HW 6.

↳ Hardik Prajapati (2678294168)

1. $\underline{\underline{\tilde{C}}}_D \rightarrow$ in-sample confusion matrix.

$(\underline{\underline{\tilde{C}}}_D)_{ij}$ = [number of data points labelled $y=j$ that were misclassified as $h=i$]

$(\underline{\underline{C}}_{out})_{ij} = P[(h=i) \text{ AND } (y=j)]$ be the ij^{th} entry of the out-of-sample confusion matrix $\underline{\underline{C}}_{out}$

a) Dataset D, Single hypothesis $h \in \{1, 2, \dots C\}$

Data points $= N$.

$\Rightarrow$ No. of Data points that were misclassified (n_{mis}) $= \sum_{i \neq j} (\underline{\underline{\tilde{C}}}_D)_{ij}$

$$\Rightarrow E_D(h) = \frac{n_{mis}}{N} = \frac{\sum_{i \neq j} (\underline{\underline{\tilde{C}}}_D)_{ij}}{N}$$

$\Rightarrow$ Total probability of error $P(h \neq y)$ $= \sum_{i \neq j} (\underline{\underline{C}}_{out})_{ij}$

(Side Note:- Events $(\underline{\underline{C}}_{out})_{ij}$ & $(\underline{\underline{C}}_{out})_{kl}$ are not mutually exclusive

Because

Single hypothesis h will hold only 1 integer value for each Data point & hence determining one cell of $(C_{out})_{ij}$ will also speak for $(C_{out})_{kl}$.

↳ $\mu = P[\text{incorrect classification}] = \sum_{i \neq j} (\underline{\underline{C}}_{out})_{ij}$

~~N~~

↳ ν = percent misclassified by h on D $= \dfrac{\sum_{i \neq j} (\hat{\underline{\underline{C}}}_D)_{ij}}{N}$

Hoeffding Inequality

$$P\left[\,|\nu - \mu| > \varepsilon\,\right] \leq 2e^{-2\varepsilon^2 N} \quad \text{for any } \varepsilon > 0$$

Now, we know that

$$E_{out}(h) = P\left[h(\underline{x}) \neq f(\underline{x})\right] = \mu$$

$E_{in}(h)$ = % of points in D that are misclassified by h $= \nu$

$$\Rightarrow \quad P\left[\,|E_D(h) - E_{out}(h)| > \varepsilon\,\right] \leq 2e^{-2\varepsilon^2 N}$$

Let $\delta = 2e^{-2\varepsilon^2 N}$

$$\Rightarrow \ln\left(\frac{\delta}{2}\right) = -2\varepsilon^2 N$$

$$\Rightarrow \frac{1}{2N} \ln\left(\frac{2}{\delta}\right) = \varepsilon^2$$

$$\Rightarrow \varepsilon = \sqrt{\frac{1}{2N} \ln\left(\frac{2}{\delta}\right)}$$

$$\Rightarrow \quad P\left[|E_{out}(h) - E_D(h)| \leq \sqrt{\frac{1}{2N} \ln\left(\frac{2}{\delta}\right)} \right] \geq 1-\delta$$

$$\Rightarrow \quad P\left[E_{out}(h) \leq E_D(h) + \sqrt{\frac{1}{2N} \ln\left(\frac{2}{\delta}\right)} \right] \geq 1-\delta$$

→ Yes, This is a generalization Error bound for test-set Error for $C > 2$ class problem.

→ As there is only a single hypothesis set, we can use $|M| = 1$ & don't need to determine growth function & VC dim of the multi-class hypothesis h.

→ This is a tighter bound.

b) Validation-Set Error on D_{val}

$|H| = M \quad , \quad 0 < M < \infty$

$$\Rightarrow \quad P\left[E_{out}(h) \leq E_{D_{val}}(h) + \sqrt{\frac{1}{2N} \ln\left(\frac{2M}{\delta}\right)} \right] \geq 1-\delta$$

2. d. ii

Case :- $\alpha = 0.5$

$$\varepsilon_{\alpha\beta}(\alpha=0.5) = 2(1-0.5)\left[0+\frac{1}{2}(0.1)\right] + 4\sqrt{\frac{(0.5)^2}{\beta}+\frac{(0.5)^2}{1-\beta}}\sqrt{\frac{2}{N}(10)\ln\left[2(N+1)\right]+\frac{2}{N}\ln\left(\frac{8}{\delta}\right)}$$

$$\Rightarrow \boxed{\varepsilon_{\alpha\beta}(\alpha=0.5) = 0.05 + (4)(0.5)\sqrt{\frac{1}{\beta}+\frac{1}{1-\beta}}\,(K)}$$

where $K = \sqrt{\frac{2}{N}(10)\ln\left[2(N+1)\right]+\frac{2}{N}\ln\left(\frac{8}{\delta}\right)}$

$$\varepsilon_{\alpha\beta}(\alpha=\beta) = 2(1-\beta)\left[0+\frac{1}{2}(0.1)\right] + 4\sqrt{\frac{\beta^2}{\beta}+\frac{(1-\beta)^2}{1-\beta}}\,(K)$$

$$\Rightarrow \boxed{\varepsilon_{\alpha\beta}(\alpha=\beta) = 0.1(1-\beta) + 4\sqrt{1}\,(K)}$$

