University of Stuttgart | ANALYTIC COMPUTING

Foundations of Machine Learning - Exercise (SS 25)

Assignment 3: Linear regression

Arvindh Arunbabu
arvindh.arun@ki.uni-stuttgart.de

Nadeen Fathallah
Nadeen.Fathallah@ki.uni-stuttgart.de

Cosimo Gregucci
cosimo.gregucci@ki.uni-stuttgart.de

Akram Sadat Hosseini
Akram.Hosseini@ki.uni-stuttgart.de

Farane Jalali
farane.jalali-farahani@ki.uni-stuttgart.de

Osama Mohammed
osama.mohammed@ki.uni-stuttgart.de

Jiaxin Pan
jiaxin.pan@ki.uni-stuttgart.de

Tim Schneider
tim.schneider@ki.uni-stuttgart.de

Jingcheng Wu
jingcheng.wu@ki.uni-stuttgart.de

Submit your theoretical solution in ILIAS as a single PDF file.[1] Make sure to list the full names of all participants, matriculation number, study program, and B.Sc. or M.Sc on the first page. Optionally, you can *additionally* upload source files (e.g., PPTX files). Submit your programming task in ILIAS as a single Jupyter notebook. If you have any questions, feel free to ask them in the exercise forum in ILIAS.

Submission is open until Monday, 12th of May, 12:00 noon.

[1] Your drawing software probably allows exporting as PDF. An alternative option is to use a PDF printer. If you create multiple PDF files, use a merging tool (like pdfarranger) to combine the PDFs into a single file.

Task 1: Existence of Unique Solutions

Let $X\mathbf{w}$ be a linear regression model with data matrix $X \in \mathbb{R}^{n \times d}$ and coefficients $\mathbf{w} \in \mathbb{R}^d$. The unscaled covariance matrix $X^T X \in \mathbb{R}^{d \times d}$ which appears in the solution to linear regression is also called the *Normal Matrix*. We will investigate some presumeably useful properties of this matrix in the following.

1. **Task** Show that the Normal matrix is symmetric and positive semi-definite.
2. **Task** Under which requirement does the *Normal Matrix* become symmetric and positive definite? What do these requirements mean for the data in a Machine Learning task?
3. **Task** Show that any positive definite matrix is always invertible.
4. **Task** Summarize the previous results in a single sentence: When does the linear regression problem $\hat{\mathbf{w}} = (X^T X)^{-1} X^T \mathbf{y}$ have a unique solution?

Task 2: Age Prediction

The *IMDB-Wiki* dataset contains images of people labeled with the current age at depiction. Assume you have implemented a loader for the images that yield colored images as arrays $p \in \mathbb{R}^{512\times512\times3}$.

(a) 63 years **(b)** 32 years **(c)** 31 years

Figure 1 Three examples from the *IMDB-Wiki* dataset. The task is to predict the age of the person depicted.

1. **Task** How would you build a linear regression model for the task of age prediction? What is the dimension d of data points?
2. **Task** Further assume that you have collected $n = 10000$ annotated images for training. Given you use 32bit floating point numbers (= 4 bytes), how much memory is required to store the matrices X and X^TX?
3. **Task** Can you use $\mathbf{w} = (X^TX)^{-1}X^T\mathbf{y}$ to compute the weights for your model on your home computer?
4. **Task** Elon Musk's company xAI built a supercomputer named *Colossus* which was launched in September 2024. It currently contains $2 * 10^5$ *Nvidia H100* GPUs and was announced to be extended to 1 million GPUs. Can you use $\mathbf{w} = (X^TX)^{-1}X^T\mathbf{y}$ to compute the weights for your model with *Colossus*?

Figure 2 The site of the Colossus Supercomputer in South Memphis

5. **Task** Suggest several options for how to deal with the problem. (And show in the best case mathematically that it works.)

Task 3: Linear Regression feature maps

Follow the instructions in the jupyter notebook.

Task 4: Linear Regression on Boston housing dataset

Follow the instructions in the jupyter notebook.