Tarea 5.
Reconocimiento Estadístico de Patrones.
Y. Sarahi García González

Problema 2.

Linear Discriminant Analysis

$$\hat{y}(x) = \begin{cases} 1 & \text{si } l^t x > c \\ 0 & \text{en otro caso} \end{cases}$$

Función discriminante
$l^t x + c$

$\Rightarrow$ x_0 es Clase 1 si $l^t x_0 \geq c$ y x_0 es Clase 0 si $l^t x_0 < c$

$\Rightarrow$ Estamos proyectando x a una dimensión con $l^t x$, esto tiene como consecuencia la pérdida de información.
Las clases podrían estar separadas en la dim original y tener mucho traslape al proyectar.

¿En que dirección debemos proyectar para obtener la mayor separación entre clases?

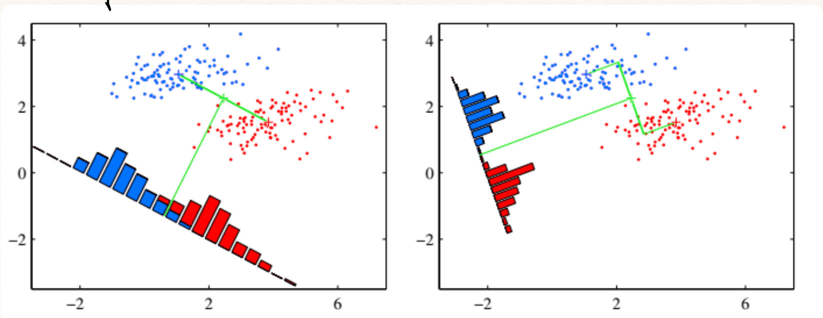
Idea 1 $\quad \underset{l}{\text{argmax}} \, (l^t c_+ - l^t c_-)^2 = \underbrace{[l^t (c_+ - c_-)]^2}$

separación entre los centroides al proyector sobre l^t

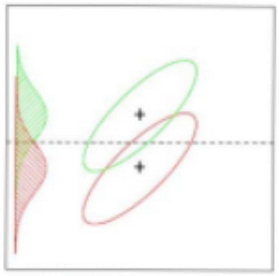
donde c_+ y c_- son los centroides de los clusters

Problemas: 1. La expresión $l^t (c_+ - c_-)$ se puede hacer arbitariamente grande cambiando la norma de l

2. Traslape de las clases al proyectar en 1D, consecuencia de que la matriz de covarianza de las clases es no diagonal



$\rightarrow$ De Bishop Pattern Recognition the above solution:



→ Diapositivas recapt 112024 pdf

Para solucionar lo anterior:

El criterio de Fisher se define como la relación entre la variación entre lasclases y la variación total.

Asumiendo que la covarianza S_w no depende de la clase, tenemos el sig. problema a resolver:

$$\underset{l\neq 0}{\operatorname{argmax}} \frac{[l^t(c_+ - c_-)]^2}{l^t S_w l} = \underset{l\neq 0}{\operatorname{argmax}} \frac{l^t \overbrace{(c_+ - c_-)(c_+ - c_-)^t}^{\text{variación entre clases}} l}{l^t S_w l}$$

empieza aquí →

$$= \underset{l\neq 0}{\operatorname{argmax}} \frac{l^t S_B l}{l^t S_w l}$$

Derivamos:

$$\frac{d}{dl}\left(\frac{l^t S_B l}{l^t S_w l}\right) = \frac{l^t S_B l (S_w l) - l^t S_w l (S_B l)}{(l^t S_w l)^2}$$

Igualamos a cero:

$$l^t S_B l (S_w l) - l^t S_w l (S_B l) = 0$$

$$\Rightarrow \underbrace{(l^t S_B l)}_{\text{ESCALAR}} S_w l = \underbrace{(l^t S_w l)}_{\text{ESCALAR}} S_B l$$

