CX 4803 CML – Spring 2022

Assignment 2

There are **four** *questions. Please show your work as if you were explaining your solution to another student. You can use any programming language you like. Submit your solutions as a single pdf file on Canvas. Include your program listings in your pdf file. For example, if you use latex, you can use the* `listings` *package.*

1. Suppose we have n observations (x_i, y_i), $i = 1, \ldots, n$. We wish to fit the model $y = f(x; w)$ to these observations. Assume that the noise in the observations is Gaussian with mean 0 and unknown variance σ^2.

   Show in detail the steps (including obtaining derivatives) for deriving the maximum likelihood estimate for σ^2,

   $$\sigma^2_{\mathrm{ML}} = \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i; w_{\mathrm{ML}}))^2.$$

   Be sure to also explain the meaning of w_{ML}.

2. Consider polynomial regression. Write a program to plot the U-shaped curve for the test MSE as a function of the polynomial order. You are free to choose how the test and training data are generated. Test your program using different values of the noise variance. Show your plots for different noise variances. (Note: the U-shaped curve for the test MSE is not always perfectly decreasing then increasing.)

3. In this question, we will try to reproduce Figure 3.5 in the textbook by Bishop. Generate sinusoidal data

   $$y_i = \sin(2\pi x_i) + u_i$$

   with noise u_i for $i = 1, \ldots, n$, with $n = 25$, the sample size for one sample. Fit this data using the *regularized* linear basis function model

   $$f(x; w) = w_0 + \sum_{j=1}^{24} w_j \phi_j(x) + \lambda \|w\|_2^2$$

   where

   $$\phi_j(x) = \exp\left(-\frac{1}{2s^2}(x - \mu_j)^2\right).$$

   Choose the centers of the Gaussians μ_j to be equally spaced in the interval 0 to 1. Also choose $s = 0.1$ and $\lambda = 0.5$ initially.

   Plot your data and the function $f(x; w)$ that is fit to the data. Experiment with different values of λ (as in the textbook by Bishop) and explain what you observe in terms of bias and variance. Show a plot like that in Figure 3.5 of Bishop (20 plots are given for 20 samples).

   Explain the effect of larger and smaller values of s.

4. Continuing the above question, plot $(\text{bias})^2$ and the model variance as a function of $\ln \lambda$ like in Figure 3.6 of Bishop. Note that in order to compute the model variance, for example, you will have to average a certain quantity over many models (each model computed via a sample or dataset). It may be that Bishop used 100 datasets (or more) with each dataset having 25 data points. For the test MSE, 1000 data points were used, and these points may be the same for testing each of the 100 models. What is the value λ that achieves the minimum test MSE?