Universität des Saarlandes
Prof. Dr. Dietrich Klakow
Lehrstuhl für Signalverarbeitung
NNTI Winter Term 2020/2021

# Assignment 3

Machine Learning Basics

**Deadline :01.12.2020 23:59**

## 1 Linear Regression

### 1.1 Part a (1 point)

Suppose we have a dataset that contains the financial status of startups that have recently entered the stock market with three regressors, $X_1$ = "raised funds" (in millions of dollars), $X_2$ = "initial stock value", $X_3$ = "debt" (in millions of dollars). The variable of interest Y is the company value after a year (in millions of dollars). Suppose we use least squares to fit the model :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 ln X_{i2} + \beta_3 X_{i3}$$

and get $\beta_0$ = 10, $\beta_1$ = 10, $\beta_2$ = 0.5, $\beta_3$ = -5.

Elaborate on the correctness of these statements :

- One unit change $X_1$ causes a 1000 percent change in Y .
- One unit change in $X_2$ causes a 50 percent change in Y .
- 100 percent change in $X_2$ causes a 50 percent change in Y .
- Higher debt implies lower future stock value.
- Is there any meaningful interpretation of the bias term $\beta_0$ ?

### 1.2 Part b : Ridge Regression (1.5 points)

In the slide "Fitting with Regularization" a regularization term is introduced to the least squares loss to avoid over/underfitting. This is also known as ridge regression.

$$\mathbf{J(w)} = \mathbf{MSE}_{train} + \lambda \mathbf{w}^{\mathsf{T}} \mathbf{w}$$

Here w are the weights and $\lambda$ is the regularization parameter. Provide a closed form solution for **w** which minimizes this loss.Please you should show all the steps that arrive at the solution given in the slides. If you have problems with vector and matrix differenntiation please take a look at the following slides.

## 2 Training Basics

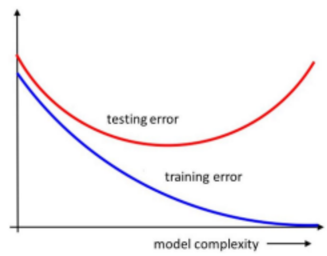


Figure 1: : Training vs test error as a function of the model complexity

**Figure 1** shows the training and test errors of a general machine learning algorithm as a function of the model complexity (richness of the hypothesis space). The model complexity is usually adjusted via a parameter, such as the regularization parameter $\lambda$ in ridge regression.

### 2.1 Part a (1 point)

Mark on Figure 1 the following :

- The ideal point to operate the machine learning algorithm.
- The range of the model complexity in which the machine learning algorithm overtrains/overfits.
- The range of the model complexity in which the machine learning algorithm undertrains/underfits.

### 2.2 Part b (0.5 points)

Consider the ridge regression problem in which the following loss function is minimized for a given $\lambda \geq 0$:

$$L(\boldsymbol{\beta}) = \sum_{i=1}^{n}(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta})^2 + \lambda\sum_{j=1}^{p}\beta_j^2$$

where n denotes the number of data instances and p denotes the dimension of the data instances as usual.

- How does the model complexity change as $\lambda$ increases?
- Given that Figure 1 represents the all possible values for the model complexity given $\lambda \; \varepsilon \; [0, \infty \,)$, mark on Figure 1 the operating point when $\lambda = 0$ and the operating point when $\lambda = \infty$.

### 2.3 Part c (0.5 points)

You are given a dataset D = ($x_i$ ,$y_i$) with i=1..n . Assume that you ran leave-one-out cross-validation on this dataset and calculated the cross-validation mean squared error (MSE) and found that it is equal to 35. Now, you randomly shuffle the dataset D and re-run leave-one-out cross-validation and calculate the new MSE. How is the new MSE related to the previous MSE which was 35? Clearly justify your reasoning.

**Note :** For solving this question it might be a good idea to read the Machine Learning Basics chapter in the Deep Learning book by Goodfellow et.al.

## 3 Maximum Likelihood

Consider independent and identically distributed (i.i.d.) random variables $X_1$,.....,$X_n$, where each $X_i$ is Poisson($\lambda$) ($\lambda$ is the parameter of the Poisson distribution). Poission distribution with parameter $\lambda$ is defined as below :

$$P\,(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

### 3.1 Part a (1 point)

Given a realization (observation) $x_1$,.......,$x_n$ of these random variables find the maximum likelihood estimate of $\lambda$.

### 3.2 Part b (1.5 points)

Assume that we only observe events $X_i = 0$ or $X_i > 0$ for each random variable. Given a realization $y_1$,..., $y_n$ of these events ($y_i = 0$ if $x_i = 0$ and $y_i = 1$ if $x_i > 0$), find the maximum likelihood estimate of $\lambda$.

## 4 Logistic Regression (1 point)

In the lecture, you were introduced to the logistic regression model. The Wikipedia page on logistic regression can be a good starting point to know more about it. Please answer in short sentences the following questions.

- When do we use logistic regression? Why can't we use linear regression in such cases?
- Is logistic regression an unsupervised learning method?
- Can we use mean squared error as an evaluation metric to compare the output of the logistic regression output with its target output? Explain why such is the case.
- How is logistic regression trained? Hint: This is a method you have exploited in this assignment. A short answer is enough.
- Why can we view the output of logistic regression as a probability?

## 5 EigenDecomposition (2 points)

**Note:** Before you solve this question as a refresher for your Linear Algebra knowledge would be better to look at Wikipedia page on eigendecomposition.

Consider the following matrix:

$$M = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

- Is the matrix M symmetric? What does it imply for its eigendecomposition? (0.5 points)
- Is the matrix M Singular? What does it imply for its eigendecomposition? (0.5 points)
- Find the eigendecomposition of M (1 point).

# 6 Submission instructions

The following instructions are mandatory. If you are not following them, tutors can decide to not correct your exercise.

- You **have to** submit a solution of this assignment sheet as a team of 2 members.
- Hand in a **single** PDF file with your solutions to the tasks.
- Therefore, make sure to write the name and matriculation ID of each of the members in your team.
- The solution must be uploaded by only **one** of your team members to the MS Teams.
- If you have any trouble with the submission, contact your tutor **before** the deadline.