Discriminante lineal

* Método de clasificación supervizado.

* Dos o más grupos son conocidos a priori

* $P(Y=k \mid X=x)$

✓ Alternativa a la regresión logística cuándo hay 2 o más niveles.

✓ Clases bien separadas → Parámetros + estables

✓ Equivalente a logística cuando hay solo 2 niveles (Tiene el mismo desempeño).

Método:

1) Dividir los datos en entrenamiento en el que se conoce a que grupo pertenece cada observación.

2) Probabilidades a Priori (proporción "esperada" en cada grupo).

4) Determinar si la varianza o matriz de covarianzas es homogenea (Prueba de homogeneidad de varianzas entre las clases).
De ello depende LDA o QDA.

5) Estimar los parámetros necesarios para las funciones de probabilidad condicional.
(verificando las condiciones

6) Calcular el resultado de la función

6) Calcular el resultado de la función discriminante. (Indica dónde se asigna la observación)

7) Usar validación cruzada para estimar las probabilidades de clasificaciones erroneas.

Teorema de Bayes para Clasificación.

2 Eventos $\rightarrow$ A

$\searrow$ B

$$P(B|A) = \frac{P(A \cap B)}{P(B)}.$$

* Probabilidad a priori $P(Y=k) = \pi_k$

$k = 1, \ldots, C$ $\quad$ C: # clases

* $f_k(x) = P(X = x \mid Y = k)$

↓
función de densidad condicional de x para una observación que pertenece a la clase K.

* $f_k(x)$ grande → Mayor probabilidad de que la observación pertenezca a K.

* Probabilidad a posteriori:

$$P(Y=K \mid X=x)$$

La probabilidad de que una observación pertenezca a K siendo x el valor predictor.

$$P(Y=K \mid X=K) = \frac{\pi_K \, P(X=x \mid Y=K)}{\sum_{i=1}^{K} \pi_i \, P(X=x \mid Y=i)}$$

$$= \frac{\pi_k \cdot f_k(x)}{\sum_{i=1}^{k} \pi_i f_i(x)}$$

La clasificación con menor error se consigue asignando la observación que maximice la probabilidad a posteriori.

Como $\sum_{i=1}^{K} \pi_i f_i(x)$ es igual para todas las clases

$$P(y=k \mid x=k) \propto \pi_k f_k(x).$$

Se selecciona el mayor $\pi_k f_k(x)$

$\downarrow$

Estimarlos

$$\hat{\pi}_k = \frac{n_k}{N}$$

H_0: las observaciones se distribuyen

de Forma normal.

$$f_k(x) = P(Y=k \mid X=x)$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_k} \cdot e^{-\frac{1}{2\sigma_k^2}\{(x-\mu)^2\}}$$

1) Asumiendo σ_k cte.

$$P(Y=k \mid X=x) = \frac{\cancel{\frac{1}{\sqrt{2\pi}\,\sigma}} e^{-\frac{1}{2\sigma^2}\{(x-\mu_k)^2\}} \cdot \pi_k.}{\sum_{j=1}^{K} \pi_j \cancel{\frac{1}{\sqrt{2\pi}\,\sigma}} e^{-\frac{1}{2\sigma^2}\{(x_i-\mu_j)^2\}}}$$

$$\hat{\delta}_k = \log(Y=k \mid X=x)$$

$$= -\frac{1}{2\sigma^2}(x^2 - 2x\mu_k + \mu_k^2) + \ln(\pi_k)$$

$$- \ln(\text{cte})$$

$$\alpha \quad \frac{x\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + \ln(\pi_k).$$

Estimadores:

$$\mu_k = \frac{1}{n_k}\sum_{i\in y_k}^{n_k} x_i$$ ← Promedio

Suma pondera la varianzas.

$$\hat{\sigma}_k = \frac{1}{N-K}\sum_{i=1}^{K}\sum_{i\in y_k}^{n_k}(x_i-\hat{\mu}_k)^2$$

$$\hat{\pi}_k = \frac{n_k}{n}$$ ← Proporción.

