CSE 546 HW #1

Sam Kowash

October 18, 2018

(1) MLE and bias–variance tradeoff

(a) If we draw $x_1, \dots, x_n \sim \text{uniform}(0, \theta)$ for some positive parameter θ, then

$$\mathbb{P}(x_1, \dots, x_n \mid \theta) = \begin{cases} \left(\frac{1}{\theta}\right)^n & \text{if } x_1, \dots, x_n \in [0, \theta] \\ 0 & \text{else} \end{cases}$$

We can see that the likelihood increases as θ decreases as long as all points drawn still lie in $[0, \theta]$, so we conclude that the MLE is $\hat{\theta} = \max(x_1, \dots, x_n)$.

(b) Let $(x_1, y_1), \dots, (x_n, y_n) \in \mathbb{R}^d \times \mathbb{R}$ be drawn from some population. We define

$$\hat{w} = \arg\min_w \sum_{i=1}^{n} (y_i - w^T x_i)^2.$$

Let $(\tilde{x}_1, \tilde{y}_1), \dots, (\tilde{x}_m, \tilde{y}_m)$ be test data drawn from the same population. We define the training and test losses

$$R_{\text{tr}} = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{w}^T x_i\right)^2, \qquad R_{\text{te}} = \frac{1}{m} \sum_{i=1}^{n} \left(\tilde{y}_i - \hat{w}^T \tilde{x}_i\right)^2.$$