Course: <u>Numerical Analysis for Machine Learning</u>

Prof. E. Miglio - February 9th 2021

Duration of the exam: 2.5 hours.

Exercise 1

We consider a database containing six characterizing measurements for batches of plastic pellets. The outcome when using this material, either "Poor" or "Adequate", is also provided. The goal is to classify material lots according to quality, starting from the measurements.

The dataset can be created with the following commands:

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

dataframe = pd.read_csv('https://openmv.net/file/raw-material-characterization.csv')
data = dataframe.to_numpy()
A = data[:,2:].astype(np.float64)
labels = data[:,1]
groups = ('Adequate', 'Poor')
```

Each row of the matrix A refers to a sample (i.e. to a material lot). Each column corresponds to a feature.

The vector labels contains the material quality of the sample with corresponding index.

1. How many samples are there in the dataset? How many poor quality and adequate quality samples are there?
2. By exploiting the SVD decomposition, perform PCA on the provided data. Then, plot the trend of the singular values σ_k in logarithmic scale.
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. Propose a simple classifier to discriminate among poor and adequate quality materials, based on the second principal component. Then compute the accuracy of the classifier (i.e. the fraction of correctly classified samples).

Exercise 2

Consider the function

$$f(x) = f_1(x) + f_2(x), \tag{1}$$

where $f_1(x) = (x-1)^2$ and $f_2(x) = (x+1)^2$.

Use the Stochastic Gradient Descent (SGD) method to find the minimum of $f(x)$; each iteration of the SGD is given by

$$x_{k+1} = x_k - \eta_k \nabla f_{i(k)}(x_k), \tag{2}$$

where $i(k) \in \{1, 2\}$ is drawn uniformly. Initialize the method with a random initial condition x_0 and perform 5000 iterations; consider the following two choices for η_k:

- $\eta_k = \eta = 10^{-3}$;
- $\eta_k = \frac{1}{10+k}$.

Plot the convergence history in the two cases.

Find experimentally how the quantity $\mathbb{E}(f(x_k)) - f(x^*)$ (where x^* is the point where $f(x)$ attains the minimum) depends on the number of performed iteration. Comment the result.

Exercise 3

Given an input $\mathbf{x} \in \mathbb{R}^2$, a weight vector $\mathbf{w} \in \mathbb{R}^2$ and a bias $b_0 \in \mathbb{R}$, draw the computational graph for the computation of the mean squared error $L = MSE(\hat{y}, y)$ of a prediction $\hat{y} = \sigma(\mathbf{w}^T\mathbf{x} + b_0)$ with respect to the true value y; σ is the sigmoid function.

Consider the following values for $\mathbf{x}, \mathbf{w}, b_0$ and y:

$$\mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} \frac{4}{5} \\ \frac{7}{5} \end{bmatrix}, \quad b_0 = \frac{3}{5}, \quad y = 1, \tag{3}$$

and the MSE function $L = (\hat{y} - y)^2$. Report the correponding values on the computational graph.

Explain how the backpropagation method can be used to compute the gradient of L with respect to the inputs.

Draw the computational graph for the computation of the gradient and calculate the gradients values for each edge and node in the computational gradient graph.