ECE 805 — MACHINE LEARNING
Spring 2021
Homework no. 1

(due on Monday, 12 April 2021 by 1:00pm, submitted via Teams to the TA Rafaella Elia)

1. (50%) Consider the following eight data points (x_i, y_i), where $i = 1, 2, \ldots, 8$:

$$\{(0, 2.5), (0.5, 1.5), (1, 1.5), (1.5, 1), (2, 1), (2.5, 0.5), (3, -0.5), (3.5, -2)\}$$

We would like to approximate this dataset by polynomials (of different order) by minimizing the weighted square error

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{8} \omega_i \left(y_i - \phi(x_i)^T \theta \right)^2$$

where ϕ represents the polynomial basis functions $\phi(x) = [1, x, x^2, \ldots x^M]^T$, with M being the order of the polynomial, ω_i is the weight associated with the i-th data point, and θ is the parameter vector (coefficients) to be determined. This is the weighted least-squares problem discussed in Lecture 5-6 (see slide 10). In all your answer, show clearly all the steps of your calculations, the final answers, as well as your MATLAB code.

(a) Assume that the weights associated with each data point is the same ($\omega_i = \omega = 1/8$) and the polynomial is first-order (i.e., a straight line). Solve by hand (not using MATLAB) the least squares problem; in other words, find the parameters θ_0, θ_1, such that $\hat{f}(x) = \theta_0 + \theta_1 x$ minimizes the above squared error function. Verify you answer using MATLAB.

(b) Now assume that the weights associated with each data point is the same ($\omega_i = \omega = 0.1$) EXCEPT the last data point $(3.5, -2)$, whose weight is $\omega_8 = 0.3$; i.e., the last point is weighted three times more than the other data points (we have more confidence in this data point). Solve again the weighted least squares problem by hand and verify using MATLAB. Make a single plot on MATLAB that includes the eight data points and the two optimal straight lines, one with equal weights (part (a)) and the other with the more emphasis on the last data point.

(c) Using MATLAB (not by hand!), repeat part (a) for the case $M = 2, 3, 4, 5, 6, 7$; i.e., as the order of the polynomial increases. Find the optimal parameter θ^* for each case, and make a single plot that includes the data points and the optimal polynomial for the five cases: $M = 1, 2, 3, 5, 7$.

(d) For each optimal weight vector θ^* corresponding to the cases $M = 1, 2, 3, 4, 5, 6, 7$, compute the optimal cost $J(\theta^*)$ and plot it versus M. Hint: the optimal cost should be decreasing as M increases.

2. (50%) In this exercise, we will investigate the issues of underfitting and overfitting, as we discussed in Lecture 1 (slides 21-24) and in subsequent lectures. The dataset (x_i, y_i), where $i = 1, 2, \ldots, N$, is generated by the following function:

$$\begin{aligned} f(x) &= 0.45\cos(2.1\pi x) + 0.6\sin(7\pi x) + 0.55 \\ y_i &= f(x_i) + \varepsilon_i, \end{aligned}$$

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. Similar to Exercise 1, we consider the square error function

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

where ϕ represents the approximation function. In this exercise, we consider polynomial basis functions and Radial Basis Functions (RBF).

(a) Let the number of data points $N = 10$ and the variance of the measurement noise be $\sigma^2 = 0.1$. Let ϕ be polynomial basis functions of varying order M. Using MATLAB, solve the least squares problem for 6 cases: $M = 1, 3, 5, 7, 8, 9$. 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 $\sigma^2 = 0.1$. In this case, vary the number of data points as follows: $N = 10, 20, 100, 500$. Plot four 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: $\sigma^2 = 0, 0.1, 0.2, 0.5$. Plot four plots corresponding to each different value of σ^2.

(d) In this experiment, we replace the polynomial function with a Gaussian Radial Basis Function (RBF) network. We assume that the number of data points is $N = 10$ and the measurement noise variance is $\sigma^2 = 0.2$. We assume that we have M radial basis functions, whose centers are chosen uniformly (equi-distant from each other, not randomly) between 0 and 1. Let $M = 10$. Try to adjust the variance γ of the RBF network to optimize the approximation performance of the RBF network. Plot the final result. Repeat for the case $M = 50$.