Trabajo Práctico Nro. 1

Analisis Multivariado

Prieto Julián 45065709

1.a

Calcular gradiente y hessiano de:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log(h_\theta(x^{(i)})) + (1-y^{(i)})\log(1-h_\theta(x^{(i)}))\right]$$

$$\text{Teniendo en cuenta que, } h_\theta(x) = g(\theta^T x),$$

Gradiente:

$$\nabla J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\frac{y^{(i)}}{g(\theta^T x^{(i)})}g'(\theta^T x^{(i)}) - \frac{1-y^{(i)}}{1-g(\theta^T x^{(i)})}g'(\theta^T x^{(i)})$$

$$\nabla J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[\frac{y^{(i)}}{g(\theta^T x^{(i)})} - \frac{1-y^{(i)}}{1-g(\theta^T x^{(i)})}\right]g'(\theta^T x^{(i)})$$

Ya que $\frac{\partial}{\partial\theta}g(\theta^T x) = g(\theta^T x)(1-g(\theta^T x))x$,

$$\nabla J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[\frac{y^{(i)}}{g(\theta^T x^{(i)})} - \frac{1-y^{(i)}}{1-g(\theta^T x^{(i)})}\right]g(\theta^T x^{(i)})(1-g(\theta^T x^{(i)}))x^{(i)}$$

$$\nabla J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[\frac{y^{(i)}g(\theta^T x^{(i)})(1-g(\theta^T x^{(i)}))}{g(\theta^T x^{(i)})} - \frac{(1-y^{(i)})g(\theta^T x^{(i)})(1-g(\theta^T x^{(i)}))}{1-g(\theta^T x^{(i)})}\right]x^{(i)}$$

$$\nabla J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}(1-g(\theta^T x^{(i)})) - (1-y^{(i)})g(\theta^T x^{(i)})\right]x^{(i)}$$

$$\nabla J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} - y^{(i)}g(\theta^T x^{(i)}) - g(\theta^T x^{(i)}) + y^{(i)}g(\theta^T x^{(i)})\right]x^{(i)}$$

$$\nabla J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} - g(\theta^T x^{(i)})\right]x^{(i)}$$

$$\boxed{\nabla J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left[g(\theta^T x^{(i)}) - y^{(i)}\right]x^{(i)}}$$

Hessiano:

$$\nabla^2 J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left[g(\theta^T x^{(i)})\right]' x^{(i)}$$

$$\boxed{\nabla^2 J(\theta) = \frac{1}{m}\sum_{i=1}^{m}g(\theta^T x^{(i)})(1-g(\theta^T x^{(i)}))x^{(i)^T}x^{(i)}}$$

Es semi-definido positivo?

Primero demostremos que $\nabla^2 J(\theta) = (\nabla^2 J(\theta))^T$

$$\frac{1}{m}\sum_{i=1}^{m}g(\theta^T x^{(i)})(1-g(\theta^T x^{(i)}))x^{(i)^T}x^{(i)} = \left[\frac{1}{m}\sum_{i=1}^{m}g(\theta^T x^{(i)})(1-g(\theta^T x^{(i)}))x^{(i)^T}x^{(i)}\right]^T$$

Teniendo en cuenta que la expresion $g(\theta^T x)(1-g(\theta^T x))$ puede ser escrita como una matriz diagonal,

$$D = \begin{pmatrix} g(\theta^T x^{(1)})(1-g(\theta^T x^{(1)})) & & \\ & \ddots & \\ & & g(\theta^T x^{(m)})(1-g(\theta^T x^{(m)})) \end{pmatrix}$$

$$D^T = D$$

y que por propiedad de la transposición,

$$(x^T x)^T = x^T x$$

se cumple que $\nabla^2 J(\theta) = (\nabla^2 J(\theta))^T$

Por ultimo, probemos que $z^{(i)^T} \cdot \nabla^2 J(\theta) \cdot z^{(i)} \geq 0$, con $z \in \mathbb{R}^n$

$$z^{(i)^T} \cdot \nabla^2 J(\theta) \cdot z^{(i)}$$

$$z^{(i)^T} \cdot \frac{1}{m} \sum_{i=1}^{m} g(\theta^T x^{(i)})(1 - g(\theta^T x^{(i)}))x^{(i)^T} x^{(i)} \cdot z^{(i)}$$

Ya que $0 < g(\theta^T x) < 1$, y que

$$\sum_{i=1}^{m} z^{(i)^T} \cdot x^{(i)^T} \cdot x^{(i)} \cdot z^{(i)} = (x^T z)^2 \geq 0$$

$$\frac{1}{m} \sum_{i=1}^{m} g(\theta^T x^{(i)})(1 - g(\theta^T x^{(i)})) \cdot (x^{(i)^T} z^{(i)})^2 \geq 0$$

Podemos decir que $\nabla^2 J(\theta) = H \geq 0$

1.c

Mostrar que

$$p(y = 1|x; \phi, \mu_0, \mu_1, \Sigma) = \frac{1}{1 + \exp(-(\theta^T x + \theta_0))}$$

$$\text{donde } \theta \in \mathbb{R}^n \text{ y } \theta_0 \in \mathbb{R} \text{ son funciones de } \phi, \mu_0, \mu_1 \text{ y } \Sigma.$$

$$p(y = 1|x; \phi, \mu_0, \mu_1, \Sigma) = \frac{p(x|y = 1) \cdot p(y = 1)}{p(x)}$$

$$p(y = 1|x; \phi, \mu_0, \mu_1, \Sigma) = \frac{p(x|y = 1) \cdot p(y = 1)}{p(x|y = 1) \cdot p(y = 1) \cdot p(x|y = 0) \cdot p(y = 0)}$$

diviendo numerador y denominador por $p(x|y = 1) \cdot p(y = 1)$,

$$p(y = 1|x; \phi, \mu_0, \mu_1, \Sigma) = \frac{1}{1 + \frac{p(x|y=0) \cdot p(y=0)}{p(x|y=1) \cdot p(y=1)}}$$

reemplazando por funcion de probabilidad,

$$p(y = 1|x; \phi, \mu_0, \mu_1, \Sigma) = \frac{1}{1 + \frac{\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_0)^T \Sigma^{-1}(x-\mu_0)\right) \cdot (1-\phi)}{\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_1)^T \Sigma^{-1}(x-\mu_1)\right) \cdot \phi}}$$

por propiedad $\frac{a^n}{a^m} = a^{n-m}$, y cancelando termino constante,

$$= \frac{1}{1 + \exp(-\frac{1}{2}(x - \mu_0)^T \Sigma^{-1}(x - \mu_0) + \frac{1}{2}(x - \mu_1)^T \Sigma^{-1}(x - \mu_1)) \frac{(1-\phi)}{\phi}}$$

