Quiz 1

(1)

(a) If $\frac{P(x|1)}{P(x|-1)} > \frac{\pi_{-1}}{\pi_1}$, then

chose $+1$, else -1.

$\Rightarrow$

$$\frac{P(x|1)}{P(x|-1)} = \begin{cases} 2.5 & x=0 \\ 1.25 & x=1 \\ 0.6667 & x=2 \\ 0.75 & x=3 \end{cases}$$

$\therefore$ Decision Rule: chose $+1$ when $x=0$ and $x=1$

chose -1 when $x=2$ and $x=3$

(b) $P_e = \pi_{-1} \cdot P(H_1 | H_{-1}) + \pi_1 \cdot P(H_{-1} | H_1)$

$$P_e = \frac{1}{2} \cdot [0.1+0.2] + \frac{1}{2} \cdot [0.2+0.3]$$

$$= \frac{1}{2} \cdot \frac{3}{10} + \frac{1}{2} \cdot \frac{5}{10} = \frac{8}{20}$$

$$\therefore P_e = 0.4$$

(2)

(a) For each new data point, let f predict the label $\hat{y}$ as the same label y of the nearest neighbour to the new data point in the training data.

i.e., f is 1-NN classifier.

(to itself)

A 1-NN classifier fits itself, and thus, perfectly classifies the training data with training error equal to zero.

(2)
(b) Prediction error is the error produced by a classifier when it classifies previously unseen data. (over all possible training data sets)

$$Err_{pred} = \frac{\#\text{ of data points incorrectly classified in } \mathcal{V}}{\text{Total \# of data points in } \mathcal{V}}$$

for 0-1 loss.

(3)

$$g_{12}(\underline{x}) = 1 + 2 \cdot 2 - 3 = 2$$

$$g_{13}(\underline{x}) = 2 \cdot 1 + 2 - 1 = 3$$

$$g_{23}(\underline{x}) = 1 - 2 + 2 = 1$$

Since, $g_{12}(\underline{x}) > 0$ and $g_{13}(\underline{x}) > 0$

$\therefore$ $\underline{x}$ is classified as class $\underline{1}$

(4)

(a) Linear classifiers:

Logistic Regression
Linear Discriminant Analysis

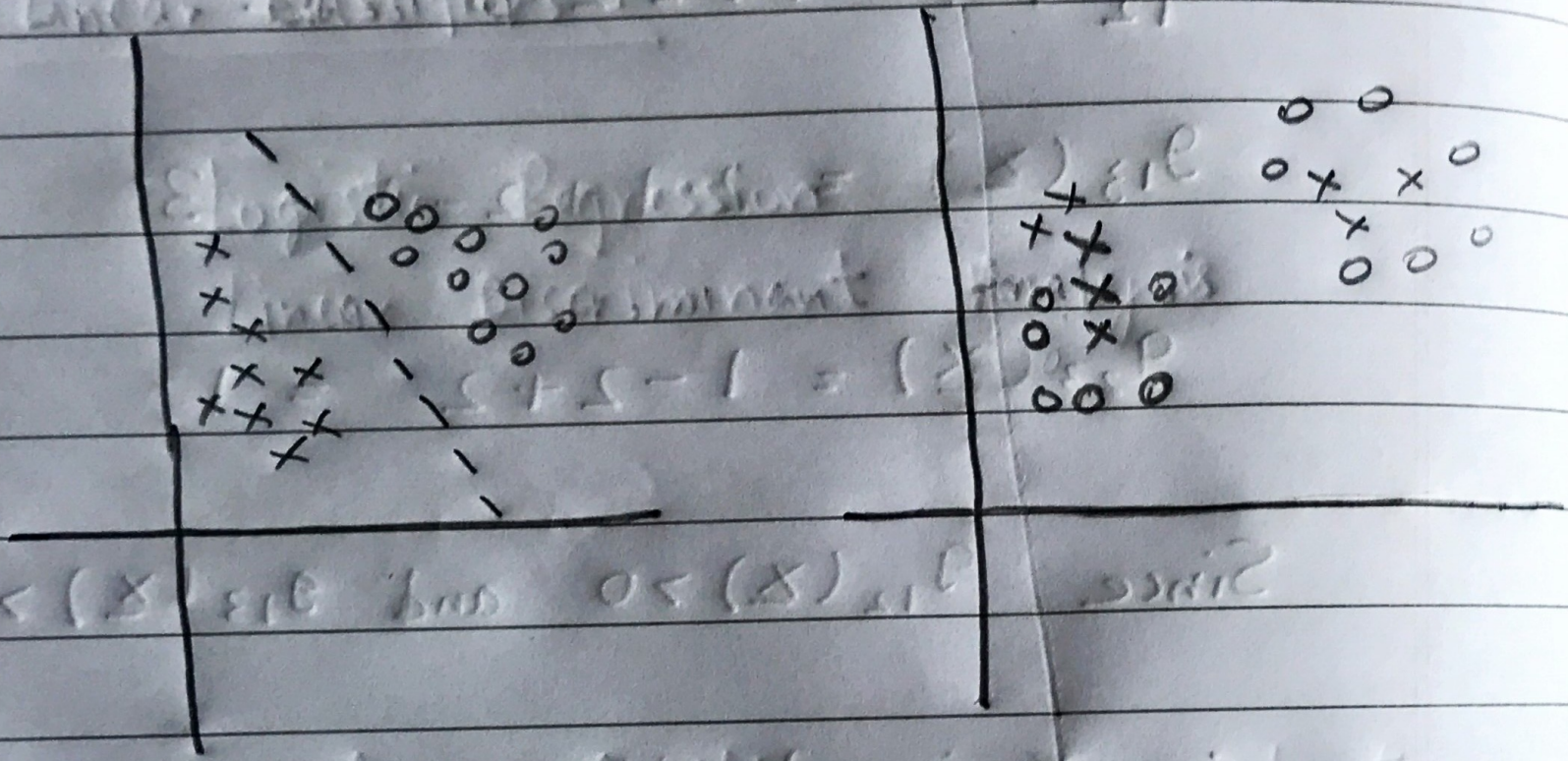
(b) Given a linear classifier, we can seperate data that is not linearly seperable by using the kernel trick.

We first write all the computations used in the linear classifier as dot products.

Then, we substitution all dot products, $x^T r$ with $K(x, r)$ where $K(\cdot)$ is the kernel function.

This helps us classify data in higher dimensions where the data is linearly seperable.

5



Linearly Seperable

Non-linearly Seperable

(6)

(a) Naive Bayes makes the assumption that the components of the feature vector $\underline{x}$ are independent, i.e.,

$$p(\underline{x}, y) = p(x_1|y) \cdot p(x_2|y) \cdot \ldots \cdot p(x_n|y)$$

The components of $\underline{x}$: $x_1, x_2, \ldots, x_n$ conditioned on y are independent.

6

(b) Discriminative model is a model which find $p(x|y)$ directly from the given data while, Generative model is a model which first finds $p(x,y)$ and finds $p(x|y)$ from $p(x,y)$.

Since Logistic Regression classifier directly finds $p(x|y)$ from the data, it is considered a discriminative model.