ADAPTIVE COMPUTATION AND MACHINE LEARNING (COMS4030A/COMS7047A)
Assignment I

Submission due date: 26th March, 2021, 17:00

March 12, 2021

1   Description

Assignment 1 will assess material covered on Linear Regression (Week 2). Assignment 1 consists of two parts: 1. Theoretical questions, 2. Programming exercise. Separate submission links will be provided for each part, and it is a student's responsibility to make appropriate submissions. Please append your submissions with your student number and name.

NOTE: Assignment 1 has to be done individually. THIS IS NOT A GROUP ASSIGNMENT.

1. Assume a hypothesis $h_\theta(x)$ that overfits given data X. Explain how minimising parameters θ or using a subset of features (from input X) may help reduce the complexity of the updated hypothesis and may "correctly" fit the hypothesis onto the data? (5 points)

2. In class, we fixed learning rate to a constant value through out the learning process. Explain how a fixed learning rate still helps gradient descent algorithm to converge to global/local minima? Use cost function plots to explain your understanding. (5 points)

3. Suppose you have a regression data generated by a polynomial of degree 5. Characterize the bias-variance of the estimates of the following models on the data with respect to the true model by selecting the appropriate choice. (5points)

Model	Bias	Variance
Linear regression	low/high	low/high
Polynomial regression with degree 2	low/high	low/high
Polynomial regression with degree 5	low/high	low/high
Polynomial regression with degree 10	low/high	low/high
Polynomial regression with degree 20	low/high	low/high

Submission: Submit a single .pdf file containing your answers.

2 Programming exercise: (85 points)

In this part, you will implement linear regression and see it work on a simple dataset of your choice.

2.1 Python Installation:

It is recommended that you use Anaconda distribution for installing python notebooks and other essential libraries. Use the following link to go to Anaconda's download page :

   https://docs.anaconda.com/anaconda/install/

Make sure to download Python 3.6 version.

2.2 Submissions:

1. jupyter python notebook containing your code
2. your sample dataset

Submission links will be posted on Ulwazi.

2.3 Grading:

Grading for programming part will be based on the following exercises:
1. Basic implementation of linear regression - 40 points
2. Exploring the effect of different learning rates on convergence - 10 points
3. Implementation of feature scaling, feature standardization and regularization for improved learning - 20 points
4. Data visualization to understand the working of algorithm and other steps - 15 points