Machine Learning I: Foundations
Exercise Sheet 8

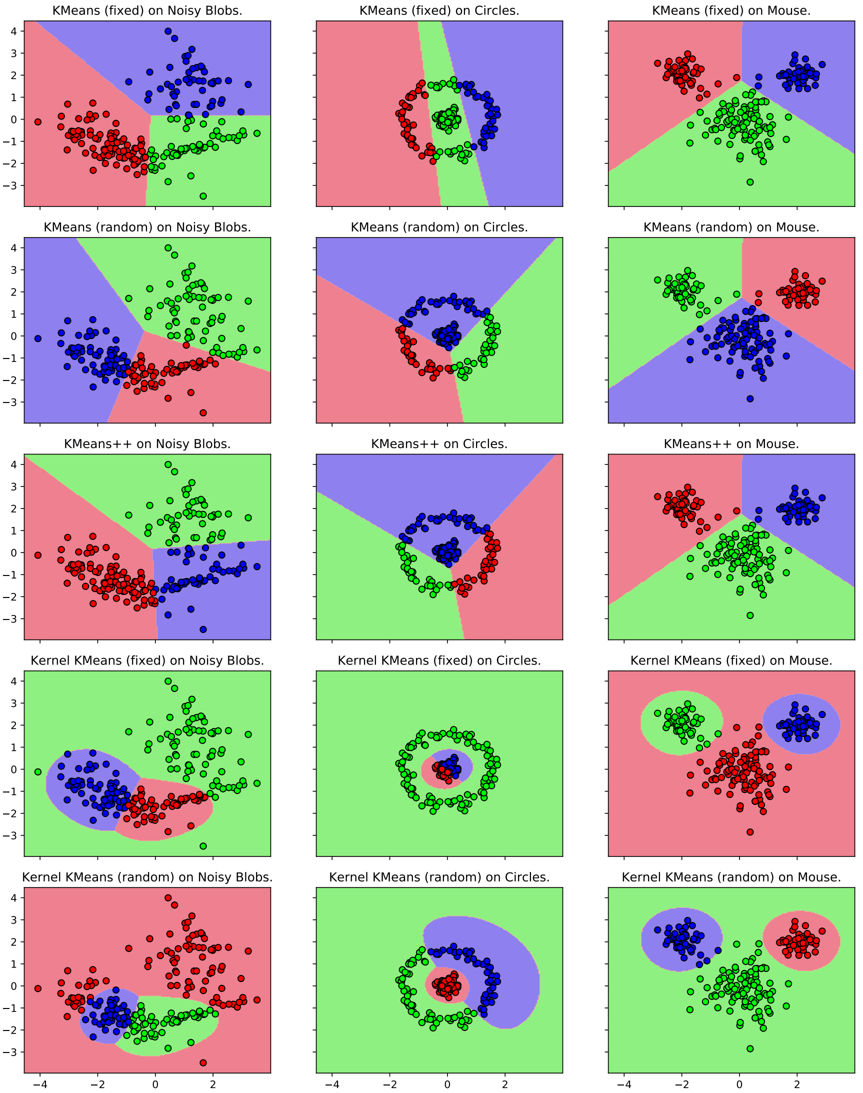
Prof. Marius Kloft TA:Billy Joe Franks

23.06.2020
Deadline: 23.06.2020

1) (MANDATORY) 10 Points

This week the mandatory exercise will be a coding exercise. As such you need to download and work on the provided Jupyter Notebook. There you have to solve Exercise 1, the exercise about k-means. At the end of this exercise you should have produced some plots, that are already present in the code skeleton. These plots are automatically saved as a pdf file named 'kmeans_plots.pdf'. You should hand in your finished .ipynb file as well as this saved pdf of the plots. You will be scored on your plots, as such do not change the code skeleton revolving around the plotting. In unclear cases we might check your code. Refer to sheet 0 if you are unsure how to set up Jupyter Notebook.

The solution to this exercise can be found in the solution jupyter notebook file.



2) I expect there to be students who are only solving the exercise sheet, but are not solving the coding exercises. Coding is an inherent part of ML1 as such I recommend you to spend the time you would have spent on this exercise on coding exercises instead, in case you have not done so anyway.

3) In the lecture the following solution to ridge regression was stated

$$\mathbf{w}_{RR} = \left(XX^\top + \frac{1}{2C}\mathrm{I}\right)^{-1} Xy.$$

The traditional linear regression has the solution $\mathbf{w}_R = (XX^\top)^{-1}Xy$. The matrix $X \in \mathbb{R}^{n\times d}$ is commonly not invertible. For example, if our problem has more features than entries the traditional linear regression is not defined since $(XX^\top)$ is singular. Ridge regression can solve this problem by adding $\frac{1}{2C}\mathrm{I}$.

a) For which values of C is $\left(XX^\top + \frac{1}{2C}\mathrm{I}\right)$ singular, thus having no solution? (Tip: consider the eingenvalues of $XX^\top$)

Note that $(XX^T)^T = (X^T)^T X^T = XX^T$. Therefore XX^T is symmetric and diagonalizable. Let $\mathbf{v}$ be an eigenvector of XX^T and λ be its respective eigenvalue. Then we have:

$$\begin{aligned}
\left(XX^T + \frac{1}{2C}\mathrm{I}\right)\mathbf{v} &= XX^T\mathbf{v} + \frac{1}{2C}\mathrm{I}\mathbf{v} \\
&= \lambda\mathbf{v} + \frac{1}{2C}\mathbf{v} \\
&= \left(\lambda + \frac{1}{2C}\right)\mathbf{v}
\end{aligned}$$

Therefore, $\mathbf{v}$ is also eigenvector of $\left(XX^T + \frac{1}{2C}\mathrm{I}\right)$ with eigenvalue $\left(\lambda + \frac{1}{2C}\right)$. A singular matrix has at least one eigenvalue equal to 0 thus:

$$\begin{aligned}
\lambda + \frac{1}{2C} &= 0 \\
2C\lambda + 1 &= 0 \\
2C\lambda &= -1 \\
C &= -\frac{1}{2\lambda}
\end{aligned}$$

b) Prove that XX^T is positive semi-definite.

Consider for any $\mathbf{v} \in \mathbb{R}^d$

$$\mathbf{v}^T XX^T \mathbf{v} = \left\|X^T\mathbf{v}\right\|^2 \geq 0$$

c) Prove that, for proper choices of C, $\left(XX^\top + \frac{1}{2C}\mathrm{I}\right)$ is always invertible.

$\boxed{\text{In part a) we found for } C = -\frac{1}{2\lambda} \text{ the matrix in question will be singular.}}$

In part a) we found for $C = -\frac{1}{2\lambda}$ the matrix in question will be singular. However proper choices for C are positive, i.e. $C > 0$. These two facts can only be rectified if $\lambda < 0$, however XX^T is PSD, so $\lambda \geq 0$, thus the matrix in question will be invertible.

4) Solve programming task 8.