

AY: 2024-2025
MIDTERM | Machine Learning
Nov. 2024

L3-S5: Dept. of Electrical Engineering
Teacher: A. Mhamdi
Time Limit: 1h

This document contains 6 pages numbered from 1/6 to 6/6. As soon as it is handed over to you, make sure it is complete. The 3 tasks are independent and can be treated in the order that suits you.

The following rules apply:

- ❶ No document is allowed in the examination room.
- ❷ Any electronic material, except basic calculator, is prohibited.
- ❸ Mysterious or unsupported answers will not receive full credit.
- ❹ Round results to the nearest thousandth (i.e., third digit after the decimal point).
- ❺ Task N°3: Each correct answer will grant a mark with no negative scoring.

Task N°1

⌚ 25mn | (6 points)

Considering this code:

```
[1]: import numpy as np
```

```
[2]: np.set_printoptions(precision=1)
```

```
[3]: X = 2 - 3 * np.random.randn(5, 4)
      X
```

```
[3]: array([[ -0.1,  -1. ,  -0.6,  -0.7],
            [-0.5,  -1.2,   1.7,   6.8],
            [-0.4,   1.3,  -1.3,  -1.4],
            [ 3. ,   4.4,   1.4,   1.3],
            [ 6.6,   7.2,  -0.2,  -1.4]])
```

```
[4]: X.min(axis=0), X.max(axis=0), X.mean(axis=0), X.std(axis=0)
```

```
[4]: (array([-0.5, -1.2, -1.3, -1.4]),
      array([6.6, 7.2, 1.7, 6.8]),
      array([1.7, 2.1, 0.2, 0.9]),
      array([2.8, 3.2, 1.1, 3.1]))
```

```
[5]: Y = np.random.randint(-5, 10, (5, 4))  
Y
```

```
[5]: array([[ 5,  8,  5,  0],  
          [-1,  9, -5,  4],  
          [-5, -5, -5,  3],  
          [-4, -2,  8,  4],  
          [ 2,  4,  8, -5]])
```

```
[6]: Y.min(axis=0), Y.max(axis=0), Y.mean(axis=0), Y.std(axis=0)
```

```
[6]: (array([-5, -5, -5, -5]),  
      array([5, 9, 8, 4]),  
      array([-0.6,  2.8,  2.2,  1.2]),  
      array([3.7, 5.5, 6. , 3.4]))
```

```
[7]: from sklearn.preprocessing import MinMaxScaler, StandardScaler
```

```
[8]: mms = MinMaxScaler()  
sc = StandardScaler()
```

```
[9]: mms.fit(X)  
mms.transform(Y)[2:,-2:]
```

```
[10]: ((Y - X.min(axis=0))/(X.max(axis=0) - X.min(axis=0)))[-1:,-1:]
```

```
[11]: sc.fit(X)  
sc.fit_transform(Y)  
sc.transform(X)[1:-1,1:-1]
```

```
[12]: ((X - Y.mean(axis=0))/Y.std(axis=0))[1,:1]
```

What does each of the following cells produce?

(a) (2 points) Cell #9?

```
[9]: array([[ -1.2,  0.5],  
          [ 3.1,  0.7],  
          [ 3.1, -0.4]])
```

(b) (1 point) Cell #10?

```
[10]: array([[ -0.4]])
```

(c) (2 points) Cell #11?

```
[11]: array([[ -0.7, -0.1],
            [-0.3, -0.6],
            [ 0.3, -0.1]])
```

(d) (1 point) Cell #12?

```
[12]: array([[0.1]])
```

Task N°2

⌚ 15mn | (4 points)

We have a dataset that includes the number of hours studied, the number of practice problems completed, and the corresponding scores received in a test. We want to perform linear regression to predict the test score based on these two features. The code is provided hereafter :

STEP 1: Import libraries

```
[1]: import numpy as np
import pandas as pd
from sklearn.linear_model import LinearRegression
```

```
[2]: np.set_printoptions(precision=3)
```

