1

(a) $\dfrac{p(x|1)}{p(x|-1)} \underset{-1}{\overset{1}{\gtrless}} \dfrac{\pi_{-1}}{\pi_1}$

$\Rightarrow \dfrac{p(x|1)}{p(x|-1)} \underset{-1}{\overset{1}{\gtrless}} \dfrac{1}{4}$

$\Rightarrow \Lambda > \frac{1}{4}$ for $x=0, x=1, x=3$

$\Lambda < \frac{1}{4}$ for $x=2$

$\therefore$ choose H_1 ($\Lambda > \frac{1}{4}$) for $\underline{x=0, x=1, x=3}$
and choose H_{-1} ($\Lambda < \frac{1}{4}$) for $\underline{x=2}$

(b) $p_e = \pi_{-1} \cdot P(\text{choose } 1 \mid -1) + \pi_1 \cdot P(\text{choose } -1 \mid 1)$

$p_e = (0.2) \cdot [p(x=0|-1) + p(x=1|-1) + p(x=3|-1)]$
$+ (0.8) \cdot [p(x=2|1)]$

$p_e = (0.2) \cdot [0.1 + 0.3 + 0.1] + (0.8) \cdot [0.1]$

$p_e = 0.2 \times 0.5 + 0.08$

$p_e = 0.18$

$\boxed{\therefore p_e = 0.18}$

(2)

$$g_{12}(x_1=1, x_2=5) = 1+5-2 = 4 > 0$$

$$g_{13}(x_1=1, x_2=5) = 2\cdot1-5-3 = -6 < 0$$

$$g_{23}(x_1=1, x_2=5) = 1-2\cdot5-1 = -10 < 0$$

$$\Rightarrow g_{12}(\underline{x}) > 0 \quad \text{and} \quad g_{31}(\underline{x}) > 0 \quad \text{and} \quad g_{32}(\underline{x}) > 0$$

$$\boxed{\therefore\ \underline{x} \text{ is classified in class } \underline{y=3}}$$

$\left(\because g_{31}(\underline{x}) > 0 \text{ and } g_{32}(\underline{x}) > 0\right)$

(3)

$$\Lambda = p(1|\underline{x}) \underset{-1}{\overset{1}{\gtrless}} p(-1|\underline{x})$$

$$\Rightarrow \Lambda = \ln\left(\frac{p(1|\underline{x})}{p(-1|\underline{x})}\right) > 0$$

$$\Rightarrow \ln\left(\frac{p(1|\underline{x})}{p(-1|\underline{x})}\right) = \beta_0 + \underline{\beta}^T \underline{x}.$$

$\therefore \Lambda = \beta_0 + \underline{\beta}^T \underline{x}$. The classifier chooses $y=1$ when $\beta_0 + \underline{\beta}^T \underline{x} \geq 0$ and $y=0$ when $\beta_0 + \underline{\beta}^T \underline{x} < 0$. Therefore, the logistic regression model is a <u>linear</u> classifier.

④

ⓐ **Prediction error** is the average error of a classifier f trained on training data τ on every possible input data $(\tilde{x}, \tilde{y})$.

$$err_{pred} = E\left[L(f(\tilde{x}), \tilde{y}) \mid f \text{ trained on } \tau \right]$$

Validation error is the average error of a classifier f trained on training data τ on an independent data set called validation data ν. Since, the classifier has not been trained on ν, it is an estimate for prediction error y_{pred} (since y_{pred} is difficult to compute)

$$err_{val} = \hat{err}_{pred} = \frac{\sum L(f(\tilde{x}), \tilde{y})}{|\nu|}$$

where $(\tilde{x}, \tilde{y})$ are data in ν and $|\nu|$ is the cardinality of ν.

(b) Eventhough Bayes classifier minimizes prediction error, bayes classifier assumes that the priors (π_y) and the probability distribution ($p(x,y)$) is available. But in supervised learning, π_y and $p(x,y)$ are not available and we need to use training data for a classifier to work. Therefore, we cannot use it in supervised learning.

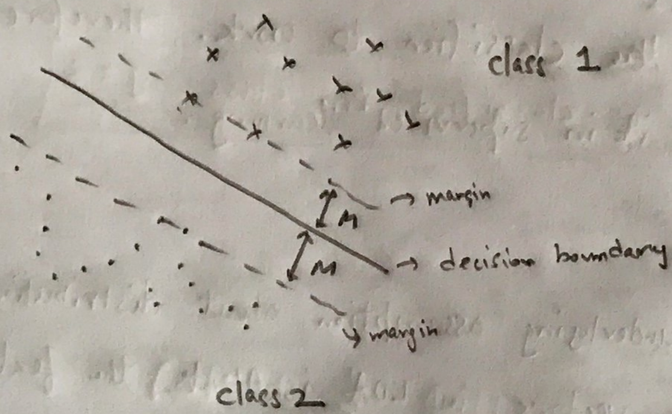
(c) The underlying assumption about distribution of the feature vectors in LDA is that the feature vectors under each class are gaussian distributions and can have varying means but should have same covariance.

(d) Binary SVM for linearly seperable data ~~maximizes~~ margin.

Margin is the set of points that are within a distance M

where M is the distance from decision boundary to the

nearest feature vector.



(e) Naive Bayes uses the assumption of conditional independence over classes, i.e.,

$$p(x|y) = p(x_1|y) \cdot p(x_2|y) \cdot \ldots \cdot p(x_k|y)$$

The advantage for Naive Bayes is that $p(x_1|y), p(x_2|y), \ldots,$ $p(x_k|y)$ are easier to compute than directly computing $p(x|y)$.

(f) Model Selection is choosing the 'best' classifier of all available classifier based on the performance of classifier on validation data.

Model Assessment is assessing how well the chosen classifier (in Model selection) generalizes well on unseen data (based on the performance on test data).