ECE 2372 - Homework 5

<< — Bias-Variance Decomposition — >>

Problem 1:

Consider a learning scenario where $x \in \mathbb{R}$ and $x \in \mathbb{R}$ is given by $y = x^2$.

Assume that the input variable x is drawn uniformly on the interval $[-1, 1]$ and we are given two independent observations $\mathcal{D} = \{(x_1, x_1^2), (x_1, x_2^2)\}$.

We'll consider two different approaches to fitting a line to this data set, and we will be interested in calculating the expected risk (in terms of squared error), the bias, and the variance.

(a) Suppose $h(x) = b$, i.e., constant line. We are considering here to fit the line by setting

$$\begin{aligned} h_{\mathcal{D}}(x) &= \frac{y_1 + y_2}{2} \\ &= \frac{x_1^2 + x_2^2}{2} \end{aligned}$$

Derive (analytically) the average hypothesis: $\bar{h}(x) = \mathbb{E}_{\mathbb{D}}\left[h_{\mathbb{D}}(x)\right]$

(b) Compute (analytically) the bias

$$\mathbb{E}_X\left[(\bar{h}(X) - X^2)^2\right]$$

(c) Compute (analytically) the variance

$$\mathbb{E}_X\left[E_{\mathbb{D}}\left[\ (h_{\mathcal{D}}(X) - \bar{h}(X))^2\right]\ \right]$$

(d) Now simulate this. Explicitly, numerically estimate average hypothesis, the bias, the variance and $\bar{h}(x) = \mathbb{E}_{\mathbb{D}}\left[R(h_{\mathbb{D}})\right]$

(e) Now consider a line of the form $h(x) = ax + b$ which we fit by selecting the line the passes through our two observations. Modify the code from part (d) to estimate the new $\bar{h}(x)$, bias, variance, and risk. Explain how the results change and why.

Problem 2: Let's consider the scenario described in Lecture 12, where x is drawn uniformly on the interval $[-1, 1]$ and $y = sin(\pi x)$. We have $n = 2$ training samples: $(x_1, y_1), (x_2, y_2)$. Here we will look at an alternative approach to fitting a line to our data based on "Tikhonov regularization". In particular, we let

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix} \quad \theta = \begin{bmatrix} b \\ a \end{bmatrix}$$

Then we consider Tikhonov regularized least squares estimators of the form

$$\hat{\theta} = (A^T A + \Gamma^T \Gamma)^{-1} A^T y \tag{1}$$

(a) How would we set Γ to reduce this estimator to fitting a constant function (i.e., finding an $h(x) = b$)?

For this problem, it is sufficient to set Γ in a way that just makes $a \approx 0$. To make $a = 0$ exactly requires setting Γ in a way that makes the matrix AT A + $A^T A + \Gamma^T \Gamma$ singular – but note that this does not mean that the regularized least-squares optimization problem cannot be solved, you must just use a different formula than the one given in Eq.(1).

(b) How would we set Γ to reduce this estimator to fitting a line $h(x) = ax + b$ that passes through the observed data points (i.e.,$(x_1, y_1), (x_2, y_2)$)?

(c) Estimate (numerically) the bias and variance for (at least approximations of) both of these estimators, and confirm that your estimates correspond to the numbers I provided in Lecture 13.

(d) Try and see if you can find a matrix Γ that results in a smaller risk than either of the two approaches we discussed in class. Report the Γ that gives you the best results. (You may restrict your search to diagonal Γ to simplify this.)