Sea $\alpha = \dfrac{l^t S_B l}{l^t S_w l}$

$$\Rightarrow \alpha S_W l = S_B l \Rightarrow \alpha S_W^{-1} S_W l = S_W^{-1} S_B l$$

$$\Rightarrow \alpha l = S_W^{-1}(S_B l) \Rightarrow \alpha l = S_W^{-1}(C_+ - C_-)\underbrace{(C_+ - C_-)^t l}_{\text{ESCALAR}}$$

Sea $\dfrac{(C_+ - C_-)^t l}{\alpha} = \lambda$

$$\Rightarrow l = \lambda S_W^{-1}(C_+ - C_-)$$

Por lo que la dirección de l es:

$$\underline{S_W^{-1}(C_+ - C_-)}$$

Problema 3

En regresio logistica tenemos el modelo

$$\pi(x) = P(Y=1 \mid X=x) = (1 + \exp(-\beta^t x))^{-1}$$

Usamos log verosimilitud para estimar los parámetros:

$$l(\beta) = \log \prod_{i=0}^{N} f_{Y_i}(y_i ; \beta)$$

Tenemos el daset $\{\pi(X_n), y_n\}$ con $n = \{1, \dots N\}$, y $y_n \in \{0,1\}$ así que usamos la dist. de Bernoulli:

$$l(\beta) = \log \prod_{i=0}^{N} \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}$$

Para aligerar la notación, usaremos $\pi_i = \pi(X_i)$

$$\Rightarrow l(\beta) = \log \prod_{i=0}^{N} \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$

$$= \sum_{i=0}^{N} \log\left[\pi_i^{y_i} (1 - \pi_i)^{1-y_i}\right]$$

$$= \sum^{N} y_i \log \pi_i + (1-y_i)\log(1-\pi_i)$$

$$= \sum_{i=0}^{N} y_i\left[\log \pi_i - \log(1-\pi_i)\right] + \log(1-\pi_i)$$

empieza aquí $\rightarrow$ $$= \sum_{i=0}^{N} y_i\left[\log\left(\frac{\pi_i}{1-\pi_i}\right)\right] - \log\left(\frac{1}{1-\pi_i}\right)$$

Sustituimos $\pi_i = \pi(x_i) = (1+\exp(-\beta^t x_i))^{-1}$

$$\Rightarrow l(\beta) = \sum_{i=0}^{N}\left[y_i \log\left(\frac{(1+\exp(-\beta^t x_i))^{-1}}{1-(1+\exp(-\beta^t x_i))^{-1}}\right)\right.$$
$$\left. - \log\left(\frac{1}{1-(1+\exp(-\beta^t x_i))^{-1}}\right)\right]$$

Notemos que el denominador de ambos términos puede reescribirse como:

$$\square = 1 - \frac{1}{1+\exp(-\beta^t x_i)} = \frac{\exp(-\beta^t x_i)}{1+\exp(-\beta^t x_i)}$$

Sustituimos:

$$\Rightarrow l(\beta) = \sum_{i=0}^{N}\left[y_i \log\left(\frac{\frac{1}{\cancel{(1+\exp(-\beta^t x_i))}}}{\frac{\exp(-\beta^t x_i)}{\cancel{1+\exp(-\beta^t x_i)}}}\right)\right.$$
$$\left. - \log\left(\frac{1}{\frac{\exp(-\beta^t x_i)}{1+\exp(-\beta^t x_i)}}\right)\right]$$

$$= \sum_{i=0}^{N} y_i \log\left(\frac{1}{\exp(-\beta^t x_i)}\right) - \log\left(\frac{1+\exp(-\beta^t x_i)}{\exp(-\beta^t x_i)}\right)$$

$$= \sum_{i=0}^{N} y_i \log\left(\exp(\beta^t x_i)\right) - \log\left(\frac{1}{\exp(-\beta^t x_i)} + 1\right)$$

$$\therefore \; l(\beta) = \sum_{i=0}^{N} y_i \beta^t x_i - \log\left[\exp(\beta^t x_i) + 1\right] \qquad \blacksquare$$

