Course: Numerical Analysis for Machine Learning
Prof. E. Miglio - January 18th 2022
Duration of the exam: 2.5 hours.

Exercise 1

Cardiovascular diseases (CVDs) are the number 1 cause of death globally, taking an estimated 17.9 million lives each year, which accounts for 31% of all deaths worlwide. This dataset contains the medical records of 299 patients who had heart failure, collected during their follow-up period, where each patient profile has 12 clinical features.

People with cardiovascular disease or who are at high cardiovascular risk (due to the presence of one or more risk factors such as hypertension, diabetes, hyperlipidaemia or already established disease) need early detection and management wherein a machine learning model can be of great help.

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

data = pd.read_csv('https://archive.ics.uci.edu/ml/machine-learning-databases/00519/
heart_failure_clinical_records_dataset.csv')
data_np = np.array(data)
A = data_np[:,:-1].astype(np.float64).T # matrix containing the data (num features x num patients)
labels = data_np[:,-1].astype(np.int32) # outcomes (0 = alive; 1 = death)
```

1. How many patients are associated with good and with bad outcome, respectively?

2. Perform PCA on the dataset by means of the SVD decomposition. Then, plot the trend of the following quantities and comment the results:

2.1. the singular values σ_k;

2.2. the cumulate fraction of singular values $\dfrac{\sum_{i=1}^{k}\sigma_i}{\sum_{i=1}^{q}\sigma_i}$;

2.3. the fraction of the "explained variance" $\dfrac{\sum_{i=1}^{k}\sigma_i^2}{\sum_{i=1}^{q}\sigma_i^2}$;

3. Generate a scatterplot of the first two principal components of the dataset, grouped by label; in light of the scatterplot:

3.1. Which one of the first two principal components correlates most with the outcome?

3.2. Is a bad outcome associated with a positive or negative value of the principal component you chose at point **3.1.** ?

3.3. Propone a risk predictor based on the sign of the principal component you chose at point **3.1.** .

3.4. Based on the risk predictor of point **3.3.**, compute the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN).

3.5. Compute the following quantities:

- sensitivity: $(TP)/(TP+FN)$;
- specificity: $(TN)/(TN+FP)$;
- accuracy: $(TP+TN)/(TP+TN+FP+FN)$.

Exercise 2

Consider the Rosenbrock function

$$f(x,y) = 100(y-x^2)^2 + (1-x)^2. \tag{1}$$

1. Compute the gradient ∇f and the Hessian of f. Prove that the function has a unique minimizer x^* and compute it.

2. Apply the Gradient Descent method with fixed learning rate (η) to find the minimum of f. Explore the effect of different values of η and different starting points. Comment on the results.

Exercise 3
Consider the neural network $\mathcal{N}$ defined by the following composite function:

$$\mathcal{N}(\mathbf{x}) = C(ReLU(B(ReLU(A\mathbf{x})))), \tag{2}$$

where:

- $ReLU$ is applied componentwise;
- $A, B \in \mathbb{R}^{50\times 50}$;
- $C \in \mathbb{R}^{5\times 50}$.

1. For the given matrices A, B and C what are the dimensions of the input and output of the neural network $\mathcal{N}$? Can the output be seen as a vector of probabilities ?

2. Consider the softmax function
$$\text{softmax}(\mathbf{x})_i = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \tag{3}$$
for $i \in \{1, \ldots, n\}$.

   For the given matrices A, B and C what are the dimensions of the input and output of the neural network $\text{softmax}(\mathcal{N})$? Can the output be seen as a vector of probabilities ?

3. If $\mathcal{N}$ is a classifier, for how many classes does it work?

4. Assume $\mathcal{N}$ has been trained and you want to use it to for classification on an input $\mathbf{x}$. Do you use $\text{softmax}(\mathcal{N})$ or $\mathcal{N}$? Justify your answer.