$$= \frac{1}{1 + \exp\left(-\frac{1}{2}(x - \mu_0)^T \Sigma^{-1}(x - \mu_0) + \frac{1}{2}(x - \mu_1)^T \Sigma^{-1}(x - \mu_1) + \ln \frac{(1-\phi)}{\phi}\right)}$$

aplicando distributiva,

$$= \frac{1}{1 + \exp\left(-\frac{1}{2}(x^T \Sigma^{-1} - \mu_0^T \Sigma^{-1})(x - \mu_0) + \frac{1}{2}(x^T \Sigma^{-1} - \mu_1^T \Sigma^{-1})(x - \mu_1) + \ln \frac{(1-\phi)}{\phi}\right)}$$

$$= \frac{1}{1 + \exp\left(-\frac{1}{2}(x^T \Sigma^{-1} x - x^T \Sigma^{-1} \mu_0 - \mu_0^T \Sigma^{-1} x + \mu_0^T \Sigma^{-1} \mu_0) + \frac{1}{2}(x^T \Sigma^{-1} x - x^T \Sigma^{-1} \mu_1 - \mu_1^T \Sigma^{-1} x + \mu_1^T \Sigma^{-1} \mu_1) + \ln \frac{(1-\phi)}{\phi}\right)}$$

ya que $x^T \Sigma^{-1} \mu_1 = \mu_1^T \Sigma^{-1} x$

$$= \frac{1}{1 + \exp\left(-\frac{1}{2}(x^T \Sigma^{-1} x - 2x^T \Sigma^{-1} \mu_0 + \mu_0^T \Sigma^{-1} \mu_0) + \frac{1}{2}(x^T \Sigma^{-1} x - 2x^T \Sigma^{-1} \mu_1 + \mu_1^T \Sigma^{-1} \mu_1) + \ln \frac{(1-\phi)}{\phi}\right)}$$

$$= \frac{1}{1 + \exp\left(-\frac{x^T \Sigma^{-1} x}{2} + x^T \Sigma^{-1} \mu_0 - \frac{\mu_0^T \Sigma^{-1} \mu_0}{2} + \frac{x^T \Sigma^{-1} x}{2} - x^T \Sigma^{-1} \mu_1 + \frac{\mu_1^T \Sigma^{-1} \mu_1}{2} + \ln \frac{(1-\phi)}{\phi}\right)}$$

$$= \frac{1}{1 + \exp\left(x^T \Sigma^{-1} \mu_0 - \frac{\mu_0^T \Sigma^{-1} \mu_0}{2} - x^T \Sigma^{-1} \mu_1 + \frac{\mu_1^T \Sigma^{-1} \mu_1}{2} + \ln \frac{(1-\phi)}{\phi}\right)}$$

$$= \frac{1}{1+\exp\left(x^T\Sigma^{-1}\mu_0 - x^T\Sigma^{-1}\mu_1 - \frac{\mu_0^T\Sigma^{-1}\mu_0}{2} + \frac{\mu_1^T\Sigma^{-1}\mu_1}{2} + \ln\frac{(1-\phi)}{\phi}\right)}$$

$$\boxed{= \frac{1}{1+\exp\left(x^T\Sigma^{-1}(\mu_0-\mu_1) - \frac{1}{2}\Sigma^{-1}(\mu_0^T\mu_0 - \mu_1^T\mu_1) + \ln\frac{(1-\phi)}{\phi}\right)}}$$

$\text{con } \theta = \Sigma^{-1}(\mu_1 - \mu_0) \text{ y con } \theta_0 = \frac{1}{2}\Sigma^{-1}(\mu_0^T\mu_0 - \mu_1^T\mu_1) - \ln\frac{(1-\phi)}{\phi}$

---

se llega a la expresion:

$$p(y=1|x;\phi,\mu_0,\mu_1,\Sigma) = \frac{1}{1+\exp(-(\theta^Tx+\theta_0))}$$

1.d

$$l(\phi,\mu_0,\mu_1,\Sigma) = \prod_{i=1}^{m} \ln p\left(x^{(i)}, y^{(i)}; \phi, \mu_0, \mu_1, \Sigma\right)$$

$$= \prod_{i=1}^{m} \ln p\left(x^{(i)} \mid y^{(i)}; \phi, \mu_0, \mu_1, \Sigma\right) p(y^{(i)};\phi)$$

$$= \sum_{i=1}^{m} \ln p\left(x^{(i)} \mid y^{(i)}; \phi, \mu_0, \mu_1, \Sigma\right) + \ln p(y^{(i)};\phi)$$

$$= \sum_{i=1}^{m} \ln\left[\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp(-\frac{1}{2}(x^{(i)}-\mu_n)^T\Sigma^{-1}(x^{(i)}-\mu_n))\right] + \ln\left[\phi^{y^{(i)}}(1-\phi)^{1-y^{(i)}}\right]$$

$$= \sum_{i=1}^{m} \ln\left[\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\right] - \frac{1}{2}(x^{(i)}-\mu_n)^T\Sigma^{-1}(x^{(i)}-\mu_n) + y^{(i)}\ln(\phi) + (1-y^{(i)})\ln(1-\phi)$$

$$= \sum_{i=1}^{m} \ln(1) - \ln\left[(2\pi)^{n/2}|\Sigma|^{1/2}\right] - \frac{1}{2}(x^{(i)}-\mu_n)^T\Sigma^{-1}(x^{(i)}-\mu_n) + y^{(i)}\ln(\phi) + (1-y^{(i)})\ln(1-\phi)$$

$$= \sum_{i=1}^{m} -\ln\left[(2\pi)^{n/2}\right] - \ln\left[|\Sigma|^{1/2}\right] - \frac{1}{2}(x^{(i)}-\mu_n)^T\Sigma^{-1}(x^{(i)}-\mu_n) + y^{(i)}\ln(\phi) + (1-y^{(i)})\ln(1-\phi)$$

$$\boxed{= \sum_{i=1}^{m} -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln(|\Sigma|) - \frac{1}{2}(x^{(i)}-\mu_n)^T\Sigma^{-1}(x^{(i)}-\mu_n) + y^{(i)}\ln(\phi) + (1-y^{(i)})\ln(1-\phi)}$$

