Relevant Repo-

https://github.com/rohinarora/EECE5644-Machine_Learning/blob/master/Hw2/

Answer 1

question! Some notation →

$$\lambda_{ij} = \begin{cases} 0 & i = j \\ \lambda_{r} & i = c+1 \\ \lambda_s & \text{else} \end{cases}$$

$i, j = 1$ to c

Truth

Decision

class	1	2	3	- - - - -	n
1	0	λ_s	λ_s		λ_s
2	λ_s	0	λ_s		λ_s
3	λ_s	λ_s	0		λ_s
. . .					
n					0
n+1	λ_r	λ_r	λ_r	- - - - -	λ_r

In general, to minimize risk →

compute $\underbrace{R(D = i \mid x)}_{\boxed{\text{cost associated in deciding label } i}} = \sum_{j=1}^{c} \underbrace{\lambda_{ij}}_{\text{cost}(i,j)} \underbrace{P(L = j \mid x)}_{\text{Posterior of Label } j}$

$$\text{decision/action} = \operatorname*{argmin}_{i} R(D = i \mid x)$$

For our problem, consider classes i and K.

Risk associated in deciding class i →

$$R(D=i|x) = \sum_{j=1}^{c} \lambda_{ij}\, P(L=j|x)$$

$$= \cancel{\lambda_s}\ \cancel{P(L=1)}\ \lambda_s \left[\sum_{\substack{j \in (\text{class labels}) \\ j \neq i}} P(L=j|x) \right] \text{——— (1)}$$

$\lambda_{ij} = 0$ when $j = i$ hence that term has to be ignored

$$\sum_{j=1}^{c} P(L=j|x) = 1$$

$$\therefore \sum_{\substack{j \in \text{class labels} \\ j \neq i}} P(L=j|x) = 1 - P[L=i|x]$$

$\therefore$ (1) becomes →

$$R(D=i|x) = \lambda_s \left[1 - P[L=i|x] \right] \text{—— (2)}$$

Risk associated with deciding class K →

$$R(D=K|x) = \lambda_s \left[1 - P(L=K|x)\right] \quad (3)$$

To decide between class i and class K →

~~R~~ Decide class i iff →

$$R(D=i|x) < R(D=K|x)$$

Rephrasing →

$$R(D=i|K) \underset{D=i}{\overset{D=K}{\gtrless}} R(D=K|x)$$

Put in the values of risk from equation (2) and (3)

$$\lambda_s \left[1 - P(L = i \mid x)\right] \lesseqgtr_{D=i}^{D=K} \lambda_s \left[1 - P(L = K \mid x)\right]$$

$$P(L = i \mid x) \lesseqgtr_{D=K}^{D=i} P(L = K \mid x)$$

$\Rightarrow$ Decide class i if $P(L = i \mid x) \geqslant P(L = K \mid x)$
for all class $K = (1 - - - c)$; $K \neq i$

$\Rightarrow$ We still need to decide between class i and $\boxed{\text{class } (c+1)}$ $\rightarrow$ reject class.

Risk associated with reject class $\rightarrow$

$$R\left[D = c+1 \mid x\right] = \sum_{j=1}^{c} \lambda_{ij} \, P(L = j \mid x)$$

$$= \lambda_r \sum_{j=1}^{c} P(L = j \mid x)$$

$$= \lambda_r \quad \text{———} \quad (4)$$

Rewriting equation 2 →

$$R[D=i \mid x] = \lambda_s [1 - P[L=i \mid x]])$$

Decision rule → minimize risk associated with decision.

∴ decide class i iff →

$$R(D=i \mid x) \leq R(D=c+1 \mid x)$$

$$\lambda_s [1 - P(L=i \mid x)] \leq \lambda_R$$

$$1 - P(L=i \mid x) \leq \frac{\lambda_R}{\lambda_s}$$

$$\Rightarrow \boxed{P(L=i \mid x) \geq \left(1 - \frac{\lambda_R}{\lambda_s}\right)} \quad (5)$$

→ If $P(L=i \mid x) < \left(1 - \frac{\lambda_R}{\lambda_s}\right)$;

choose reject class $(c+1)$.

QED

1. $\lambda_{r} = 0$

consider equation 5 →

$$P(L=i \mid x) \geqslant 1 - \frac{0}{\lambda_s}$$

$$\Rightarrow P(L=i \mid x) \geqslant 1$$

This means decide class i only if posterior $\geqslant 1$.

→ This is very unlikely.

→ since $\lambda_{r} = 0$; cost associated with reject class is 0. Hence decision rule will always decide the reject class.

(2) $\lambda_r > \lambda_s$

consider equation 5 $\rightarrow$

Right hand side of equation

$$P(L = i \mid x) \geq 1 - \frac{\lambda_r}{\lambda_s}$$

If $\lambda_r > \lambda_s$; RHS is less than 0.

$$P(L = i \mid x) \geq \underbrace{\boxed{1 - \frac{\lambda_R}{\lambda_s}}}_{< 0}$$

$\downarrow$

This is true for all probabilities.
$(P(x) > 0)$

$\rightarrow$ decision rule will always choose some class i

$\rightarrow$ It will never select reject class

$\rightarrow$ This is ~~unintuitive~~ intuitive as cost of reject class is high, & we wish to minimize risk.

Answer 2

Answer 2

→ generate random uniform samples. Use it to threshold a prior, and draw samples from class 1 or class 2.

→ class 1 and class 2 are gaussians. draw samples ~~with~~ from zero mean; I variance. Use linear transform $A\vec{x} + b$ to get desired class label i.

where $b = \mu_i$;

$A =$ cholesky (Σ_i) where μ_i & Σ_i are the desired mean variance of class i

$\sim N(0,1)$

$z_i = A x + b$

samples for class i

MAP classification rule:

$$d_{MAP} = \underset{i}{\arg\max}\ P(x|L=i)\,P(L=i)$$

In case of 2 classes →

Decide class 1 if $$\underbrace{P(x|L=1)}_{\text{likelihood}}\ \underbrace{P(L=1)}_{\text{prior of class 1}} > \underbrace{P(x|L=2)}_{\text{likelihood}}\ \underbrace{P(L=2)}_{\text{prior of class 2.}}$$

likelihood → gaussian for class 1

likelihood → gaussian of class 2

Else decide class 2.

↓

This shall be our decision rule.

Based on this, count the number of misclassifications for class 1 & 2.

$$P(\text{error}) = \frac{\text{misclassification of class 1 and 2}}{400}$$

400 → Total samples

Reference code for this question- Q2_Git_repo

Answer 3.

q3 from the samples, estimate μ & Σ of each class. We know these belong to gaussian.

via MLE, we can find same mean & sample cov.

Let class 1 $\sim N(\mu_1, \Sigma_1)$

class 2 $\sim N(\mu_2, \Sigma_2)$

applying LDA on these $\rightarrow$

$$\underbrace{S_W^{-1} S_B}_{\substack{\downarrow \\ \text{find the greatest eignvalue, eignvector of this matrix.}}} W = \underbrace{\left(\frac{W^T S_B W}{W^T S_W W}\right)}_{\text{eign value}} \cdot \underbrace{W}_{\text{eignvector.}}$$

where $S_W = \Sigma_1 + \Sigma_2$

$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$

Reference code for this question- Q3_Git_repo