STEP 2: Create the dataset

```
[3]: data = {
    'Hours_Studied': [1, 2, 3, 4, 5, 6],
    'Practice_Problems': [5, 10, 15, 20, 25, 30],
    'Test_Score': [50, 55, 65, 70, 75, 80]
}
df = pd.DataFrame(data)
```

```
[4]: df
```

```
[4]:   Hours_Studied  Practice_Problems  Test_Score
0              1                5         50
1              2               10         55
2              3               15         65
3              4               20         70
4              5               25         75
5              6               30         80
```

STEP 3: Fit the linear regression model

```
[5]: X = df[['Hours_Studied', 'Practice_Problems']].values  
y = df['Test_Score'].values
```

```
[6]: model = LinearRegression()  
model.fit(X, y)
```

```
[6]: LinearRegression()
```

```
[7]: model.coef_, model.intercept_
```

```
[7]: (array([0.236, 1.181]), 44.33333333333333)
```

STEP 4: Make predictions

```
[8]: sample_data = np.array([[4, 30]])
```

```
[9]: model.predict(sample_data)
```

(a) (2 points) What will be the output of cell #9?

```
[9]: array([80.718])
```

It is equivalent to doing the following:

```
[10]: np.sum( model.coef_ * sample_data ) + model.intercept_
```

```
[10]: 80.71794871794872
```

(b) (2 points) For someone who solved 20 problems and spent 5 hours studying. What would his test result be?

```
[11]: test_data = np.array([[5, 20]])
```

```
[12]: np.sum( model.coef_ * test_data ) + model.intercept_
```

```
[12]: 69.14102564102564
```

```
[13]: model.predict(test_data)
```

```
[13]: array([69.141])
```

AY: 2024-2025

L3-S5: Dept. of Electrical Engineering

MIDTERM | Machine Learning

Nov. 2024

Teacher: A. Mhamdi

Full Name:

ID:

Class:

Room:

Time Limit: 1h

ANSWER SHEET

Task N°3

⌚ 20mn | (10 points)

- (a) (1 point) Which of the following numpy line of code would sum the values in a matrix \mathcal{A} horizontally?
- ☐ A.sum() ☐ A.sum(axis=0) ☒ A.sum(axis=1)
- (b) (1 point) On a straight line, the function's derivative "_____".
- ☐ changes as values on axis increase/decrease ☒ doesn't change
- (c) (1 point) A convex function always has multiple local optima.
- ☐ True ☒ False
- (d) (1 point) Examine the following code.

```

IPython: home/mhamdi
(venv) >> ipython
Python 3.10.12 (main, Sep 11 2024, 15:47:36) [GCC 11.4.0]
Type 'copyright', 'credits' or 'license' for more information
IPython 8.24.0 -- An enhanced Interactive Python. Type '?' for help.

In [1]: def bar(a, b=[]):
...:     b.append(a)
...:     return b
...:

In [2]: bar(1);

In [3]: bar(2);

In [4]: bar(3, []);

In [5]: 

```



The values of `bar(1)`, `bar(2)` and `bar(3, [])` are respectively

☒ `[1]`, `[1, 2]`, `[3]`

☐ `[1]`, `[2]`, `[3]`

☐ `[1]`, `[1, 2]`, `[1, 2, 3]`

☐ `[1]`, `[2]`, `[]`

(e) (1 point) The main purpose of the numpy library is to:

☐ Construct machine learning models.

☐ Achieve scientific computations.

☒ Perform computations on arrays efficiently.

☐ Visualize results in 2D and 3D plots.

(f) (1 point) What is the correct order for using a model.

① fit the model.

② clean the data.

③ split the data into training and test sets.

④ evaluate model accuracy.

☐ ① → ② → ③ → ④

☐ ④ → ③ → ② → ①

☒ ② → ③ → ① → ④

☐ ② → ① → ③ → ④

☐ ③ → ② → ④ → ①

(g) (1 point) What kind of array has dimensions in this format: `(7,)`?

☐ A rank 0 array ☒ A rank 1 array ☐ An identity matrix

(h) (1 point) When we talk about the class of regression (*linear or non linear*), we are talking about the “_____”.

☐ logs ☐ characters ☒ coefficients

(i) (1 point) Is polynomial linear regression a linear model?

☒ Yes ☐ No

(j) (1 point) Which statistical measure is used to determine the central tendency of a dataset?

☒ Mean ☐ Variance ☐ Standard deviation