Ejemplo 2 Clases

$$P(Y=k \mid X=x)$$

↳ x se asigna a la clase 1 con probabilidad

$$P(Y=1|X=x)$$

y a la clase 2 con probabilidad

$$P(Y=2|X=x)$$

$$\boxed{\frac{P(Y=1|X=x)}{P(Y=2|X=x)} > 1}$$

Límite de descación:

$$x = \frac{\mu_1 + \mu_2}{2}$$

$$\frac{x\mu_1}{\sigma^2} - \frac{\mu_1^2}{2\sigma^2} + \ln(\pi_1).$$

$$\underline{\hspace{6cm}} > 1.$$

$$\frac{x\mu_2}{\sigma^2} - \frac{\mu_2^2}{2\sigma^2} + \ln(\pi_2)$$

$$\frac{2x\mu_1 - \mu_1^2 + 2\sigma^2\ln(\pi_1)}{2x\mu_2 - \mu_2^2 + 2\sigma^2\ln(\pi_2)} > 1.$$

$$2x\mu_1 - \mu_1^2 + 2\sigma^2\ln(\pi_1) > 2x\mu_2 - \mu_2^2 + 2\sigma^2\ln(\pi_2)$$

$$2x(\mu_1 - \mu_2) - (\mu_1^2 - \mu_2^2) + 2\sigma^2(\ln(\pi_1) - \ln(\pi_2)) > 0$$

$$x > \frac{(\mu_2^2 - \mu_2^2) - 2\sigma^2(\ln(\pi_1) - \ln(\pi_2))}{2(\mu_1 - \mu_2)}.$$

$$x > \frac{(\mu_1^2 - \mu_2^2) - 2\sigma^2(\ln(\pi_1/\pi_2))}{2(\mu_1 - \mu_2)}$$

$$x > \frac{(\cancel{\mu_1 - \mu_2})(\mu_1 + \mu_2) - 2\sigma^2(\ln(\pi_1/\pi_2))}{2(\cancel{\mu_1 - \mu_2})}$$

$$x > \frac{(\mu_1+\mu_2)}{2} - \frac{\sigma^2 \ln(\pi_1/\pi_2)}{(\mu_1-\mu_2)}.$$

Si $\pi_1 = \pi_2$ entonces.

$$x > \frac{\mu_1+\mu_2}{2}$$

El algoritmo de clasificación LDA.

1) Ajustar $\hat{\delta}_k(x)$ para $k = 1, 2, \ldots, C$.

$$\hat{\delta}_k(x) = x\frac{\hat{\mu}_k}{\hat{\sigma}^2} - \frac{\mu_k^2}{2\hat{\sigma}^2} + \log(\hat{\pi}_k)$$

2) Buscar el valor de k donde $\hat{\delta}_k(x)$ sea máximo

Se denomina análisis discriminante lineal por $\hat{\delta}_k(x)$ es lineal para el valor de x.

$$\hat{\delta}_k(x) = ax + b$$

$$a = \frac{\hat{\mu}_k}{\hat{\sigma}^2} \qquad b = \frac{\hat{\mu}_k^2}{2\hat{\sigma}^2} + \log(\hat{\pi}_k).$$

CASO MULTIVARIADO

$$X \sim N(\mu, \Sigma)$$

$$E(X) = \mu \qquad Cov(X) = \Sigma$$

$$f(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)' \Sigma^{-1} (x-\mu)\right\}$$

$$P(Y=k \mid X=x)$$

$$= \frac{\dfrac{1}{\cancel{(2\pi)^{p/2}}\,|\Sigma|^{1/2}}\, e^{\left\{-\frac{1}{2}(x-\mu_k)'\Sigma^{-1}(x-\mu_k)\right\}} \cdot \pi_k}{\displaystyle\sum_{i=1}^{c} \pi_i \frac{1}{\cancel{(2\pi)^{p/2}}\,|\Sigma|^{1/2}}\, e^{\left\{-\frac{1}{2}(x-\mu_i)'\Sigma(x-\mu_i)\right\}}}.$$

$$\ln(P(Y=k \mid X=x)) =$$

$$1/2 \ln(|\Sigma|) - \frac{1}{2}(x-\mu_k)'\Sigma^{-1}(x-\mu_k) + \ln(\pi_k)$$

$$- \ln(cte)$$

$$1/2 \ln(|\Sigma|) - \frac{1}{2}x'\Sigma^{-1}x + \frac{1}{2}\mu_k'\Sigma^{-1}x$$

$$\left(\frac{1}{2}\mu_k'\Sigma^{-1}x\right)' \quad + \underbrace{\frac{1}{2}x'\Sigma^{-1}\mu_k}_{} - \frac{1}{2}\mu_k'\Sigma^{-1}\mu_k$$

