CSCI 4100 Fall 2018
Assignment 8 Answers

Damin Xu
661679187

October 29, 2018

Exercise 4.3

(a) As the complexity of f increase, and H is unchanged, the deterministic noise increases, and the tendency to overfitting will decrease.

(b) As the complexity of H decrease, and f is unchanged, the deterministic noise decrease, and because H is going to be more complex than f the tendency to overfitting will increase.

Exercise 4.5

(a) We already know that $\sum_{q=0}^{Q} w_q^2 = w^T w$, so we need to let $w^T \Gamma^T \Gamma w = w^T w$, because $I^T = I$ and $w \times I = I \times w = w$, so $\Gamma = I$.

(b) Here we want $w^T \Gamma^T \Gamma w = (w^T w)^2$, than $\Gamma \times w = \begin{bmatrix} \sum_{q=0}^{Q} w_q^2 \\ 0 \\ ... \\ 0 \end{bmatrix}$, So $\Gamma = \begin{bmatrix} 1 & 1 & ... & 1 \\ 0 & 0 & ... & 0 \\ & ... & ... & \\ 0 & 0 & 0 & 0 \end{bmatrix}$

Exercise 4.6
The hard-order constraint is be better than the soft-order constraint for binary classification using perceptron model.

The hard-order constraint count some weight as 0, so this must have some effect on E_{in}.

For the soft-order constraint, it constraint $w^T w \leq C$, so $E_{in}(w) = \sum_{i=1}^{n}[sign(w^T x_i) \neq y_i] + w^T w, wherew^T w \leq C$.

With out constraint, $E'_{in}(w) = \sum_{i=1}^{n}[sign(w^T x_i) \neq y_i]$, and let the best weight for this problem be w', than $E_{in}(w') = min(E_{in}(w))$ and $E'_{in}(w') = min(E'_{in}(w))$.

So, for a constant a, $E_{in}(aw') = E'in(aw') + (aw')T(aw') = E'_{in}(w') + ||aw'||^2$.

Clearly, when a is very small, $E_{in}(aw') \approx E'_{in}(w')$, and soft-order constraint has no effect.

Therefore, the hard-order constraint is better.

Exercise 4.7

(a) Because $E_{val}(g^-) = \frac{1}{K}\sum_{x_n \in D_{val}} e(g^-(x_n), y_n)$,

$$\begin{aligned}
\sigma_{val}^2 &= VAR_{D_{train}}[E_{val}(g^-)] \\
&= VAR_{D_{train}}[\frac{1}{K}\sum_{x_n \in D_{val}} e(g^-(x_n), y_n)] \\
&= \frac{1}{K^2}\sum_{x_n \in D_{val}} VAR_{x_n} e(g^-(x_n), y_N) \\
&= \frac{1}{K^2} K\sigma^2(g^-) \\
&= \frac{1}{K}\sigma^2(g^-)
\end{aligned}$$

(b) If $g^-(x) = y, e(g^-(x), y) = 1$, else $e(g^-(x), y) = 0$.
So $E[x][e(g^-(x), y)] = P[g^-(x) \neq y]$.

$$\begin{aligned}
\sigma_{g^-}^2 &= VAR_x[e(g^-(x), y)] \\
&= E[x][(e(g^-(x), y) - E[x][e(g^-(x), y)])^2] \\
&= E[x][(e(g^-(x), y) - P[g^-(x) \neq y])^2] \\
&= (1 - P[g^-(x) \neq y]) \times P[g^-(x) \neq y]
\end{aligned}$$

(c) From part (a) and (b), $\sigma_{val}^2 = \frac{1}{K}(1 - P[g^-(x) \neq y]) \times P[g^-(x) \neq y]$.
Because $(1-x)x \leq \frac{1}{4}$, so

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

(d) There is not any upper bound for $Var[E_{val}(g^-)]$

(e) $\sigma^2(g^-)$ is expected to be higher.

(f) In some range, the increase of K decreases E_{out}, and out of the range, increase of K increases E_{out}.

Exercise 4.8
Because $E_m = E_{val}(g_m^-)$,

$$E_{D_{val}}(E_m) = E_{D_{val}}(E_{val}(g-_m)) = E_{out}(g-_m)$$

So E_m is an unbiased estimate for $E_{out}(g_m^-)$