

Due data: 2/20/2019, end of the day.

For questions 1 – 3, please submit a PDF file via Canvas.

For question 4 (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.
2. [6 points] What is *overfitting* of training data? What is *regularization*?
3. [6 points] Prove Bayes' Theorem.

Programming Problem:

4. [40 points] In this problem, we write a program to find the coefficients for a linear regression model. You need to use Python to implement the following methods of finding the coefficients:
 - 1) Normal equation, and
 - 2) Gradient descent (batch or stochastic mode)
 - a) Print the cost function vs. iterations
 - b) Discuss how you choose the right alpha (learning rate). For example, you can plot cost function vs. learning rate to determine the best learning rate.

A simulated dataset will be provided, your job is to find the coefficients that can accurately estimate the true coefficients. Your solution should be 2 for intercept and 3, 4, 5 for the three coefficients.

Please do NOT use the `fit()` function of the Scikit-Learn library in your program. (You need to implement the Gradient Descent algorithm in your code.)

Simulated data is given as follows in Python:

```
import numpy as np

x1 = 2 * np.random.rand(100, 1)

x2 = 2 * np.random.rand(100, 1)

x3 = 2 * np.random.rand(100, 1)

y = 3 * x1 + 4 * x2 + 5 * x3 + np.random.randn(100, 1) + 2
```