Assignment 8

Alberto Mejia
RIN: 661514960

CSCI 4100 - Machine Learning from Data

October 27, 2019

1. (200) LFD Exercise 4.3

(a) With H being fixed, if we increase the complexity of f, then the deterministic noise will go up, giving the tendency of overfitting to go up.

(b) With f being fixed, if we decrease the complexity H, then the deterministic noise will decrease as will the tendency to overfit.

2. (200) LFD Exercise 4.5

(a) To obtain the constraint $\sum_{q=0}^{Q} w_q^2 \leq C$, let Γ be the identity matrix with dimension d+1, $\Gamma = I^{d+1}$, as it will . To obtain the constraint, we can do the following:

$$w^T \Gamma^T \Gamma w = w^T I^T I w = w^T w = \sum_{q=0}^{Q} w_q^2$$

(b) To obtain the constraint $(\sum_{q=0}^{Q} w_q^2 \leq C)^2 \leq C$, let Γ be the matrix of all ones, $\Gamma =$ [1,1,...,1].

$$w^T \Gamma^T \Gamma w = (w^T w)^2 \text{ -> } \Gamma \times w = (\sum_{q=0}^{Q} w_q^2)^2$$

3. (100) LFD Exercise 4.6

For a binary classification using the perceptron model, I expect the hard-order constraint to be more useful. The soft order constraint would not have any effect and will not impact classification. The better option is to regularize with the hard order constraint (Hard order constraint can still produce an accurate classification for the model even if the model does not represent all the points in a data set).

4. (200) LFD Exercise 4.7

$$\sigma^2(g^-) = \mathrm{Var}_x[e(g^-(x), y)]$$

(a)

$$\begin{aligned}
\sigma^2_{val} &= \mathrm{Var}_{Dval}[E_{val}(g^-)] \\
&= \mathrm{Var}_{Dval}[\tfrac{1}{K} \textstyle\sum_{x \in D_{val}} e(g^-(x_n), y_n)] \\
&= \tfrac{1}{K^2} \textstyle\sum_{x \in D_{val}} \mathrm{Var}_{x_{Dval}}(g^-(x_n), y_n) \\
&= \tfrac{1}{K^2} K\, \sigma^2(g^-) \\
&= \tfrac{K}{K \times K} \sigma^2(g^-) \\
&= \tfrac{1}{K} \sigma^2(g^-)
\end{aligned}$$

(b)

$$\mathrm{Var}[E_{val}(g^-)] = \tfrac{1}{K} \sigma^2(y^-)$$

$e(g^-(x), y) = [g^-(x) \neq y]$
$e(g^-(x), y) = [g^-(x) \neq y]^2$ as it will not contribute to the variance when $g(x) = y$.
$\therefore \mathrm{Var}[E_{val}(g^-)] = \frac{1}{K}P[g^-(x) \neq y]^2$

(c)
$\mathrm{Var}[E_{val}(g^-)] = \frac{1}{K}P[g^-(x) \neq y]^2$
$= \frac{1}{K}P[\frac{1}{2}]^2$
$= = \frac{1}{K} \times (\frac{1}{2} \times \frac{1}{2})$
$= \frac{1}{K} \times \frac{1}{4}$
$= \frac{1}{4K}$
$\mathrm{Var}[E_{val}(g^-)] \leq \frac{1}{4K}$

(d) No; the square error is unbounded, the variance of $[E_{val}(g^-)]$ is infinite and therefore unbounded.

(e) I expect it to be higher because training with fewer points gives a higher probability of overfitting (A small dataset typically overfits the data as we continue to try and generalize the data). Not only that, our out of sample error should also increase, giving a worse approximation to the target function.

(f) Increasing the size of the validation set will increase k could give us a lower out of sample error. However, we would need the sweet spot of training vs validation because if we increase it too much it will decrease the size of our training set, which could give us a worse g or worse out of sample error.

5. (100) LFD Exercise 4.8

Yes, it is an unbiased estimate as g_m^- is not influenced by the data, it only depends on the training set.

6. (200) LFD Problem 4.26 (Graduate Only)