

Due data: 9/23/2020, end of the day.

For question 1 , please submit a PDF file via Canvas.

For question 2 (programming question), please submit an .ipynb file via Canvas.

Please answer the following questions related to Machine Learning concepts:

1. [18 points] Explain the following concepts:

- 1) supervised learning,
- 2) unsupervised learning,
- 3) online learning,
- 4) batch learning,
- 5) model-based learning,
- 6) instance-based learning.

Programming Problem:

2. [42 points] In this problem, we write a program to estimate the parameters for an unknown polynomial using the `polyfit()` function of the `numpy` package.

- 1) Please plot the noisy data and the polynomial you found (in the same figure). You can use any value of `m` selected from 2, 3, 4, 5, 6.
- 2) Plot MSE versus order `m`, for `m = 1, 2, 3, 4, 5, 6, 7, 8` respectively. Identify the best choice of `m`.
- 3) Change variable `noise_scale` to 150, 200, 400, 600, 1000 respectively, re-run the algorithm and plot the polynomials with the `m` found in 2). Discuss the impact of noise scale to the accuracy of the returned parameters. [You need to plot a figure like in 1) for each choice of `noise_scale`.]
- 4) Change variable `number_of_samples` to 40, 30, 20, 10 respectively, re-ran the algorithm and plot the polynomials with the `m` found in 2). Discuss the impact of the number of samples to the accuracy of the returned parameters. [You need to plot a figure like in 1) for each choice of `number_of_samples`.]

A simulated dataset will be provided as below. The polynomial used is $y = 5 * x + 20 * x^2 + x^3$.

Simulated data is given as follows in Python:

```
import matplotlib.pyplot as plt

plt.style.use('seaborn-whitegrid')

import numpy as np

noise_scale = 100

number_of_samples = 50

x = 25*(np.random.rand(number_of_samples, 1) - 0.8)

y = 5 * x + 20 * x**2 + 1 * x**3 + noise_scale*np.random.randn(number_of_samples, 1)

plt.plot(x,y,'ro')
```