Course: Numerical Analysis for Machine Learning
Prof. E. Miglio - January 19th 2021
Duration of the exam: 2.5 hours.

Exercise 1

We consider a database containing geometrical features of iris plants. The dataset can be loaded with the following commands:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

data = pd.read_csv('http://archive.ics.uci.edu/ml/machine-learning-databases/iris/
iris.data', header = None).to_numpy()

A = data[:, :4].T.astype(np.float64)
labels = data[:,4]
groups = ('Iris-setosa','Iris-versicolor','Iris-virginica')
```

Each column of the matrix A refers to a sample. Each rows corresponds to a feature. Specifically:

- the 1st row contains the sepal length in cm;
- the 2nd row contains the sepal width in cm;
- the 3rd row contains the petal length in cm;
- the 4th row contains the petal width in cm.

The vector labels contains the class of iris plants each samples belongs to. There are three classes: 'Iris-setosa', 'Iris-versicolor' and 'Iris-virginica'.

1. How many samples are there in the dataset? How many samples belong to each class?
2. Perform PCA on the dataset by means of the SVD decomposition. Then, plot the trend of
   - the singular values σ_k;
   - the cumulate fraction of singular values $(\sum_{i=1}^{k} \sigma_i)/(\sum_{i=1}^{q} \sigma_i)$;
   - the fraction of the "explained variance" $(\sum_{i=1}^{k} \sigma_i^2)/(\sum_{i=1}^{q} \sigma_i^2)$.
3. Compute a matrix containing the principal components associated with the dataset.
4. Generate a scatterplot of the first two principal components of the dataset, grouped by label.
5. Comment on the results of point 4, in light of the results of point 2.

Exercise 2 Give a brief explaination of the Gradient Descent method and motivate the introduction of the Stochastic Gradient Descent (SGD).

Consider the following dataset

```
import numpy as np
m = 100
noise = 1.0
coeff_exact = np.array([5.0, 1.0])

np.random.seed(0)
X = np.c_[[1]*100, 13.5 * np.random.rand(m, 1)]
y = X @ coeff_exact + noise * np.random.randn(m)
```

Use the SGD to fit a linear model to these data. Initialize the two unknown parameters using `np.random.randn(2)` and find suitable values for the learning rate and for the number of epochs; motivate your choices.

Exercise 3

Consider the neural network in Figure 1 with input $x \in \mathbb{R}$, 3 hidden layers with one node each and one output $y \in \mathbb{R}$.

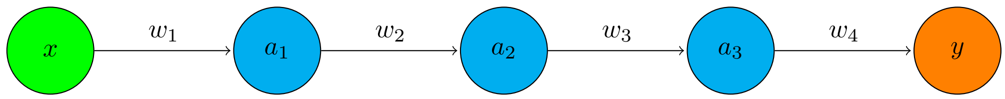

Figure 1: Simple neural network.

In the network each node corresponds to the sigmoid of the previous node multiplied by some weight *i.e.* $a_i = \sigma(w_i a_{i-1})$, $i = 1, \ldots, 4$ where $a_0 = x$ and $a_4 = y$.

- By using the chain rule compute $\dfrac{\partial y}{\partial x}$.
- Compute the maximum of σ' and discuss how this is related to the vanishing gradients problem.