Ahora las derivadas:

- $$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=0}^{N} y_i x_i - \frac{1}{\exp(\beta^t x_i)+1} * \left(x_i \exp(\beta^t x_i)\right)$$

$$= \sum_{i=0}^{N} y_i x_i - x_i\left(\frac{\exp(\beta^t x_i)}{1+\exp(\beta^t x_i)}\right)$$

Pero podemos reescribir la expresion

$$= \frac{\frac{1}{\exp(-\beta^t x_i)}}{1 + \frac{1}{\exp(-\beta^t x_i)}} = \frac{\frac{1}{\cancel{\exp(-\beta^t x_i)}}}{\frac{1+\exp(-\beta^t x_i)}{\cancel{\exp(-\beta^t x_i)}}}$$

$$= \frac{1}{1+\exp(-\beta^t x_i)} = \left(1 + \exp(-\beta^t x_i)\right)^{-1} = \pi(x_i)$$

Sustituimos y usamos $\pi_i = \pi(x_i)$

$$\Rightarrow \frac{\partial l(\beta)}{\partial \beta} = \sum_i y_i x_i - x_i \pi(x_i) = \sum_i x_i (y_i - \pi_i) \qquad \blacksquare$$

- $$\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta^t} = \frac{\partial}{\partial \beta}\left(\sum_i x_i (y_i - \pi_i)\right)$$

$$= -\sum x_i \frac{\partial}{\partial \beta} \pi_i$$

$$= -\sum x_i \frac{\partial}{\partial \beta} \left(1 + \exp(-\beta^t x_i)\right)^{-1}$$

Pero: $\boxed{} = \left(1+\exp(-\beta^t x_i)\right)^{-2} * \left(-x_i^t \exp(-\beta^t x_i)\right)$

$$= \underbrace{\left(1+\exp(-\beta^t x_i)\right)^{-1}}_{\pi_i} * \left(\frac{-x_i^t \exp(-\beta^t x_i)}{1+\exp(-\beta^t x_i)}\right)$$

$$= -\pi_i * \frac{\exp(-\beta^t x_i)}{1+\exp(-\beta^t x_i)} x_i^t$$

$$= -\pi_i * \frac{\exp(-\beta^t x_i)+(1-1)}{1+\exp(-\beta^t x_i)} x_i^t$$

$$= -\pi_i * \frac{1+\exp(-\beta^t x_i)-1}{1+\exp(-\beta^t x_i)} x_i^t$$

$$= -\pi_i * \left(1 - \frac{1}{\underbrace{1+\exp(-\beta^t x_i)}_{\pi_i}} x_i^t\right)$$

$$= -\pi_i (1-\pi_i) x_i^t$$

Sustituimos:

$$\Rightarrow \frac{\partial^2 l(\beta)}{\partial\beta\partial\beta^t} = -\sum x_i \left(-\pi_i(1-\pi_i)\right) x_i^t$$

$$= \sum_i x_i \left[-\pi(x_i)(1-\pi(x_i))\right] x_i^t$$

Bayesian naive Classifier

$$\hat{Y}(x) = \mathbb{I}\left(\frac{P(X=x \mid Y=1)}{P(X=x \mid Y=0)} > c_1 \right) *$$

asumimos que $X \mid Y=y$ son independientes

$$P(X=x \mid Y=y) = \prod_i P(X_i = x_i \mid Y=y) *$$

Para cada caracteristica

$$\underline{P(X_1 = x_1 \mid Y=0)} = e^{-\frac{(x_1-\mu)^2}{\sigma^2}}$$

sin integral porque NO es la acumulada

hay que iterar sobre cada caracteristica

para cada x_j en el conjunto test

μ y σ se calculan con el conjunto train.

$P(Y=y)$ es la marginal.