Ahora debemos derivar parcialmente e igualar a 0 para cada parametro

Con respecto a ϕ:

$$\frac{\partial l}{\partial \phi} = \sum_{i=1}^{m}\left[y^{(i)}\ln(\phi) + (1-y^{(i)})\ln(1-\phi)\right]'$$

$$\frac{\partial l}{\partial \phi} = \sum_{i=1}^{m}\left[\frac{y^{(i)}}{\phi} - \frac{1-y^{(i)}}{1-\phi}\right]$$

$$\frac{\partial l}{\partial \phi} = \sum_{i=1}^{m}\left[\frac{y^{(i)}\phi(1-\phi)}{\phi} - \frac{(1-y^{(i)})\phi(1-\phi)}{1-\phi}\right]$$

$$\frac{\partial l}{\partial \phi} = \sum_{i=1}^{m}\left[y^{(i)}(1-\phi) - (1-y^{(i)})\phi\right]$$

$$\frac{\partial l}{\partial \phi} = \sum_{i=1}^{m}\left[y^{(i)} - y^{(i)}\phi - \phi + y^{(i)}\phi\right]$$

$$\frac{\partial l}{\partial \phi} = \sum_{i=1}^{m}\left[y^{(i)} - \phi\right] = \sum_{i*1}^{m} y^{(i)} - \sum_{i=1}^{m}\phi$$

$$\frac{\partial l}{\partial \phi} = \sum_{i=1}^{m} y^{(i)} - \sum_{i*1}^{m}\phi \ = \boxed{\sum_{i=1}^{m} y^{(i)} - m\cdot\phi}$$

$$\sum_{i=1}^{m} y^{(i)} - m\cdot\phi = 0$$

$$-m\cdot\phi = -\sum_{i=1}^{m} y^{(i)}$$

$$\phi = \frac{1}{m}\sum_{i=1}^{m} y^{(i)} = \boxed{\frac{1}{m}\sum_{i=1}^{m} 1\{y^{(i)} = 1\}}$$

Con respecto a μ_n, con $n \in \{0, 1\}$:

$$\frac{\partial l}{\partial \mu_n} = \left[-\frac{1}{2}\sum_{i=1}^{m}(x^{(i)} - \mu_n)^T \Sigma^{-1}(x^{(i)} - \mu_n)\right]'$$

$$\frac{\partial l}{\partial \mu_n} = \left[-\frac{1}{2}\sum_{i=1}^{m}(x^{(i)} - \mu_n)^T \Sigma^{-1}(x^{(i)} - \mu_n)\right]'$$

aplicando distributiva,

$$\frac{\partial l}{\partial \mu_n} = \left[-\frac{1}{2}\sum_{i=1}^{m} x^{(i)^T} \Sigma^{-1} x^{(i)} - 2\mu_n^T \Sigma^{-1} x^{(i)} + \mu_n^T \Sigma^{-1} \mu_n\right]'$$

$$\frac{\partial l}{\partial \mu_n} = -\frac{1}{2}\sum_{i=1}^{m} -2\Sigma^{-1} x^{(i)} + 2\Sigma^{-1}\mu_n$$

aplicando nuevamente distributiva,

$$\boxed{\frac{\partial l}{\partial \mu_n} = \sum_{i=1}^{m} \Sigma^{-1} x^{(i)} - \sum_{i=1}^{m} \Sigma^{-1}\mu_n}$$

---

igualando a 0,

$$\sum_{i=1}^{m} \Sigma^{-1} x^{(i)} - \sum_{i=1}^{m} \Sigma^{-1}\mu_n = 0$$

$$\sum_{i=1}^{m} 1\{y^{(i)} = n\}\Sigma^{-1} x^{(i)} - \sum_{i=1}^{m} 1\{y^{(i)} = n\}\Sigma^{-1}\mu_n = 0$$

$$\Sigma^{-1}\sum_{i=1}^{m} 1\{y^{(i)} = n\} x^{(i)} - \Sigma^{-1}\sum_{i=1}^{m} 1\{y^{(i)} = n\}\mu_n = 0$$

$$\Sigma^{-1}\sum_{i=1}^{m} 1\{y^{(i)} = n\} x^{(i)} = \Sigma^{-1}\sum_{i=1}^{m} 1\{y^{(i)} = n\}\mu_n$$

$$\frac{\Sigma^{-1}\sum_{i=1}^{m} 1\{y^{(i)} = n\} x^{(i)}}{\Sigma^{-1}\sum_{i=1}^{m} 1\{y^{(i)} = n\}} = \mu_n$$

$$\frac{\sum_{i=1}^{m} 1\{y^{(i)} = n\} x^{(i)}}{\sum_{i=1}^{m} 1\{y^{(i)} = n\}} = \mu_n$$

Es decir, nos quedaria de la siguiente manera:

$$\boxed{\mu_0 = \frac{\sum_{i=1}^{m} 1\{y^{(i)} = 0\} x^{(i)}}{\sum_{i=1}^{m} 1\{y^{(i)} = 0\}} \text{ y } \mu_1 = \frac{\sum_{i=1}^{m} 1\{y^{(i)} = 1\} x^{(i)}}{\sum_{i=1}^{m} 1\{y^{(i)} = 1\}}}$$

Con respecto a Σ:

$$\frac{\partial l}{\partial \Sigma} = \sum_{i=1}^{m}\left[-\frac{1}{2}\ln|\Sigma| - \frac{1}{2}(x^{(i)} - \mu_n)^T \Sigma^{-1}(x^{(i)} - \mu_n)\right]'$$

$$\frac{\partial l}{\partial \Sigma} = -\frac{1}{2}\sum_{i=1}^{m}\left[\ln|\Sigma| + (x^{(i)} - \mu_n)^T \Sigma^{-1}(x^{(i)} - \mu_n)\right]'$$

Con $\Sigma = [\sigma^2]$ y $|\Sigma| = \sigma^2$

$$\frac{\partial l}{\partial \Sigma} = -\frac{1}{2}\sum_{i=1}^{m}\left[\ln(\sigma^2) + (x^{(i)} - \mu_n)^T \frac{1}{\sigma^2}(x^{(i)} - \mu_n)\right]'$$

$$\frac{\partial l}{\partial \Sigma} = -\frac{1}{2}\sum_{i=1}^{m}\left[\frac{2\sigma}{\sigma^2} - \frac{2}{\sigma^3}(x^{(i)} - \mu_n)^T (x^{(i)} - \mu_n)\right]$$

$$\frac{\partial l}{\partial \Sigma} = -\frac{1}{2}\sum_{i=1}^{m}\left[\frac{2}{\sigma} - \frac{2}{\sigma^3}(x^{(i)} - \mu_n)^T (x^{(i)} - \mu_n)\right]$$

$$\frac{\partial l}{\partial \Sigma} = \sum_{i=1}^{m}\left[-\frac{2}{2\sigma} + \frac{2}{2\sigma^3}(x^{(i)} - \mu_n)^T (x^{(i)} - \mu_n)\right]$$

$$\boxed{\frac{\partial l}{\partial \Sigma} = \sum_{i=1}^{m}\left[-\frac{1}{\sigma} + \frac{1}{\sigma^3}(x^{(i)} - \mu_n)^T (x^{(i)} - \mu_n)\right]}$$

$$\sum_{i=1}^{m}\left[-\frac{1}{\sigma}+\frac{1}{\sigma^3}(x^{(i)}-\mu_n)^T(x^{(i)}-\mu_n)\right]=0$$

$$\sum_{i=1}^{m}\left[\frac{1}{\sigma^3}(x^{(i)}-\mu_n)^T(x^{(i)}-\mu_n)\right]=\sum_{i=1}^{m}\frac{1}{\sigma}$$

$$\sum_{i=1}^{m}\left[(x^{(i)}-\mu_n)^T(x^{(i)}-\mu_n)\right]=\frac{\sum_{i=1}^{m}\frac{1}{\sigma}}{\sum_{i=1}^{m}\frac{1}{\sigma^3}}$$

$$\sum_{i=1}^{m}\left[(x^{(i)}-\mu_n)^T(x^{(i)}-\mu_n)\right]=\sum_{i=1}^{m}\frac{\sigma^3}{\sigma}$$

$$\sum_{i=1}^{m}\left[(x^{(i)}-\mu_n)^T(x^{(i)}-\mu_n)\right]=\sum_{i=1}^{m}\sigma^2$$

