Caneta

EN 2017

1. (a) They are orthogonal, we could prove this by calculating the dot product between them (it should be close to 0)

(b) We can calculate this w/ $R(\%) = \frac{\lambda_1^2}{\lambda_1^2+\lambda_2^2+\lambda_3^2} \times 100 \approx$ $\approx 95,5\%$

By only using one PC.

(c) Scaler: $f_{norm} = \frac{f-u}{\sigma} =$

$$= \left(\frac{146-324,033}{201,353} \quad \frac{238-740,327}{384.216} \quad \frac{8,63-8,68}{2,277} \right)^T \approx$$

$$\approx \begin{pmatrix} -0,884 \\ -1,308 \\ -0,022 \end{pmatrix}$$

$$\text{Output} = f_{norm}^T \, \omega_1 \approx \dots$$

$$\begin{pmatrix} -0,594 \\ -0,58 \\ -0,558 \end{pmatrix}$$

eigenvector associated w/ largest eigenvalue

(2)

(a) $$\Lambda(x) = \frac{p(x|\omega_1)}{p(x|\omega_2)} > \frac{P(\omega_2)}{P(\omega_1)}$$

if this condition is true, classify as ω_1, else ω_2

It compares the ratio of the likelihoods to the ratio of the priors, if evidence for ω_1 outweighs the priors against it (threshold), we decide ω_1

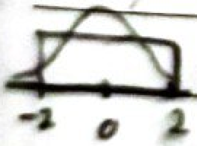
(b) $$p(x|\omega_1) = \mathcal{N}(0,1) = \frac{1}{\sqrt{2\pi}} e^{-(x^2/2)}$$

1- Priors are equal, so the threshold $P(\omega_2)/P(\omega_1) = 1$

2- So we just classify as ω_1 when $P(x|\omega_1) > P(x|\omega_2)$

$P(x|\omega_2) = 0$ outside the interval (classify as ω_1)
else, we need to find intersection points:

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} > \frac{1}{4} \Leftrightarrow e^{-\frac{x^2}{2}} > \frac{\sqrt{2\pi}}{4} \Leftrightarrow$$



$$\Leftrightarrow -\frac{x^2}{2} > \ln\left(\frac{\sqrt{2\pi}}{4}\right) \Leftrightarrow x^2 < -2\ln(0{,}626) \Leftrightarrow$$

$$\Leftrightarrow x^2 < 0{,}937 \Leftrightarrow |x| \leq \sqrt{0{,}937} \approx 0{,}968$$

Rule: ω_1 if $|x| < 0{,}968$ or $|x| > 2$ (outside interval)

ω_2 if $0{,}968 < |x| < 2$

③ RBF Kernel:

ⓐ $K(x_i, x) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right)$

σ controls width of Gaussian dist

Role of σ:

Radius of influence around SV where it significantly affects boundary

Large σ: Wide Gaussian ⇒ the influence of each support vector extends over large distance.
When summing the wide Kernels to create decision surface, the result is smooth & irregularities are average'd out.

Small σ: Narrow Gaussian ⇒ Influence of a SV is restricted to it's vicinity.

A → Large σ, smooth line, captures global trend
B → Small σ, complex line, captures local variations

ⓑ I'd choose A. It has better generalization.

④

ⓐ (i) Bias
(ii) Variance

ⓑ ① Probably bias, Fisher Discriminant and MDC are simple, linear models
② We could sample the set with replacement to create different sets. Then, we could see how the performance varies across sets.

⑤ Answered in ER2023