Delprov 2

24/10 until 6/11 2023

Turn in on Canvas:
One common pdf-file of your Answers and your python code and your Name, personal number and date.

i) Your answers should comment all your code to explain what you do and why.
ii) Reflect, comment and discuss your results.
iii) Your grade will be an overall assessment of your answers.

You may either use own data that is suitable for the following tasks or you can choose to work on the MNIST Fashion data according to:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split

from keras.datasets import fashion_mnist as fashion_mnist
(X_train,y_train),(X_test,y_test)=fashion_mnist.load_data()
```

Task 1

a) Put all data into one common DataFrame, both features and targets. Show that you obtained the correct dataframe.
Hints: check array's dimensions with the 'shape' command.
convert 3D to 2D arrays with the 'reshape' command.
Concetanete two dataframes (row-vise) with 'concat' and axis=0.

b) Select out targets 0 to 3 if your last name starts with A to G.
Select out targets 4 to 7 if your last name starts with H to M.
Select out targets 6 to 9 if your last name starts with N to Ö.
Continue the rest of the assignment with only your targets !

Task 2

a) Make a PCA plot of your training data, with datapoints colored by the target class.
b) In your own words, explain how PCA works and what mathematical Variance mean!

Task 3

Make a Random Forest classification prediction on your Data and make a Confusion matrix of your results.
What is your accuracy ? Discuss your results!

Task 4

Repeat the above tast 3, but use Logistic regression instead. Which one worked better on your data ?

Task 5 (for higher grade)

Perform an investigation of your choice of different neuron architectures in a basic classification NN on your dataset and discuss your results.

Good Luck !

/Niklas