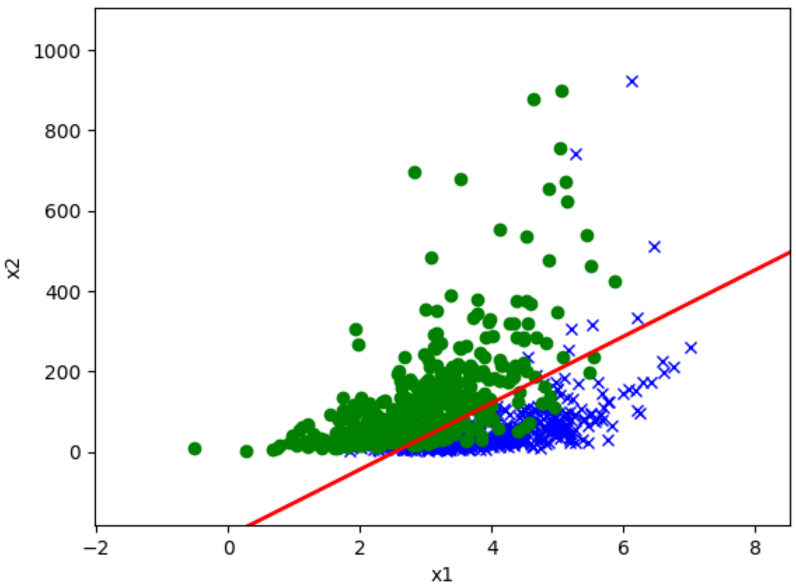
$$\sum_{i=1}^{m}\left[(x^{(i)}-\mu_n)^T(x^{(i)}-\mu_n)\right]=m\cdot\sigma^2$$

$$\sum_{i=1}^{m}\left[(x^{(i)}-\mu_n)^T(x^{(i)}-\mu_n)\right]\cdot\frac{1}{m}=\sigma^2$$

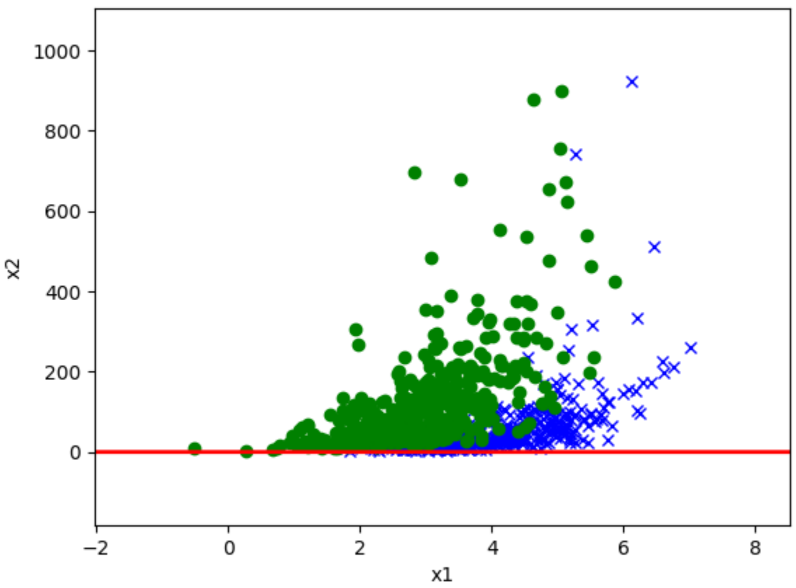
$$\boxed{\frac{1}{m}\sum_{i=1}^{m}\left[(x^{(i)}-\mu_n)^T(x^{(i)}-\mu_n)\right]=\Sigma}$$

1.f

Regresion Logistica (DS1)

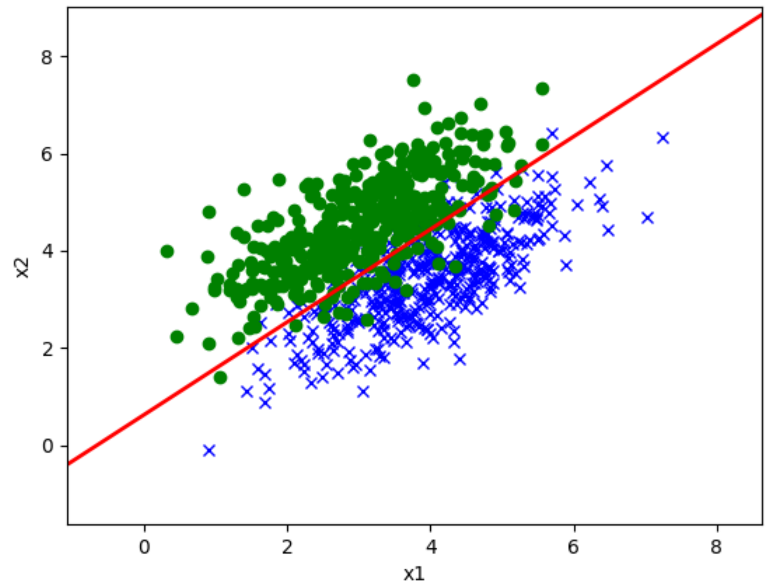


Analisis Discriminante Gaussiano (DS1)

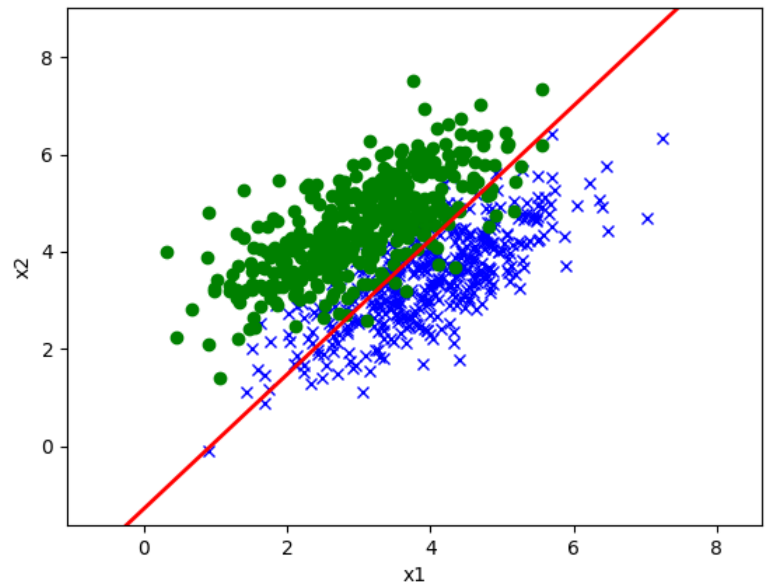


1.g

Regresion Logistica (DS2)



