Homework 2

Tuesday, February 23, 2021 1:41 PM

Problem 2.1

a) let S = training data from D

f_S = the best performing algorithm

$$h_S = \arg\min_{f \in F} l_S(f)$$

$$L(f_{bayes}) = 0 \quad (\text{realizability})$$

$$L_D(f_S) - L(f_{bayes}) = (L_D(f_S) - \cancel{L_D(f_f)}^{0}) + (\cancel{L_S}^{0}(f_S) - \cancel{L_D(f_{Bayes})}^{0})$$

$$\epsilon = L_D(f_S)$$

PAC learnable Condition

$$P[L_D(f_S) \leq \epsilon] \geq 1 - \delta$$

$$1 \geq 1 - \delta \ ✓$$

PAC learnable given agnostic PAC + realizability

b) $$L_D(f_A) \leq \arg\min_{f \in F} L_D(f) + \epsilon$$

Since $f_{Bayes} \in F \rightarrow \arg\min_{f \in F} L_D(f) = 0$

$L(f_A) \leq \epsilon$ $\therefore$ it is a successful pac learner

$> 1-\delta$

Problem 2.2

a) $E[\ell_D(h_{Best})] \leq E[L_D(h_S)]$

any function's loss is expected to be lower bounded by the best performing algorithm

Since $\ell_D(h_{Best})$ is a constant

$\ell_D(h_{Best}) \leq E[L_D(h_S)]$

$h_S = \arg\min_{h \subset F} L_S(h)$

h_S is the best possible performance on the sampled data i any other function (including h_{Best}) would be lower bounded by h_S on the training data.

$$E[L_S(h_S)] \leq E[L_S(h_{Best})]$$

$$E[L_S(h_S)] \leq L_S(h_{Best})$$

$$L_S(h_{Best}) = L_D(h_{Best})$$

$$E[L_S(h_S)] \leq L_D(h_{Best}) \leq E[L_D(h_S)]$$

b) $$2\exp\left(\frac{2\epsilon^2}{\sum c_i^2}\right) \geq 1-\delta$$

$$\ln\left(2\exp\left(\frac{2\epsilon^2}{\ldots}\right)\right) \geq \ln(1-\delta)$$

$$\ln\left(2\exp\left(\frac{-2\epsilon^2}{\sum_i c_i^2}\right)\right) \geq \ln(1-\delta)$$

$$\frac{-2\epsilon^2}{\sum_i c_i} \ln 2 \geq \ln(1) - \ln(\delta)$$

$$\frac{-2\epsilon^2}{\sum_i c_i} \geq \frac{\ln(1) - \ln\delta}{\ln 2}$$

$$-2\epsilon^2 \geq \frac{\ln(1/\delta)}{\frac{\ln 2}{c_i} m}$$

$$\epsilon \leq \sqrt{\frac{\ln(1/\delta)}{\frac{\ln 2}{2c_i} m}}$$

c) Confidence increases with a decrease in c_i and m.

Problem 2 ... $\bar{w} = \begin{bmatrix} b \\ w \end{bmatrix}$

Problem 2.3

let $\overline{w} = \begin{bmatrix} b \\ w \end{bmatrix}$

a) $\min \sum (w^T x_i + b - y_i)^2 + \lambda \|w\|^2 + \epsilon b^2$

$x^T x w + b - y_i^2 + \lambda w^T (\lambda I) w + b^T \epsilon b$

$+ \overline{w}^T Q \overline{w}$

$\nabla_w (x^T x w + b - y_i)^2 + \overline{w}^T Q w$

$x^T x \overline{w} - x^T y + Q \overline{w} = 0$

$x^T x \overline{w} + Q \overline{w} = x^T y$

$(x^T x + Q) \overline{w} = x^T y$

$C \overline{w} = x^T y$

e) the indirect RLS is computationally slower and slows down a M. It has the benefit of not requiring a direct calculation of C^{-1}, which may or may not be possible