EE 660 **Homework Week 2** Posted: Wed., 9/4/2019
Jenkins Due: Wed., 9/11/2019, 2:00 PM

1. This problem may be solved using MATLAB or Python; the functions/commands stated below are for MATLAB implementations.

   You are to implement a simple curve fitting problem using 1D regression. In this problem you are to **code the assigned portions of the regressions yourself**; using a package's regression or curve-fit function will not suffice.

   (a) Model the curve to be fit, $\hat{f}(x)$, as a d^{th} order polynomial. Write down the mean- squared error objective function for curve fitting, in terms of x_i, y_i, and w_m, in which i is the data point index ($i = 1, 2, \cdots, N$), and m is the weight index ($m = 0, 1, \cdots, d$).

   (b) Write the objective function in matrix form in terms of $\underline{\underline{\Phi}}, \underline{y}$, and $\underline{w}$. ($\underline{\underline{\Phi}}$ is the basis-set expansion version of $\underline{\underline{X}}$; the i^{th} row is $\underline{\phi}^T(x_i)$.)

   (c) Download the provided data from the dropbox and plot only the points of `x_train` *vs.* `y_train` (use the command `scatter(x, y)`).

   (d) Let the hypothesis set be polynomials in x of degree [1, 2, 3, 7, 10]. Find the curve parameters (using only data from `x_train` and `y_train`) for each of these polynomial degrees, using pseudo-inverse. (You can use commands `hold on` and `plot(x, y)` to visualize how well the curve fits to the training data, but this is not mandatory.) Show the computed weight vectors $\underline{w}_1, \underline{w}_2, \underline{w}_3$, $\underline{w}_7, \underline{w}_{10}$ where $\underline{w}_d$ denotes the weight vector for the d^{th} order polynomial.

   Hint: to set this up as a pseudo-inverse problem, use the basis function expansion of part (b) above.

   (e) Compute the mean squared error (MSE) on the training set for each one, i.e.,

   $$MSE_d = \frac{1}{N}\sum_{i=1}^{N}\left[y_i - \underline{w}_d^T \underline{\phi}(\underline{x}_i)\right]^2.$$

   Plot error *vs*. polynomial degree. Which polynomial degree seems to be the best model based on the training sample MSE only?

   (f) Using the same weights ($\underline{w}_1, \underline{w}_2, \underline{w}_3, \underline{w}_7, \underline{w}_{10}$), compute the MSE for the test samples, *i.e.*, using `x_test` and `y_test`. Plot error *vs.* polynomial degree again. Which polynomial degree seems to be the best model based on the test sample MSE only?

   (g) Now, let's fix the polynomial degree to 7. Solve using ridge regression with penalty term $\lambda = [10^{-5}, 10^{-3}, 10^{-1}, 1, 10]$. Show the computed weights.

   (h) Compute train and test MSE of the fit from part (g) and plot both vs $\log(\lambda)$. What are your conclusions?

2. Murphy Exercise 7.4. **Hint:** Start from Murphy Eq. (7.8), and assume $\underline{\hat{w}}$ is given.

Problems 4-5 below involve reading and related short exercises, for upcoming lectures.

3. *Bayesian concept learning.* Read Murphy 3.1, 3.2 up to first paragraph of 3.2.4, inclusive. The rest of 3.2 is optional.

   Key concepts (to focus on during reading):

   – What learning is
   – Hypothesis space
   – Version space
   – Strong sampling assumption
   – Likelihood
   – Prior
   – Posterior
   – Posterior predictive distribution
   – How these combine to give a prediction probability

   (a) For the numbers game, take as the hypothesis set $\mathcal{H}$:

   $$\mathcal{H} = \left\{h_{odd}, h_{even}, h_2, h_{P2}, h_5, h_{P5}, h_7, h_{P7}\right\}$$

   in which

   $$\begin{aligned} h_{odd} &= \text{all odd numbers} \\ h_{even} &= \text{all even numbers} \\ h_2 &= \text{all numbers ending in 2} \\ h_{P2} &= \text{all powers of 2 (excluding } 2^0\text{)} \\ h_5 &= \text{all numbers ending in 5} \\ h_{P5} &= \text{all powers of 5 (excluding } 5^0\text{)} \\ h_7 &= \text{all numbers ending in 7} \\ h_{P7} &= \text{all powers of 7 (excluding } 7^0\text{)} \end{aligned}$$

   such that all hypotheses are limited to numbers between 1 and 100 (inclusive). Suppose the training data is $\mathcal{D} = \left\{5, 25\right\}$. What is the version space?

   (b) Also for the numbers game, let the training data $\mathcal{D} = \left\{16\right\}$. Suppose the hypothesis space $\mathcal{H} = \left\{h_{P2}, h_{P4}\right\}$, in which:

   $$h_{P2} = \left\{2,4,8,16,32,64\right\}$$

$$h_{P4} = \{4,16,64\}$$

Assume priors are $p(h_2) = 0.6$, $p(h_4) = 0.4$, and use the strong sampling assumption.

(i) Calculate the likelihood and the posterior for h_2.

(ii) Calculate the likelihood and the posterior for h_4.

(iii) Which posterior is larger?

4. *Bayesian linear regression.* Read Murphy 7.6.0, 7.6.1, 7.6.2.

To get an overview of the algebra from Eq. (7.54) to Eq. (7.55), show that $p\left(\underline{w} \middle| \underline{\underline{X}}, \underline{y}, \sigma^2\right)$ can be written in terms of $p\left(\underline{y} \middle| \underline{\underline{X}}, \underline{w}, \sigma^2\right)$ and a prior term. Label the posterior, likelihood, and prior terms. **Do not** assume Gaussian densities in this problem.