Analisis Discriminante Gaussiano (DS2)

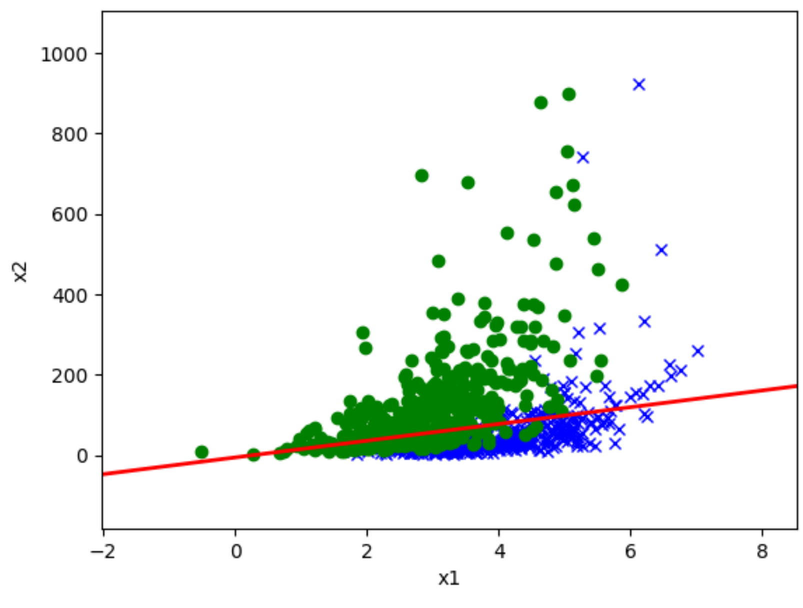


Guiandome por los graficos de GDA y Regresion Logistica de los dataset DS1 y DS2, el modelo de Discriminante Gaussiano tiene un peor funcionamiento en el dataset numero 1, ya que el borde de decision apenas es movido, es decir, que los parametros calculados no son los mas adecuados aparentemente para estos datos en especifico. Esto se puede deber a diferentes motivos, y todos dependen del dataset utilizado. Pero claramente el metodo utilizado no es el adecuado para estos datos. Una teoria podría ser que el modelo de GDA asume y da por hecho que las variables siguen una distribución normal/gaussiana. Esto puede ser lo que este causando la mala performance expecificamente con el dataset 1, el cual posiblemente sus variables no sigan este tipo de distribución

---

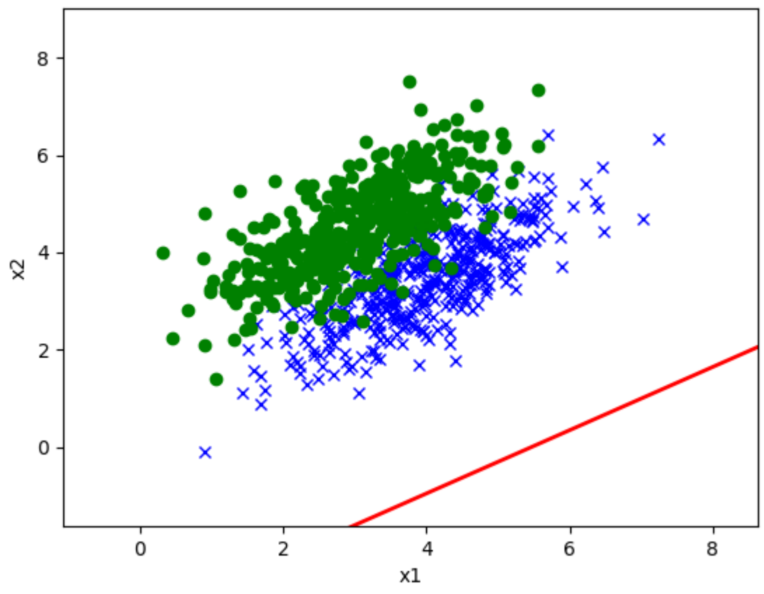
1.h

Analisis Discriminante Gaussiano (DS1), con transformación recíproca



Realizando la siguiente transformacion de x: $\frac{-1}{x}$, llamada **reciprocidad**, **recíproca**, **inversa** (todas negativas) se logra una notoria mejoría en el borde de decisión en el dataset numero 1. Sin embargo, la linea en el dataset numero 2 se modifica de gran manera empeorando las predicciones. Este tipo de tranformación (según internet) es recomendable si el tamaño de muestra es pequeño o los datos estan sesgados hacia una distribucion no normal. En este caso, se intentó probar con la transformacion inversa convencional (1/x), pero los resultados demejoraban de gran manera. Por esto es que se probó realizar la inversa negativa (-1/x), con la cual el borde de desicion mejoró bastante visualmente. Esto esta fuertemente relacionado a lo que teorizamos en el punto anterior, ya que si los datos del DS1 no sigue una distribucion normal, la transformacion inversa justamente apunta a "solucionar" esto y acercar los datos a una distribucion gaussiana.

Analisis Discriminante Gaussiano (DS2), con transformación recíproca



En este gráfico (dataset 2) es observable como el borde de desicion esta errado.Sin embargo, la pendiente de este pareciera ser acertada, solo que la posicion de la linea no es la correcta. Esto tiene sentido si se sabe que para cada modelo y para cada dataset es diferente que transformaciones funcionan. Es posible que el DS2 ya tenga una distribucion gaussiana y que al aplicarle la transformacion inversa esto arruine las predicciones.

