ECE 805 – Machine Learning

Homework #01

Spring 2023

(due on Thursday, 02 February 2023, 16:00, submitted via Teams)

General instructions

This assignment consists of two questions. You are requested to submit both your code and a report; the report should contain the requested visualisation plots and your answers to possible questions.
Please submit one compressed file (zip), named ECE805_HW1_Yourname.

Exercise 1 (50%):

a) Explain the principle of the Gradient Descent (GD) algorithm. Explain all the terms that you introduce.

b) Given the cost function for Linear Regression $J(\theta) = \frac{1}{2N}\sum_{i=1}^{N}\left(\theta^T x^i - y^i\right)^2$, where N is the number of examples, (x^i, y^i) is an example, and θ the parameters of the linear regression model to be learnt, derive the GD update rule (i.e., Widrow-Hoff learning rule) for a particular parameter θ_j:

$$\theta_j \leftarrow \theta_j - \alpha \frac{1}{N}\sum_{i=1}^{N}\left(\theta^T x^i - y^i\right)x_j^i$$

Exercise 2 (50%):

In this exercise, we will investigate the issues of underfitting and overfitting, as we discussed in Lecture 1 (slides 41-44) and in subsequent lectures. The dataset (x_i, y_i), where $i = 1, 2, \dots N$, is generated by the following function:

$$f(x) = 0.4\cos(2.1\pi x) + 0.55\sin(7\pi x) + 0.5$$

$$y_i = f(x_i) + \varepsilon_i$$

where the input points x_i are generated randomly (uniformly) in the region [0, 1] and ε_i is the measurement noise, which is a random variable assumed to satisfy a normal distribution with zero mean and variance σ^2. We consider the square error function

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{N}(y_i - \varphi(x_i)^T\theta)^2$$

where φ is the approximation function. In this exercise, we consider polynomial basis functions.

a) Let the number of data points $N = 10$ and the variance of the measurement noise be $\sigma^2 = 0.2$. Let φ be polynomial basis functions of varying order M. Using Python, solve the least squares problem for 6 cases: $M = 1, 3, 6, 9, 12, 15$. Plot six plots, corresponding to each different value of M. In each plot include the 10 data points, the true function $f(x)$ and the polynomial approximation with the optimal coefficients.

b) Now, fix $M = 9$. The measurement noise is again $\sigma^2 = 0.2$. In this case, vary the number of data points as follows: $N = 10, 50, 100, 200, 500$. Plot five plots corresponding to each different value of N. In each plot, include the N data points, the true function $f(x)$ and the polynomial approximation with the optimal coefficients.

c) In this experiment, we fix $M = 9$ and $N = 100$ and vary the variance of the measurement noise as follows: $\sigma^2 = 0, 0.2, 0.5, 0.8, 1$. Plot five plots corresponding to each different value of σ^2.