Assignment 3

Mustererkennung/Machine Learning WiSe 20/21

Do **one** of the following two exercises:

Excercise 1. Linear Regression: Trump vs. Biden

There has been an "unexplainable" shift in the US presidential election. Especially in Georgia and Pennsylvania, Trump had a substantial lead before getting to the last 10 percent of collected votes. Use the dataset[1] and do a Linear Regression using the provided notebook in the KVV to debunk this statistical "impossibility".

(a) Make a prediction: at what percentage is Biden going to lead in the both states, respectively.

Excercise 2. Logistic Regression

Logistic Regression can be interpreted as a neural network with just one layer. It uses the Cross Entropy to measure the performance of the layer (i.e. of the "trained" weight **w** and bias **b**). In ML we call this the **Loss function**.

Implement Logistic Regression using Python (incl. Numpy etc.) and use it on the "ZIP-Code"-Dataset [2]. Implement the Cross Entropy and the Sigmoid function from scratch. Use gradient descent to optimize.

(a) What happens when you take the Means Squared Error (MSE) instead of the Cross Entropy? Does this also work? Implement MSE and try for yourself.

(b) (*Optional*) Can you think of a way to classify more than one class (in this case 10 classes)? How would you change the way **w** and **b** is defined?

Recommended Reading (Preparation for Neural Networks) If you are unfamiliar with matrix calculus, read sections 3 and 4 from *The Matrix Calculus You Need For Deep Learning*[3]. This is also good as a refresher.

Submit as a fully executed Jupyter Notebook as .ipynb and as PDF in KVV. Deadline: 07.12.20, 10.00h.

[1] `https://github.com/alex/nyt-2020-election-scraper` battleground-state-changes.csv

[2] `https://web.stanford.edu/~hastie/ElemStatLearn/data.html`

[3] `https://arxiv.org/pdf/1802.01528.pdf`