2.a

Demostrar que

$$p(t^{(i)} = 1|x^{(i)}) = \frac{p(y^{(i)} = 1|x^{(i)})}{\alpha}$$

$$p(t^{(i)} = 1|x^{(i)}) = p(t^{(i)} = 1 \wedge y^{(i)} = 1|x^{(i)})$$

$$p(t^{(i)} = 1|x^{(i)}) = p(t^{(i)} = 1|x^{(i)}) \cdot p(y^{(i)} = 1|t^{(i)} = 1, x^{(i)})$$

$$p(t^{(i)} = 1|x^{(i)}) = p(t^{(i)} = 1|x^{(i)}) \cdot p(y^{(i)} = 1|t^{(i)} = 1)$$

Entonces, con $\alpha = p(y^{(i)} = 1|t^{(i)} = 1)$, se cumple que,

$$\boxed{p(t^{(i)} = 1|x^{(i)}) = \frac{p(y^{(i)} = 1|x^{(i)})}{p(y^{(i)} = 1|t^{(i)} = 1)}}$$

2.b

Mostrar que

$$h(x^{(i)}) \approx \alpha \text{ para todo } x^{(i)} \in V_+$$

dado que,

$$h(x^{(i)}) \approx p(y^{(i)} = 1|x^{(i)})$$

$$h(x^{(i)}) \approx p(t^{(i)} = 1|x^{(i)}) \cdot \alpha$$

Ya que estamos en V_+, $p(t^{(i)} = 1|x^{(i)}) \approx 1$

$$h(x^{(i)}) \approx \alpha$$

2.e

Grafico utilizando labels **t**:

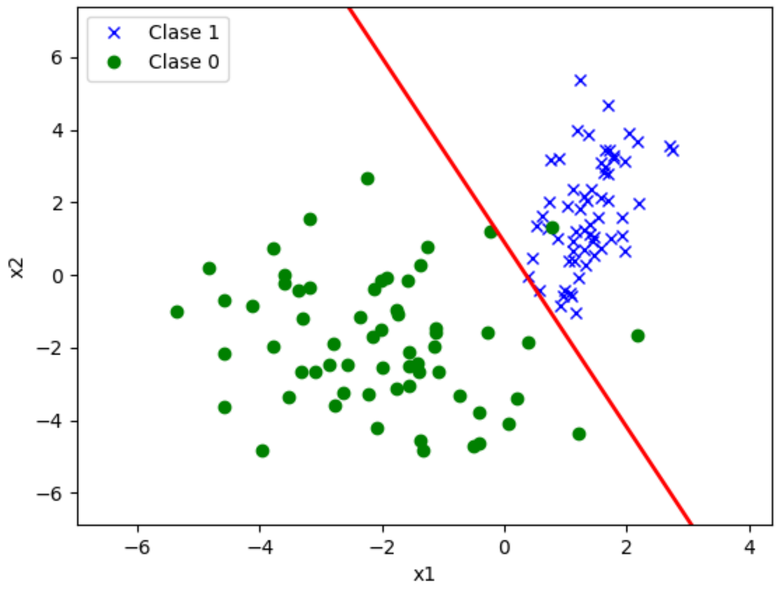


Grafico utilizando labels **y**:

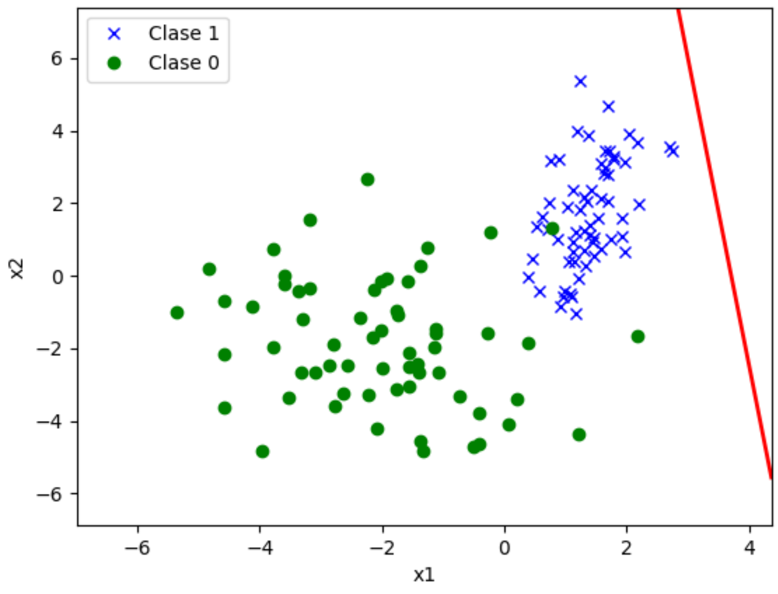
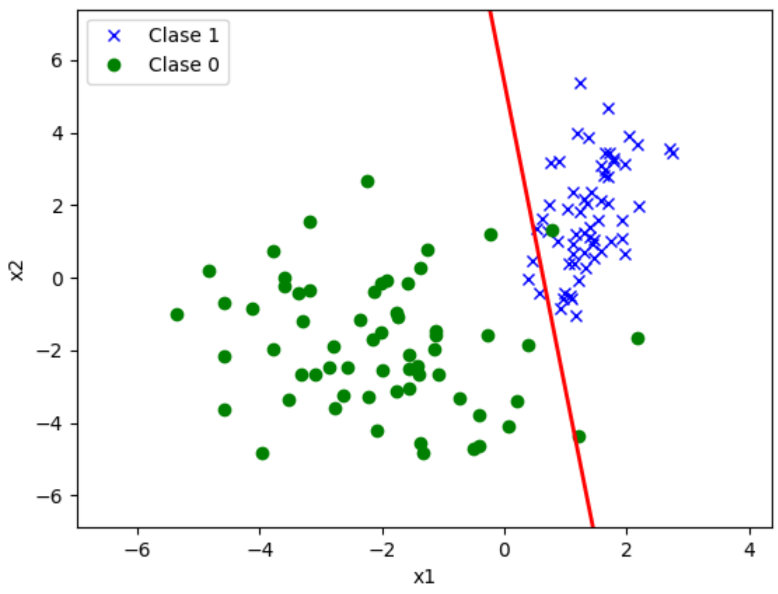


Grafico utilizando labels **y** con constante de correccion alpha:



3.a

Demostrar que la distribucion de Poisson pertenece a la familia de los exponenciales

$$p(y;\lambda) = \frac{e^{-\lambda}\lambda^y}{y!}$$

$$p(y;\lambda) = \frac{1}{y!}e^{-\lambda}\lambda^y = \frac{1}{y!}e^{-\lambda + \ln\lambda^y}$$

$$p(y;\lambda) = \frac{1}{y!}e^{-\lambda + y\ln\lambda}$$

$$\boxed{p(y;\lambda) = \frac{1}{y!}e^{\ln\lambda \cdot y - \lambda}}$$

Entonces,

$$b(y) = \frac{1}{y!}$$

$$T(y) = y$$

$$\eta = ln(\lambda)$$

$$a(\eta) = \lambda = e^{\eta}$$

3.b

Segun la teoria, la funcion de respuesta canónica $g(\eta) = E[y;\eta]$, y dado que una variable aleatoria de Poisson con parametro λ tiene esperanza λ,

$$g(\eta) = E[y;\eta] = \lambda$$

$$g(\eta) = E[y;\eta] = e^{\ln \lambda}$$

$$\boxed{g(\eta) = E[y;\eta] = e^{\eta}}$$

3.c

$$l(\theta) = \frac{1}{m}\sum_{i=1}^{m} \ln\left[p(y^{(i)};\lambda)\right]$$

$$l(\theta) = \frac{1}{m}\sum_{i=1}^{m} \ln\left[\frac{e^{-\lambda}\cdot\lambda^{y^{(i)}}}{y^{(i)}!}\right]$$

$$l(\theta) = \frac{1}{m}\sum_{i=1}^{m} \ln e^{-\lambda} + \ln\lambda^{y^{(i)}} - \ln y^{(i)}!$$

$$l(\theta) = \frac{1}{m}\sum_{i=1}^{m} -\lambda + y^{(i)}\ln\lambda - \ln y^{(i)}!$$

