ECE 2372 - Homework 5

<< — Bias-Variance Decomposition — >>

Jiyang Liu 4731134

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

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

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

a)

$$\bar{h}(x) = E_D[h_D(x)] = E_D\left[\frac{x_1^2 + x_2^2}{2}\right] = \frac{1}{2}E_D[x_1^2] + \frac{1}{2}E_D[x_2^2]$$

Calculate the expectation $E[x^2]$:

$$E_D[x^2] = \int_{-1}^{1} p(x) * x^2\, dx$$

The x is drawn unformly on the interval $[-1,1]$, *therefore*

$$p(x) = \frac{1}{2}$$

$$E_D[x^2] = \int_{-1}^{1} \frac{1}{2} * x^2\, dx = \frac{1}{2} * \frac{x^3}{3}\Big|_{-1}^{1} = \frac{1}{3}$$

Substitute to the equation:

$$\bar{h}(x) = \frac{1}{2} * \frac{1}{3} + \frac{1}{2} * \frac{1}{3} = \frac{1}{3}$$

(b) Compute (analytically) the bias

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

b)

$$E_x\left[\left(\bar{h}(x) - x^2\right)^2\right] = E_x\left[\left(\frac{1}{3} - x^2\right)^2\right] = E_x\left[\frac{1}{9} - \frac{2}{3}x^2 + x^4\right]$$

$$= \frac{1}{9} - \frac{2}{3}E_x[x^2] + E_x[x^4]$$

Calculate the expectation $E[x^4]$:

$$E_x[x^4] = \int_{-1}^{1} p(x) * x^4 \, dx = \int_{-1}^{1} \frac{1}{2} * x^4 \, dx$$

$$= \frac{1}{2} * \frac{x^5}{5}\Big|_{-1}^{1} = \frac{1}{5}$$

Substitute to the equation:

$$Bias = E_x\left[\left(\bar{h}(x) - x^2\right)^2\right] = \frac{1}{9} - \frac{2}{3} * \frac{1}{3} + \frac{1}{5} = \frac{4}{45}$$

(c) Compute (analytically) the variance

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

c)

$$E_x\left[E_D\left[\left(h_D(x) - \bar{h}(x)\right)^2\right]\right] = E_x\left[E_D\left[\left(\frac{x_1^2 + x_2^2}{2} - \frac{1}{3}\right)^2\right]\right] = E_x\left[E_D\left[\frac{1}{9} - \frac{x_1^2 + x_2^2}{3} + \frac{x_1^4 + 2x_1^2x_2^2 + x_2^4}{4}\right]\right]$$

Calculate the expectation $E[x_1^2x_2^2]$:

$$E_D[x_1^2x_2^2] = \int_{-1}^{1}\int_{-1}^{1} p(x_1)p(x_2) * x_1^2x_2^2 \, dx_1dx_2 = \frac{1}{3} * \frac{1}{3} = \frac{1}{9}$$

Substitute to the equation:

$$Variance = E_x\left[\frac{1}{9} - \frac{\left(\frac{1}{3} + \frac{1}{3}\right)}{3} + \frac{\frac{1}{5} + 2 * \frac{1}{9} + \frac{1}{5}}{4}\right] = E_x\left[\frac{1}{9} - \frac{2}{9} + \frac{7}{45}\right] = E_x\left[\frac{2}{45}\right] = \int_{-1}^{1} \frac{1}{2} * \frac{2}{45} \, dx = \frac{2}{45}$$

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

The code is in HW5_1_d .

From interval $[-1,1]$, *drawn* 10000 *data set uniformly, where each set has* 2 *points. The results are as below.*

```
Scenario (a) - Constant line h(x) = b:
Average hypothesis: 0.33222032788657796
Bias: 0.08963788654539513
Variance: 0.043749727725588584
Risk: 0.13338761427098372
```

Analytically: *Average hypothesis* $= \frac{1}{3} \approx 0.3333$; *Bias* $= \frac{4}{45} \approx 0.0888$; *Variance* $= \frac{2}{45} \approx 0.0444$;

Risk $= \frac{4}{45} + \frac{2}{45} = \frac{6}{45} \approx 0.1333$

The analytically and numerically solution are pretty closed.

