Dal punto C abbiamo definito la nostra decisione statistica come la distanza euclidea dal centro tramite i valori delle 2 PCA

$T = PCA_1^2 + PCA_2^2$ se questo valore supera una certa soglia allora sappiamo che il punto è abbastanza lontano dal centro $(Y=1)$

Nel nostro caso $T_i = x_i$

Possiamo ipotizzare che ogni classe, $(-1,1)$ segue una distribuzione gaussiana sulla variabile T

Abbiamo definito dalla teoria che il classificatore MAP

- C_1 $Y=-1$
- C_2 $Y=1$

$$\frac{P(C_1|X)}{P(C_2|X)} \xrightarrow{\text{Bayes}} \frac{\pi(C_1)\,\ell(X|C_1)}{\pi(C_2)\,\ell(X|C_2)} \underset{C_2}{\overset{C_1}{\gtrless}} 1$$

Ho suddiviso il Dataset TRAIN e TEST con i valori del PCA_1, PCA_2 sono sufficienti a spiegare la varianza dei dati x

Calcolando le prior delle due classi sul dataset Train

$$\pi(C_i) = \frac{N° \text{ dati in } Y}{N° \text{ dati } (Y = C_i)}$$

esce che sono simili per le due classi $(0{,}50)$ e $(0{,}49)$

Quindi il MAP diventa

$$\frac{\ell(x|C_1)}{\ell(x|C_2)} \gtrless \cancel{\frac{\pi(C_2)}{\pi(C_1)}}$$

Adesso si va a calcolare σ, μ

- $\ell(x, C_1) \sim N(\mu_1, \sigma_1^2)$ $\qquad \mu_1 = \frac{N° \text{ dati}}{N° \text{ dati } (Y=1)}$

m = Campioni per classe C_1 $\quad$ (Varianza) $= \sigma_1^2 = \frac{1}{m} \sum_{i=1}^{m} (T_i - \mu)^2$

$$\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{ -\frac{(T-\mu_1)^2}{2\sigma_1^2} \right\} = p(T|C_1)$$

- $p(x|C_2) \sim N(\mu_2, \sigma_2^2)$

$$p(T|C_2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left\{ -\frac{(T-\mu_2)^2}{2\sigma_2^2} \right\}$$

$$\frac{p(T|C_1)}{p(T|C_2)} \underset{C_2}{\overset{C_1}{\gtrless}} 1$$

Se il likelihood Ratio dell'osservazione T è >1 allora T è $y=1$, senno T è $y=-1$

il grafico delle prior è visibile sul codice