Sabiendo que $\lambda = e^{\theta^T x}$

$$l(\theta) = \frac{1}{m}\sum_{i=1}^{m} -e^{\theta^T x^{(i)}} + y^{(i)}\ln e^{\theta^T x^{(i)}} - \ln y^{(i)}!$$

$$l(\theta) = \frac{1}{m}\sum_{i=1}^{m} -e^{\theta^T x^{(i)}} + y^{(i)}\theta^T x^{(i)} - \ln y^{(i)}!$$

Luego, tomamos la derivada,

$$\frac{\partial l}{\partial\theta} = \frac{1}{m}\sum_{i=1}^{m} -x^{(i)}e^{\theta^T x^{(i)}} + x^{(i)}y^{(i)}$$

$$\frac{\partial l}{\partial\theta} = \frac{1}{m}\sum_{i=1}^{m}\left(-e^{\theta^T x^{(i)}} + y^{(i)}\right)x^{(i)}$$

$$\boxed{\frac{\partial l}{\partial\theta} = \frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - e^{\theta^T x^{(i)}}\right)x^{(i)}}$$

4.a

Sabiendo que $p(y;\eta) = b(y)\cdot exp(\eta\cdot y - \alpha(\eta))$,

$$\int \frac{\partial}{\partial\eta}p(y;\eta)dy = \int \frac{\partial}{\partial\eta}p(y;\eta)dy$$

$$0 = \int \frac{\partial}{\partial\eta}b(y)\cdot e^{\eta\cdot y - \alpha(\eta)}dy$$

$$0 = \int b(y)\cdot e^{\eta\cdot y - \alpha(\eta)}(y - \alpha'(\eta))dy$$

$$0 = \int y\cdot p(y;\eta) - \alpha'(\eta)\cdot p(y;\eta)dy$$

$$0 = \int y\cdot p(y;\eta) - \int \alpha'(\eta)\cdot p(y;\eta)dy$$

$$0 = E[y] - \int \alpha'(\eta) \cdot p(y;\eta)dy$$

$$0 = E[y] - \alpha'(\eta) \cdot \int p(y;\eta)dy$$

$$0 = E[y] - \alpha'(\eta) \cdot 1$$

$$\boxed{\alpha'(\eta) = E[y]}$$

4.b

Siguiendo la logica del punto anterior:

$$\frac{\partial}{\partial^2 \eta} \int p(y;\eta)dy = \int \frac{\partial}{\partial^2 \eta} p(y;\eta)dy$$

$$0 = \int \frac{\partial}{\partial^2 \eta} b(y) \cdot e^{\eta \cdot y - \alpha(\eta)} dy$$

$$0 = \int \frac{\partial}{\partial \eta} b(y) \cdot e^{\eta \cdot y - \alpha(\eta)} (y - \alpha'(\eta))dy$$

$$0 = \int b(y) \cdot e^{\eta \cdot y - \alpha(\eta)} \cdot ((y - \alpha'(\eta))^2 - \alpha''(\eta))dy$$

$$0 = \int (y - \alpha'(\eta))^2 \cdot b(y) \cdot e^{\eta \cdot y - \alpha(\eta)} - \alpha''(\eta) \cdot b(y) \cdot e^{\eta \cdot y - \alpha(\eta)} dy$$

$$0 = \int (y - \alpha'(\eta))^2 \cdot p(y;\eta) - \int \alpha''(\eta) \cdot p(y;\eta)dy$$

$$0 = \int (y - \alpha'(\eta))^2 \cdot p(y;\eta) - \alpha''(\eta) \cdot \int p(y;\eta)dy$$

$$0 = Var(y) - \alpha''(\eta) \cdot 1$$

$$\boxed{\alpha''(\eta) = Var(y)}$$

4.c

$$l(\theta) = -\sum_{i=1}^{m} ln(p(y^{(i)}|x^{(i)};\theta))$$

$$l(\theta) = -\sum_{i=1}^{m} ln\left(b(y) \cdot e^{\theta^T x^{(i)} \cdot y - \alpha(\theta^T x^{(i)})}\right)$$

$$l(\theta) = -\sum_{i=1}^{m} ln(b(y)) + \theta^T x^{(i)} y^{(i)} - \alpha(\theta^T x^{(i)})$$

Luego, derivamos:

$$\frac{\partial l}{\partial \theta} = -\sum_{i=1}^{m} x^{(i)} y^{(i)} - \alpha'(\theta^T x^{(i)}) x^{(i)}$$

$$\boxed{\frac{\partial l}{\partial \theta} = -\sum_{i=1}^{m} \left(y^{(i)} - \alpha'(\theta^T x^{(i)})\right) x^{(i)}}$$

Luego, la segunda derivada:

$$\frac{\partial l}{\partial^2 \theta} = -\sum_{i=1}^{m} -\alpha''(\theta^T x^{(i)}) x^{(i)^T} x^{(i)}$$

$$\boxed{\frac{\partial l}{\partial^2 \theta} = \sum_{i=1}^{m} \alpha''(\theta^T x^{(i)}) x^{(i)^T} x^{(i)}}$$

Probamos que es Semidefinido positivo:

$$(x^T x)^T = x^T x$$

$$\alpha''(\theta^T x^{(i)}) = Var[y; \theta, x] \rightarrow Escalar$$

Entonces,

$$\nabla^2 l(\theta) = (\nabla^2 l(\theta))^T$$

$$z^T \cdot \nabla^2 l(\theta) \cdot z$$

$$z^T \cdot \alpha''(\theta^T x^{(i)}) x^{(i)^T} x^{(i)} \cdot z$$

ya que,

$$z^T \cdot x^{(i)^T} \cdot x^{(i)} \cdot z = (x^T z)^2 \geq 0$$

y que,

$$\alpha''(\theta^T x^{(i)}) = Var[y; \theta, x] \geq 0$$

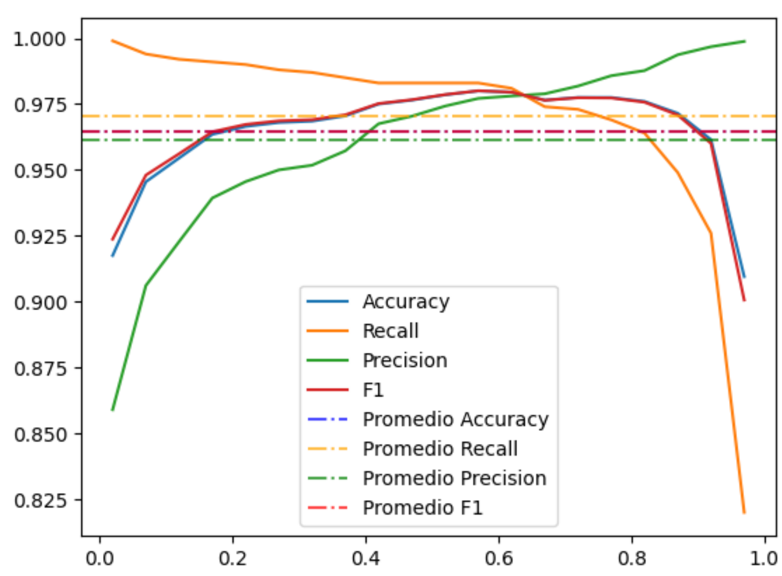
$$\nabla^2 l(\theta) \geq 0$$

5.a

Para analizar este modelo, se utilizaran las siguientes 4 metricas:

