Homework 8

Karan Sarkar
sarkak2@rpi.edu

October 28, 2019

Exercise 4.3.

(a) As you increase the complexity of the target function, the deterministic noise will go up. This will result is a higher tendency to overfit.

(b) As you decrease the complexity of the hypothesis set, the deterministic noise will increase. However, a simpler hypothesis set will nevertheless reduce the tendency to overfit.

Exercise 4.5.

(a) The Tikhonov regularizer Γ should be the identity matrix. This ensures only the diagonal entries w_q^2 are counted.

(b) The Tikhonov regularizer Γ should be a matrix where every term is $1/\sqrt{n}$. Thus each entry of $\Gamma^T \Gamma = 1$. This ensures each term $w_p w_q$ occurs twice and diagonal entries occur once.

Exercise 4.6. The soft-order constraint would be more useful because it allows us to turn the problem into an unconstrained optimization of an augmented loss function. Moreover, it allows us to not have to choose from beforehand which variables to restrict.

Exercise 4.7.

(a) Let R be the random variable $\mathrm{e}(\bar{g}(x), y)$ where x is chosen randomly. Therefore, $\mathrm{Var}[R] = \sigma^2(\bar{g})$. Note that $\sigma^2_{\text{val}} = \mathrm{Var}\left[\frac{1}{K}(R_1 + R_2 + \cdots + R_k)\right]$. Therefore, $\sigma^2_{\text{val}} = \frac{1}{K^2} \cdot K \cdot \sigma^2(\bar{g}) = \frac{1}{K}\sigma^2(\bar{g})$.

(b) Let R be the random variable $\mathrm{e}(\bar{g}(x), y)$ where x is chosen randomly. Let $P[\bar{g}(x) = y] = p = E[R]$. Note that $\mathrm{Var}[R] = E[R^2] - E[R]^2$. Because R is 0 or 1 and $0^2 = 0$ and $1^2 = 1$, $E[R^2] = E[R] = p$. Thus, $\mathrm{Var}[R] = p - p^2$. Therefore, $\sigma^2_{\text{val}} = \frac{1}{K}(p - p^2)$.

(c) $\frac{1}{K}(p - p^2)$ is a quadratic in p. It is maximized at 0.25 when $p = 0.5$. Thus, $\sigma^2_{\text{val}} \leq \frac{1}{4K}$.

Exercise 4.8. E_m is not an unbiased estimate because you are choosing the model with the least error. However, we can bound the effect using the VC-bound for finite hypothesis sets.