diagonalization, i.e.

$$K = P\Delta P^{-1} = P\sqrt{\Delta}\sqrt{\Delta}P^{-1} = P\sqrt{\Delta}P^{-1}P\sqrt{\Delta}P^{-1} = \sqrt{K}\sqrt{K}.$$

Here $\sqrt{\Delta}$ is the element-wise application of the square root, note this is only possible because the eigenvalues are non-negative. From this it also follows, that $\sqrt{K}$ is symmetric and positive-semi-definite and has the same eigenvectors as K. Using this the optimization problem can be reformulated as

$$\max \left(\sqrt{K}\alpha\right)^T K \left(\sqrt{K}\alpha\right)$$
$$\text{s.t. } \left(\sqrt{K}\alpha\right)^T \left(\sqrt{K}\alpha\right) = 1.$$

As was shown in the slides this is achieved by

$$\sqrt{K}\alpha = \mathbf{v}$$

where $\mathbf{v}$ is the top eigenvector of K. Since $\mathbf{v}$ is an eigenvector of K, it is an eigenvector of $\sqrt{K}$ and $\sqrt{K}^{-1}$. so we have

$$\begin{aligned} \alpha &= \sqrt{K}^{-1}\mathbf{v} \\ \implies \alpha &= \frac{1}{\sqrt{\lambda_{\mathbf{v}}}}\mathbf{v}. \end{aligned}$$

Enforcing orthogonality will give us the next eigenvector and so forth.

3) Please fill out the VLU. If you do not have a @cs.uni-kl.de email adress you might not have received an invitation to fill it out. You can get access by entering a university email adress under `https://vlu.informatik.uni-kl.de/teilnahme/`. The VLU is a lecture survey, if enough of you fill this out we might be able to use it to improve ML1 in the future.

Secondly during your exam preparations it would be appreciated if you could note any mistakes you find in the slides, exercise sheets, or the lecture script. Whenever you feel like you have finished you can send this list of mistakes to b_franks12@cs.uni-kl.de

4) Solve programming task 10.