Now, let $\varepsilon_{\alpha\beta}(\alpha = \beta) < \varepsilon_{\alpha\beta}(\alpha = 0.5)$

$$\Rightarrow 0.1(1-\beta) + 4K < 0.05 + 2\sqrt{\frac{1}{\beta} + \frac{1}{1-\beta}}\, K$$

Simplifying & just comparing the two terms individually, we get

$$\hookrightarrow 0.1(1-\beta) < 0.05$$
$$\Rightarrow 1-\beta < 0.5$$
$$\Rightarrow 0.5 < \beta \qquad - (1)$$

$$\hookrightarrow 4k < 2\sqrt{\frac{1}{\beta} + \frac{1}{1-\beta}}\, K$$
$$\Rightarrow 2 < \sqrt{\frac{1}{\beta} + \frac{1}{1-\beta}}$$

$$\Rightarrow \beta \in (0, 1/2) \cup (\tfrac{1}{2}, 1) \qquad - (2)$$

$\Rightarrow$ If $\beta > 1/2$, then surely,

$$\varepsilon_{\alpha\beta}(\alpha = \beta) < \varepsilon_{\alpha\beta}(\alpha = 0.5)$$

Q3. a) Covariate shift :-

$$P_S(y|\underline{x}) = P_T(y|\underline{x}) \quad , \quad P_S(y) = P_T(y) \quad , \quad P_S(x|y) = P_T(x|y)$$

For covariate shift :- $P_S(\underline{x}) \neq P_T(\underline{x})$

↳ $P_T(x,y) = P_T(x|y)\,P_T(y) = P_T(y|x)\,P_T(x)$

Similarly,

$$P_S(x,y) = P_S(x|y)\,P_S(y) = P_S(y|x)\,P_S(x)$$

Now, if $P_S(y) = P_T(y)$ & $P_S(x|y) = P_T(x|y)$

Then,

$$P_S(x|y)\,P_S(y) = P_T(x|y)\,P_T(y)$$

$$\Rightarrow \quad P_S(x,y) = P_T(x,y)$$

$$\Rightarrow \quad P_S(y|x)\,P_S(x) = P_T(y|x)\,P_T(x)$$

Now, if $P_S(y|x) = P_T(y|x)$

$$\Rightarrow \quad P_S(x) = P_T(x)$$

$\Rightarrow$ No covariate shift.

Ans:- Not possible

b. Covariate Shift:-

Given:- $P_S(y|x) = P_T(y|x)$

For covariate shift:- $P_S(x) \neq P_T(x)$

Now, $P_S(x,y) = P_S(y|x)\,P_S(x) = P_S(x|y)\,P_S(y)$

$P_T(x,y) = P_T(y|x)\,P_T(x) = P_T(x|y)\,P_T(y)$

Now,

if $P_S(x,y) \neq P_T(x,y)$

$\Rightarrow P_S(y|x)\,P_S(x) \neq P_T(y|x)\,P_T(x)$

Given, $P_S(y|x) = P_T(y|x)$

$\Rightarrow P_S(x) \neq P_T(x)$

Hence, Covariate shift possible.

Ans:- Yes

eg. Sampling bias:-
- Poll prediction from people who had telephone. (Source Domain)
- Actual polls from unbiased domain (Target)

c. Concept Shift :-

Given :- $P_S(y) = P_T(y)$ & $P_S(x) = P_T(x)$

For concept shift :- $P_T(y|x) \neq P_S(y|x)$

Now, $P_S(x,y) = P_S(x|y)\, P_S(y) = P_S(y|x)\, P_S(x)$

$$P_T(x,y) = P_T(x|y)\, P_T(y) = P_T(y|x)\, P_T(x)$$

If $$P_S(x,y) \neq P_T(x,y)$$

$$\Rightarrow P_S(y|x) \cdot P_S(x) \neq P_T(y|x) \cdot P_T(x)$$

Given, $P_S(x) = P_T(x)$

$$\Rightarrow P_S(y|x) \neq P_T(y|x)$$

$\therefore$ Here, Concept Shift possible.

Ans :- Yes.

eg. $x \in$ {Months of year}

$y_S \in$ {Summer, Winter, Monsoon} $\quad y_T \in$ {Summer, Winter, Monsoon}

→ Each season is equally probable (4 months of year)

→ India :- Summer :- March - June

Monsoon :- July - October

Winter :- November - February.

→ In XYZ country, this schedule is different.