**Course: Numerical Analysis for Machine Learning**

Prof. E. Miglio - June 14th 2024
Duration of the exam: 2.5 hours.

**Exercise 1**

Consider the Breast Cancer Wisconsin dataset. It contains features computed from digitized images of fine needle aspirate (FNA) of breast masses. These features are used to classify the tumors into malignant (cancerous) or benign (non-cancerous).

Here's an overview of the dataset:

- Number of Instances: 569,
- Number of Features: 30 numeric (real-valued features),
- Target Variable: Binary (0 for malignant, 1 for benign).

1. How many patients with and without cancer are there in the dataset ?
2. Choose two features, and make a scatterplot of the corresponding values plotting with different colors the data corresponding to patients with and without cancer.
3. Perform the PCA on the data, the plot the trend of the following quantities:
   - the singular values $\sigma_k$;
   - the cumulate fraction of singular values: $\frac{\sum_{i=1}^k \sigma_i}{\sum_{i=1}^q \sigma_i}$;
   - the cumulate fraction of the "explained variance": $\frac{\sum_{i=1}^k \sigma_i^2}{\sum_{i=1}^q \sigma_i^2}$;
4. Make a scatterplot of the first two principal components of the patients.
5. Implement a function for computing the randomized SVD of rank $k$ for a generic matrix.
6. Set $k = 10$ and compute the randomized SVD of the dataset and the principal components.
7. Select and display the top five most influential features for each principal axis. (The names of the features can be found in `data.feature_names`).
8. Compute the relative reconstruction error between the original matrix $X$ and the approximation $\hat{X}$ obtained using the randomized SVD and by varying $k$. The expression of the relative reconstruction error is given by
$$\epsilon_R = \frac{\|X - \hat{X}\|_F}{\|X\|_F}.$$
   Visualize the trend of the error with respect to the rank $k$ and comment the results.

**Exercise 2**

Load the data contained in file `ex2.txt` as follows:

```
import pandas as pd
df = pd.read_csv('./ex2.txt', sep=',', header=None)
df.columns = ['Score_1', 'Score_2', 'label']
m = df.shape[0]
X = np.hstack((np.ones((m,1)),df[['Score_1', 'Score_2']].values))
y = np.array(df.label.values).reshape(-1,1)
```

The dataset contains 100 rows that represents students. There are 3 columns: the first two columns are the marks obtained by the students in the 2 most difficult exams of the first year of the Master; the value in the third column is 1 if the student has been able to complete the Master in 2 years otherwise is 0.

Consider a logistic regression model given by:

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

where $\theta \in \mathbb{R}^3$ is the parameter vector. The cost function for logistic regression is given by the log-loss:

$$J(\theta) = -\frac{1}{100} \sum_{i=1}^{100} \left[ y_i \log(h_\theta(x_i)) + (1 - y_i) \log(1 - h_\theta(x_i)) \right].$$

1. Derive the gradient of the cost function $J(\theta)$ with respect to the parameter vector $\theta$.
2. Implement the Stochastic Gradient Descent (SGD) algorithm to minimize $J(\theta)$.
3. Apply the SGD in order to compute the optimal value of $\theta$.

**Exercise 3**

Consider the following function $g : \{-1, +1\}^N \to \{-1, 1\}$:

$$g(\mathbf{x}) = \begin{cases} 1 & \text{if } \sum_{i=1}^{N} x_i \in [S_{min}, S_{max}], \\ -1 & \text{otherwise}, \end{cases} \tag{1}$$

where $S_{min}, S_{max} \in \mathbb{Z}$ and $-N \leq S_{min} \leq S_{max} \leq N$.

1. Show that in general $g(\mathbf{x})$ cannot be reproduced using a single perceptron.
2. Show that the function $g(\mathbf{x})$ can be reproduced using a network with one hidden layer and two neurons using
$$\sigma(z) = \text{sign}(z) = \begin{cases} +1 & \text{if } z \geq 0, \\ -1 & \text{otherwise}, \end{cases} \tag{2}$$
with all weights and biases integers.
3. Show that the function $g(\mathbf{x})$ can be reproduced using a network with one hidden layer and two neurons using $\sigma(z) = ReLU(z)$.