Ya que Σ es $\quad + \ln(\pi_k) - \ln(cte)$

ya que Σ es simétrica.

$$+\ln(\pi_k) - \ln(cte)$$

$$= \frac{1}{2}\ln(|\Sigma|) - \frac{1}{2}x'\Sigma^{-1}x + \mu_k'\Sigma^{-1}x - \frac{1}{2}\mu_k'\Sigma^{-1}\mu_k$$

$$+\ln(\pi_k) - \ln(cte)$$

$$\delta_k(x) = \underbrace{\mu_k'\Sigma^{-1}x - \frac{1}{2}\mu_k'\Sigma^{-1}\mu_k + \ln(\pi_k)}_{\text{Versión vectorial}}.$$

Nota:

- El desarrollo anterior supone que Σ es cte para todas las clases, bajo suposición de homogeneidad de varianzas.

Límite de decisión.

$$\mu_k'\Sigma^{-1}x - \frac{1}{2}\mu_k'\Sigma^{-1}\mu_k + \ln(\pi_k) >$$
$$\mu_l'\Sigma^{-1}x - \frac{1}{2}\mu_l'\Sigma^{-1}\mu_l + \ln(\pi_l)$$

Para los valores de x tales que se cumpla la condicion la clasificación se hace para el grupo k envez del grupo $\underline{\underline{l}}$.

Se calcula $\hat{\delta}_k$ para todos los valores de k.

Clasificación:

$$\hat{\delta}_k(x) = x'\hat{\Sigma}^{-1}\hat{\mu}_k - \frac{1}{2}\hat{\mu}_k'\hat{\Sigma}^{-1}\hat{\mu}_k + \log \hat{\pi}_k$$

$\hat{\Sigma} \leftarrow$ Matriz de varianzas y covarianzas

$\hat{\mu}_k \leftarrow$ Vector de medias

$\hat{\pi}_k \leftarrow$ Probabilidades a priori

Seleccionar como clase el valor más grande de δ_k:

Análisis discriminante Cuadrático

1. Se asume que $\Sigma_1 \neq \Sigma_p$ para algún par de grupos.

2. Cada $X \sim N(\mu_k, \Sigma_k)$

3.

$$\delta_k(x) = -\frac{1}{2}(x-\mu_k)^T \cdot \Sigma_k^{-1}(x-\mu_k) - \frac{1}{2}\log|\Sigma_k| + \log \pi_k.$$

$$= \underbrace{-\frac{1}{2}x'\Sigma_k^{-1}x}_{} + x'\Sigma_k^{-1}\mu_k - \frac{1}{2}\mu_k'\Sigma_k^{-1}\mu_k - \frac{1}{2}\log|\Sigma_k| + \log \pi_k.$$

Aparece este término.

$$P(Y=k \mid X=x)$$

$$= \frac{\dfrac{1}{\cancel{(2\pi)^{p/2}}|\Sigma_k|^{1/2}} e^{\left\{-\frac{1}{2}(x-\mu_k)'\Sigma_k^{-1}(x-\mu_k)\right\}} \cdot \pi_k}{\sum_{i=1}^{c} \pi_i \dfrac{1}{\cancel{(2\pi)^{p/2}}|\Sigma_i|^{1/2}} e^{\left\{-\frac{1}{2}(x-\mu_i)'\Sigma_i^{-1}(x-\mu_i)\right\}}}$$

mismo termino

$$\ln(P(Y=k \mid X=x)) =$$

$$= -\frac{1}{2}\ln(|\Sigma_k|) - \frac{1}{2}(x'-\mu_k)'\Sigma_k^{-1}(x-\mu_k)$$
$$+\ln(\pi_k) - \ln(cte)$$

$$\propto -\frac{1}{2}\ln(|\Sigma_k|) - \frac{1}{2}x'\Sigma_k^{-1}x + \frac{1}{2}x'\Sigma_k^{-1}\mu_k$$
$$+\frac{1}{2}\mu_k'\Sigma_k^{-1}x - \frac{1}{2}\mu_k'\Sigma_k^{-1}\mu_k + \ln(\pi_k)$$