(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.

The code is in HW5_1_e .

```
Scenario (b) - linear line h(x) = ax+b:
Average hypothesis a: -0.008069736345342441
Average hypothesis b: 0.005221417846659537 5
Bias: 0.19466450093153284
Variance: 0.19786874186841863
Risk: 0.3925332427999515
```

Analytically: *Average hypothesis* $= \frac{1}{3} \approx 0.3333$; *Bias* $= \frac{4}{45} \approx 0.0888$; *Variance* $= \frac{2}{45} \approx 0.0444$;

Risk $= \frac{4}{45} + \frac{2}{45} = \frac{6}{45} \approx 0.1333$

For the analytically and numerically solution, the bias and the variances got much larger. This shows that linear estimator is not good for fitting $y = x^2$. *The reason is if we use the average coefficient of* $y = ax + b$ *for 2 hypothesis points, the average slope a will turns to 0, because the* $y = x^2$ *is symmetric about the* $y-axis$. *For the same reason, the average constant b also turns to 0. Therefore, the final estimator would be like* $y = 0$, *which is obviously worse estimator than* $y = b$ $(b = 0.33)$ *in previous question.*

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).

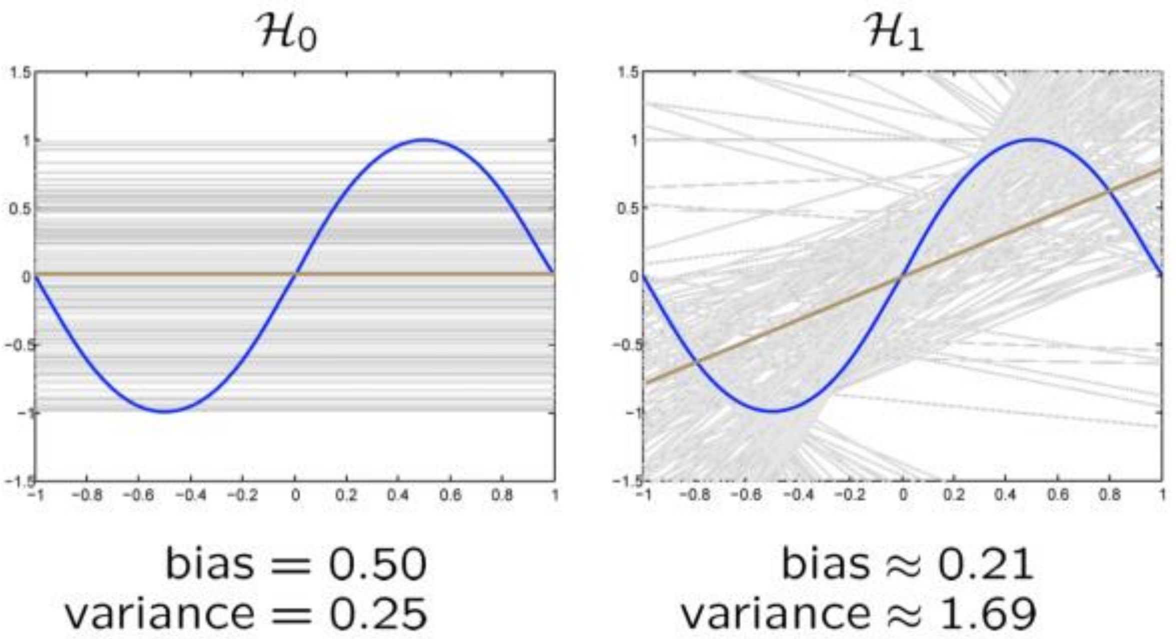
To let the estimator fitting the constant function, we set penalty factor γ *very large, which prevent the estimator to have much features.*

(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.,$(x_1, y_1), (x_2, y_2)$)?

To let the estimator fitting a linear function, we set penalty factor γ less than the previous one, which will let the estimator has less punishment for features.

(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.

The code is in HW5_2_c .




```
bias_of_avg_theta_hat: 0.5002577355538642
avg_theta_hat: [0.02746831 0.00310861]
avg_predictions: [0.02763696 0.02721668]
overall_bias: 0.7520053927363686
overall_variance: 9.510887533817765e-06
```