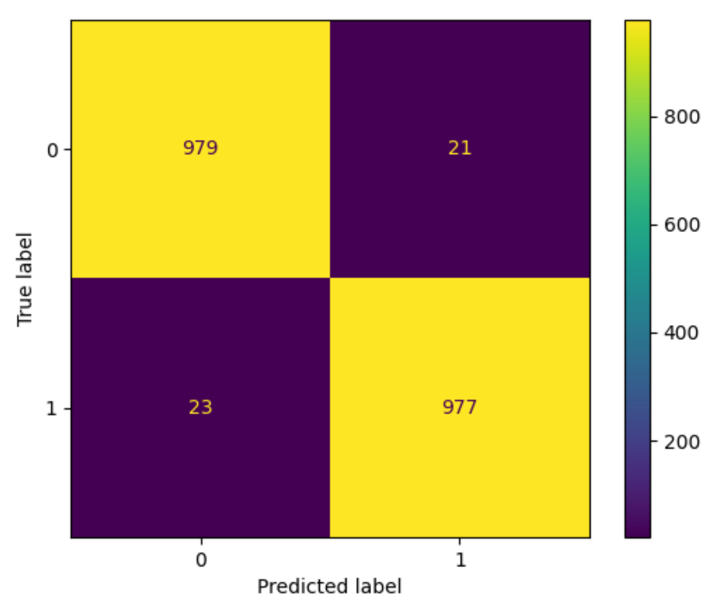
1. Accuracy: Exactitud del modelo (aciertos)
2. Precision: Habilidad del modelo de clasificar como positivo a un caso negativo
3. Recall: Habilidad del modelo de acertar todos los casos positivos
4. F1-Score: Media armonica del Precision y del Recall

Aqui, podemos observar como estas varian segun el punto de corte que sea aplicado a las predicciones:



Se puede notar que, si se busca un equilibrio entre estas metricas, un buen punto de corte podria ser el 0.65. En este, las 4 metricas convergen haciendo que esten lo mas parecidas posibles.

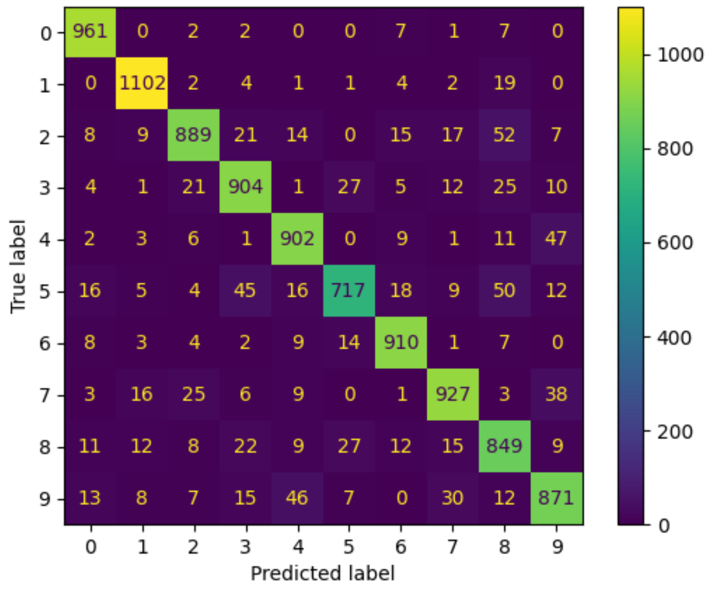
A continuacion, se muestra una matriz de confusion utilizando este punto de corte:



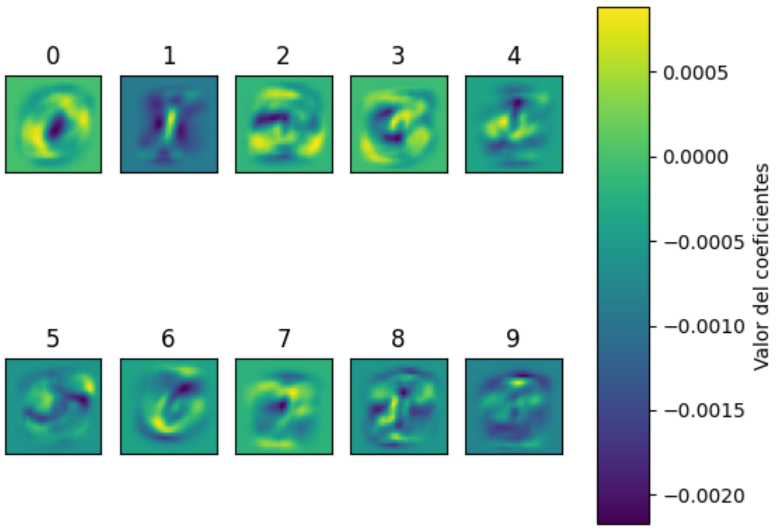
Aqui, se puede ver como la cantidad de falsos-positivos y falsos-negativos es minimizada utilizando el punto de corte mencionado anteriormente.

5.b

Creando y entrenando los 10 modelos diferentes se logra obtener un vector con las 10 predicciones. Y por lo que se puede ver en la siguiente imagen (Matriz de confusión), este modelo se desempeño muy bien a la hora de predecir las imagenes.



Continuando con el analisis, es posible graficar en un mapa de calor los coeficientes de cada modelo, para poder visualizar que "pixeles" o que variable influyen de mayor manera en cada modelo a la hora de predecir el label correspondiente:



Viendo esto, es curioso observar como en los casos de los labels 0, 1, 3 y 6 es muy notorio como se "dibuja" el mismo numero el cual intenta predecir. Sin embargo, con el resto de modelos, es mas confuso intentar saber que numero esta intentando predecir segun sus coeficientes. No esta de mas mencionar que los coeficientes de los modelos 2 y 8 tambien dibujan, aunque de manera mas dificil de comprender, los numeros los cuales intentan predecir.