ECE 2372 Midterm

Avery Peiffer

Note: I'm still sick so apologies if my handwriting isn't great

1. (a) The VC bound is a generalization of the Hoeffding Inequality

In VC, many hypotheses are performing (virtually) the same, while Hoeffding checks all hypotheses

(b) When $\Gamma = \sqrt{\lambda} I$, $\hat{\theta} = (A^T A + \lambda I)^{-1} A^T y$

Tikhonov regularization shrinks the least squares estimate to the origin (reducing coefficient values)

(c) It is the plane that provides the maximum distance to the closest points of either class (i.e. halfway between for 2 classes), so it can generalize somewhat well

(d) The Kernel trick lets us map the data to a higher dimensional feature space, where it is linearly separable (easier to get a separating hyperplane)

2. $g_0(x) = \frac{1}{2} e^{-|x|}$ $\quad g_1(x) = \frac{1}{2} e^{-|x-1|}$ $\quad P[y=0] = P[y=1] = \frac{1}{2}$

$$\frac{g_1(x)}{g_0(x)} \underset{1}{\overset{0}{\lessgtr}} \frac{\pi_0}{\pi_1} \rightarrow \frac{\frac{1}{2} e^{-|x-1|}}{\frac{1}{2} e^{-|x|}} \underset{1}{\overset{0}{\lessgtr}} 1$$

$$e^{|x| - |x-1|} \underset{1}{\overset{0}{\lessgtr}} 1 \rightarrow \text{classification Rule}$$

$$\text{Risk} = \frac{1}{2} \int_{-\infty}^{\frac{1}{2}} g_1(x)\, dx + \frac{1}{2} \int_{\frac{1}{2}}^{\infty} g_0(x) = \Phi(-\tfrac{1}{2})$$

3. $h(x) = \begin{cases} +1 & \|x - c\| \le r \\ -1 & \text{otherwise} \end{cases}$ "+1 inside circle, -1 outside"

Linear Classifier: $d_{VC} = 3$

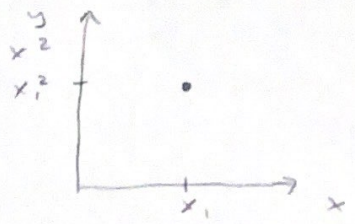
Must be at least this big because $m_H(3) = 8$ (shatters set) → can draw out configurations for possible labelings

$m_H(4) < 16$ → can't reach 16 dichotomies because of when a point is inside a triangle

4. $y = x^2$

$x \in [0, 1]$ uniformly

$D = \{(x_1, x_1^2)\}$ $\quad h(x) = ax$



Since there is no offset, we only care about a, and we will pick the line that goes through (x_1, x_1^2)

(a) $\bar{h}(x) = E_D[h_D(x)]$

$h_D(x)$ = function we pick

$a = \frac{x_1^2 - 0}{x_1 - 0}$ = slope of line $= x_1$

$\bar{h}(x) = x_1 \cdot x$

(b) $E_x[(\bar{h}(x) - x^2)^2]$

$= E[(x_1 \cdot x - x^2)^2] = E[x^4 - 2x_1 x^3 + x_1^2 x^2]$

$= E[x^4] - 2x_1 E[x^3] + x_1^2 E[x^2]$

x is drawn uniformly from 0 to 1

$E[x] = 0.5 = \int_0^1 x\, dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2} - 0 = \frac{1}{2}$

$E[x^2] = \int_0^1 x^2\, dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3}$

$E[x^3] = \int_0^1 x^3\, dx = \frac{x^4}{4}\Big|_0^1 = \frac{1}{4}$

$E[x^4] = \int_0^1 x^4\, dx = \frac{x^5}{5}\Big|_0^1 = \frac{1}{5}$

$= \frac{1}{5} - 2x_1\left(\frac{1}{4}\right) + x_1^2\left(\frac{1}{3}\right)$

$= \frac{x_1^2}{3} - \frac{x_1}{2} + \frac{1}{5}$

(c) $E_x[E_d[(h_D(x) - \bar{h}(x))^2]]$

$= E_x[E_d[(h_D(x) - x_1 x)^2]]$

$= E_x[E_d[h_D(x)^2 - x_1 x h_D(x) + x_1^2 x^2]]$

$= E_x[E_d[h_D(x)^2] - x_1 x E_D[h_D(x)] + x_1^2 x^2]$

$= E_x[E_d[h_D(x)^2] - x_1 x (x_1 x) + x_1^2 x^2]$

$= E_x\left[\int_0^1 x_1 x\, dx\right] = x_1 \cdot E_x\left[\frac{x^2}{2}\Big|_0^1\right] = \frac{x_1}{2}$

5. X_1, X_2 uniform on $[0, 1]$

X_1 and X_2 are estimates of probability of error of h_1 and h_2

use h_1 if $X_1 < X_2$, h_2 if $X_2 < X_1$

(a) $E[X_1]$ = average error of h_1

$= E[1 - acc_{h_1}]$ $\quad acc_{h_1} = \frac{1}{n}\sum_{i=1}^{n} 1_{h_1(x_i) = y_i(i)}$

$= E\left[1 - \frac{1}{n}\sum_{i=1}^{n} 1_{h_1(x_i) = y_i(i)}\right]$

$= \int_0^1 1 - \frac{1}{n}\sum_{i=1}^{n} 1_{h_1(x_i) = y_i(i)} \, dx$

$= 1 - \frac{1}{n}\int_0^1 \sum_{i=1}^{n} 1_{h_1(x_i) = y_i(i)} \, dx$

$= 1 - \frac{1}{n}\sum_{i=1}^{n} 1_{h_1(x_i) = y_i(i)}$

similarly, $E[X_2]$

$= 1 - \frac{1}{n}\sum_{i=1}^{n} 1_{h_2(x_i) = y_i(i)}$

(b) $x^* = \min(X_1, X_2)$ $\quad P[x^* \le x]$ where $x \in [0, 1]$ (arbitrary)

x^* = probability of error of better classifier

$P[x^* \le x] = 1 - P[x^* > x] = 1 - 2(2)\, e^{-2\varepsilon^2 n}$

$= 1 - 4e^{-2\varepsilon^2 n}$

(c) $E[x^*] = \int pdf = cdf = P[x^* \le x] = 1 - 4e^{-2\varepsilon^2 n}$

(d) No - the hypothesis set is not rich enough to reliably use x^* to predict how well the chosen classifier will perform in the future.