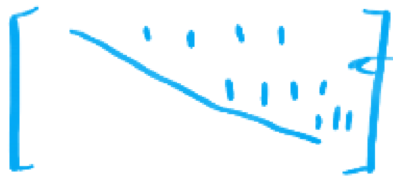
$$\propto -\frac{1}{2}\ln(|\Sigma_k|) - \frac{1}{2}x'\Sigma_k^{-1}x + x'\Sigma_k^{-1}\mu_k$$

$$-\frac{1}{2}\mu_k' \Sigma_k^{-1}\mu_k + \ln(\pi_k) = \delta(k)$$

1. Término cuadrático.
2. Depende del valor del $|\Sigma_k|$
3. Requiere estimar

$\frac{k\,p(p+1)}{2}$ parámetros para Σ_k

$p \rightarrow$ # variables. $\quad \sum_{i=1}^{n} i = \frac{n(n+1)}{2}.$



$\frac{n(n+1)}{2}$ valores diferentes:

Ejemplo LDA. – Generando Normales Aleatorias.

$U_1 \sim U(0,1) \qquad U_2 \sim U(0,1)$

$$\theta = 2\pi U_1 \qquad S = -\ln(u_2)$$

$$X = R\cos\theta \qquad R = \sqrt{2S}$$

$$Y = R\,\text{sen}\,\theta$$

$$X = \sqrt{2(-\ln(U_2))} \cdot \cos(2\pi(u_1))$$

X_1 = [0,18; 0,78; -0,73; -1.32;
1,07; 0.81; -1.1418; -2,03;
1.40; -2.01; -2.082; 3.66;
2.42; 3.45; 3.52; 3,74; 3.89;
4,11; 3.29; 2.51; 2.96; 0,31;
-0,11; 0.29; -0,01; 0.59; -0,22;
0,99]

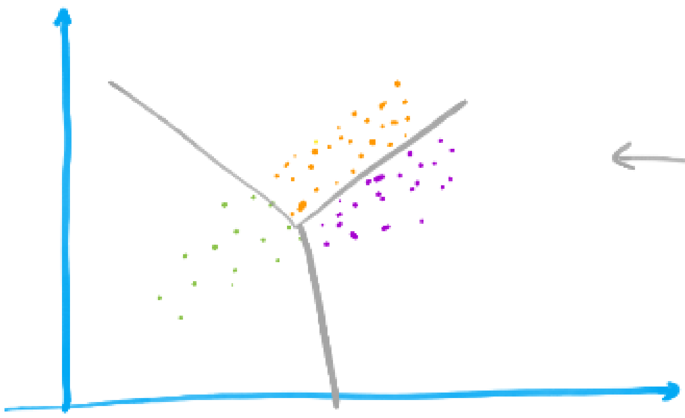
Discriminante lineal. - Ejemplo # 1.

1. Generar normales multivariadas (con la misma varianza).

$$\hat{\delta}_k(x) = x' \hat{\Sigma}^{-1} \hat{\mu}_k - \frac{1}{2} \hat{\mu}_k' \hat{\Sigma}^{-1} \hat{\mu}_k + \log \hat{\pi}_k$$

1. Simular una normal multivariada con $p=2$. con un vector de medias específico y una matriz de varianzas común. (con 3 clases)

20 observaciones x cada clase.



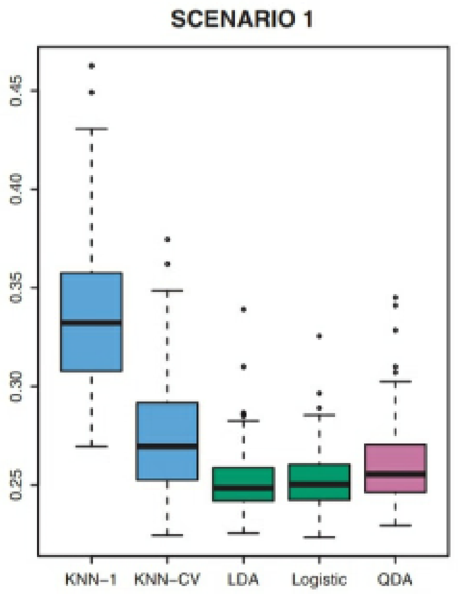
LDA y Regresión logística.

Simular los escenarios.

Escenario 1. $n = 20$ $K = 2.$

Cada clase tiene distribución normal de forma no correlacionada.

$X_1 \sim N(\mu_1, \sigma)$ $X_2 \sim N(\mu_2, \sigma)$



1) Generar 2 normales incorreladas con medias diferentes.

2) Clasificar por KNN, NNN-CV. LDA, logística y QDA.

3) Repetir y calcular los errores

4) Comparar.

LOOCV. → $\boxed{CV}$

ross
tation.
trap.