

Q1/ (a) Definition of generative & discriminative classifiers {0.5 m}
 $GMM \leftarrow$ generative
 $SVM \leftarrow$ discriminative } with justification {0.5 m}

(b) Training procedure SVM & GMM {1 m}

Yes, can able to handle (with justification) {1 m}

(c) Bernoulli distribution is preferred for binary data - (with appropriate justification) {1 m}

(d) Refer in Bishop (Pattern Recognition & Machine Learning book)
Page no. (441-445).

exⁿ 9.47, 9.48, 9.56, 9.57, 9.58
9.59, 9.60, 9.51 is
enough {2 m}

(e) $x = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$, $\lambda_1 = \{0.3, 0.6, 0.6, 0.6\}$
 $\lambda_2 = \{0.1, 0.8, 0.6, 0.1\}$

$$P(x/\lambda_1) = \prod_{i=1}^4 (\lambda_{1i})^{x_i} (1-\lambda_{1i})^{1-x_i} \quad \checkmark 0.75$$

$$= 0.7 \times 0.6 \times 0.4 \times 0.6 = 0.1008$$

$$P(x/\lambda_2) = \prod_{i=1}^4 (\lambda_{2i})^{x_i} (1-\lambda_{2i})^{1-x_i} \quad \checkmark 0.75$$

$$= 0.9 \times 0.8 \times 0.4 \times 0.9 \\ = 0.2592$$

As, priors are equiprobable,
and $P(x/\lambda_2) > P(x/\lambda_1)$
0.5

x is belongs to λ_2 class
2 m

2) (a)

We need dimensionality reduction :-

(i) When we deal with real data we often deal with high dimensional data that can go up to millions. But these dimension can be highly correlated in itself. Hence, we can reduce to lower dimension. This will speed up the computation.

(ii)

Pros

- i) Removes correlated features.
- ii) Reduces overfitting & improves visualization

Cons

- i) Information is lost.
- ii) Independent variables becomes less interpretable.

① marks

Steps to dimensionally reduce using PCA :-

1) Pre-processing:

Scale 64×64 image matrix into 1 column.

Mean ~~re~~ and variance normalization.

$$\Rightarrow \text{data} = \text{data} - \text{mean}(\text{data});$$

2) Construct covariance matrix

3) Decompose the covariance matrix into eigenvectors & eigenvalues

4) Sort the eigenvalues in decreasing order to rank corresponding eigenvectors.

5) Select k eigenvectors which correspond to the k -largest eigenvalues.

- 6) Construct a projection matrix W from "top" k eigenvectors.
- 7) Transform the d -dimensional input dataset X using the projection matrix W to obtain the new k -dimensional feature subspace.

— (1) marks

Steps of dimensionality reduction using LDA.

- 1) Compute d -dimensional mean vector for different classes from the dataset.
- 2) Compute the "in between class" & "within-class matrix".
- 3) Compute the eigenvectors (e_1, e_2, \dots, e_d) & corresponding eigenvalues ($\lambda_1, \lambda_2, \dots, \lambda_d$).
- 4) Sort the eigenvectors by decreasing eigenvalues and choose k -eigenvectors with the largest eigenvalues to form $d \times k$ dimensional matrix W .
- 5) Use this $d \times k$ eigenvector matrix to transform the samples onto the new subspace.

$$Y = X \times W$$

\swarrow \downarrow \searrow
 $n \times k$ $n \times d$ $d \times k$

$k \ll d$

— (1) marks

2)(b) Difference b/w LDA & PCA — (1) marks

Q To measure the performance we can apply the reverse transformation and measure the reconstruction error.

~~Also the performance can~~

If dimensionality reduction is used as preprocessing step before another ML algorithm, then we can simply measure the performance of the second algorithm. If dimensionality reduction did not lose too much information, the second algorithm should perform well.

Note:- Multiple solutions are allowed.

2)(c) -

— 1 marks

It is not possible to perfectly reverse the operation because some information gets lost during dimensionality reduction.

1 marks

4)

Multiple solutions are accepted.

Method — 1 marks

Training with block diagram — 2 marks

Testing with appropriate block diagram — 2 marks

Entropy $H(S) = -\sum_{i=1}^N p_i \log_2(p_i)$

Information Gain $IG(S, A) = H(S) - H(S|A)$

1] $H(D)$ (Discovered) $= -\left(\frac{9}{14}\right) \log_2\left(\frac{9}{14}\right) - \left(\frac{5}{14}\right) \log_2\left(\frac{5}{14}\right) = +0.94$

2] Entropy for Discovered for each condition of caught fever

$H(D|F = \text{Absent}) = -\frac{2}{4} \log_2\left(\frac{2}{4}\right) - \frac{2}{4} \log_2\left(\frac{2}{4}\right) = +1$

$H(D|F = \text{mild}) = -\frac{4}{6} \log_2\left(\frac{4}{6}\right) - \frac{2}{6} \log_2\left(\frac{2}{6}\right) = +0.92$

$H(D|F = \text{severe}) = -\frac{3}{4} \log_2\left(\frac{3}{4}\right) - \frac{1}{4} \log_2\left(\frac{1}{4}\right) = +0.81$

Weighted Entropy

$H(D|F) = \frac{4}{14} \times 1 + \frac{6}{14} \times 0.92 + \frac{4}{14} \times 0.81 = 0.91$

$IG(D, F) = 0.94 - 0.91 = 0.03$

3] Entropy for Discovered for each condition of caught

$H(D|C = \text{Absent}) = -\frac{2}{5} \log_2\left(\frac{2}{5}\right) - \frac{3}{5} \log_2\left(\frac{3}{5}\right) = 0.97$

$H(D|C = \text{mild}) = -\frac{4}{4} \log_2\left(\frac{4}{4}\right) - \frac{0}{4} \log_2\left(\frac{0}{4}\right) = 0$

$H(D|C = \text{severe}) = -\frac{3}{5} \log_2\left(\frac{3}{5}\right) - \frac{2}{5} \log_2\left(\frac{2}{5}\right) = 0.97$

Weighted entropy

$H(D|C) = \frac{5}{14} \times 0.97 + \frac{4}{14} \times 0 + \frac{5}{14} \times 0.97 = 0.69$

$IG(D, C) = 0.94 - 0.69 = 0.25$

$$I_4(D, J) = 0.152$$

$$I_4(D, G) = 0.048$$

Cough has largest entropy gain.
 \therefore forms the root node.

If Cough = Mild, Discard = True

\therefore No need to split further. (entropy = zero)

$$H(\text{Absent}) =$$

$$H(\text{Absent}) = -\left(\frac{2}{5}\right) \log\left(\frac{2}{5}\right) - \left(\frac{3}{5}\right) \log\left(\frac{3}{5}\right) = 0.97$$

$$H(\text{Cough} = \text{Absent} / \text{Fever}) = 0.4$$

$$I_4(\text{Cough} = \text{Absent} / \text{Fever}) = 0.571$$

$$I_4(\text{Cough} = \text{Absent} / \text{Travel}) = 0.971$$

$$I_4(\text{Cough} = \text{Absent} / \text{Gender}) = 0.020$$

split based on Travel.

