Ejercicios Tema 3 APR

Nombre: Fabián Scherle Carboneres

3.1. (p.3.9, 0.5 puntos) Obtener los estimadores de máxima verosimiltud del vector media de una gaussiana bi-variada, cuya matriz de covarianza es fija y conocida partir de una muestra de de vectores bi-dimensionales x_1,x_2,...,x_N.

Partiendo de que la matriz de covarianza es fija y conocida: $\Sigma=\begin{pmatrix} o_{1,1}^2 & o_{2,1} \\ o_{12} & o_{2,2}^2 \end{pmatrix}$

$$p(x_1,x_2;u_1,\mu_2)=A\cdot\exp(-B\cdot[\frac{(x_1-\mu_1)^2}{o_{1,1}^2}+\frac{(x_2-\mu_2)^2}{o_{2,2}^2}-\frac{2\cdot o_{1,2}\cdot(x_1-\mu_1)\cdot(x_2-\mu_2)}{(o_{1,1}\cdot o_{2,2})^2}])$$

donde:

$$A=\frac{1}{2\cdot\pi\cdot o_{1,1}\cdot o_{2,2}\cdot\sqrt{1-(\frac{o_{1,2}}{o_{1,1}\cdot o_{2,2}})^2}}\qquad B=\frac{1}{2\cdot(1-(\frac{o_{1,2}}{o_{1,1}\cdot o_{2,2}})^2)}$$

- <u>Obtengo logaritmo de la verosimilitud:</u>

$$Ls(\mu_1,\mu_2)=\log(\prod_{n=1}^{N}(p(x_n;u_1,\mu_2)))$$

$$\rightarrow\log(\prod_{n=1}^{N}(A\cdot\exp(-B\cdot[\frac{(x_{n,1}-\mu_1)^2}{o_{1,1}^2}+\frac{(x_{n,2}-\mu_2)^2}{o_{2,2}^2}-\frac{2\cdot o_{1,2}\cdot(x_{n,1}-\mu_1)\cdot(x_{n,2}-\mu_2)}{(o_{1,1}\cdot o_{2,2})^2}])))$$

$$\rightarrow\log(A^N\cdot\exp(-B\cdot\sum_{n=1}^{N}[\frac{(x_{n,1}-\mu_1)^2}{o_1^2}+\frac{(x_{n,2}-\mu_2)^2}{o_2^2}-\frac{2\cdot o_{12}\cdot(x_{n,1}-\mu_1)\cdot(x_{n,2}-\mu_2)}{(o_1\cdot o_2)^2}]))$$

$$\rightarrow\log(A^N)+\log(\exp(-B\cdot\sum_{n=1}^{N}[\frac{(x_{n,1}-\mu_1)^2}{o_{1,1}^2}+\frac{(x_{n,2}-\mu_2)^2}{o_{2,2}^2}-\frac{2\cdot o_{1,2}\cdot(x_{n,1}-\mu_1)\cdot(x_{n,2}-\mu_2)}{(o_{1,1}\cdot o_{2,2})^2}]))$$

$$\rightarrow\log(A^N)+(-B\cdot\sum_{n=1}^{N}[\frac{(x_{n,1}-\mu_1)^2}{o_{1,1}^2}+\frac{(x_{n,2}-\mu_2)^2}{o_{2,2}^2}-\frac{2\cdot o_{1,2}\cdot(x_{n,1}-\mu_1)\cdot(x_{n,2}-\mu_2)}{(o_{1,1}\cdot o_{2,2})^2}])$$

$$\rightarrow-N\cdot\log(A^{-1})-B\cdot\sum_{n=1}^{N}[\frac{(x_{n,1}-\mu_1)^2}{o_{1,1}^2}+\frac{(x_{n,2}-\mu_2)^2}{o_{2,2}^2}-\frac{2\cdot o_{1,2}\cdot(x_{n,1}-\mu_1)\cdot(x_{n,2}-\mu_2)}{(o_{1,1}\cdot o_{2,2})^2}]$$

$$\rightarrow-N\cdot\log(2\cdot\pi\cdot o_{1,1}\cdot o_{2,2}\cdot\sqrt{1-(\frac{o_{1,2}}{o_{1,1}\cdot o_{2,2}})^2})-(\frac{1}{2\cdot(1-(\frac{o_{1,2}}{o_{1,1}\cdot o_{2,2}})^2)})\cdot\sum_{n=1}^{N}[\frac{(x_{n,1}-\mu_1)^2}{o_{1,1}^2}+\frac{(x_{n,2}-\mu_2)^2}{o_{2,2}^2}-\frac{2\cdot o_{1,2}\cdot(x_{n,1}-\mu_1)\cdot(x_{n,2}-\mu_2)}{(o_{1,1}\cdot o_{2,2})^2}]$$

- <u>Obtengo la derivada parcial en función de **µ1:**</u>

$$\frac{\partial}{\partial\mu 1}=0-(\frac{1}{2\cdot(1-(\frac{o_{1,2}}{o_{1,1}\cdot o_{2,2}})^2)})\cdot\sum_{n=1}^{N}[\frac{2\cdot(x_{n,1}-\mu_1)}{o_{1,1}^2}+0-\frac{2\cdot o_{1,2}\cdot(0-1)\cdot(x_{n,2}-\mu_2)}{(o_{1,1}\cdot o_{2,2})^2}]$$

$$\rightarrow(\frac{1}{2\cdot(1-(\frac{o_{1,2}}{o_{1,1}\cdot o_{2,2}})^2)})\cdot(\frac{2}{o_{1,1}^2})\cdot(\sum_{n=1}^{N}(x_{n,1}-\mu_1)+(\frac{o_{1,2}}{o_{2,2}^2})\cdot\sum_{n=1}^{N}(x_{n,2}-\mu_2))$$

Igualo a 0:

$$0=\left(\frac{1}{2\cdot\left(1-\left(\frac{o_{1,2}}{o_{1,1}\cdot o_{2,2}}\right)^2\right)}\right)\cdot\left(\frac{2}{o_{1,1}^2}\right)\cdot\left(\sum_{n=1}^{N}(x_{n,1}-\mu_1)+\left(\frac{o_{1,2}}{o_{2,2}^2}\right)\cdot\sum_{n=1}^{N}(x_{n,2}-\mu_2)\right)$$

Se puede apreciar que ni $\left(\frac{1}{2\cdot\left(1-\left(\frac{o_{1,2}}{o_{1,1}\cdot o_{2,2}}\right)^2\right)}\right)$ ni $\left(\frac{2}{o_{1,1}^2}\right)$ valdrán 0 por lo que se eliminan de la ecuación.

$$0=\sum_{n=1}^{N}(x_{n,1}-\mu_1)+\left(\frac{o_{1,2}}{o_{2,2}^2}\right)\cdot\sum_{n=1}^{N}(x_{n,2}-\mu_2)$$

$$0=\sum_{n=1}^{N}(x_{n,1}-\mu_1)+\left(\frac{o_{1,2}}{o_{2,2}^2}\right)\cdot\sum_{n=1}^{N}(x_{n,2}-\mu_2)$$

$$0=\sum_{n=1}^{N}x_{n,1}-\sum_{n=1}^{N}\mu_1+\left(\frac{o_{1,2}}{o_{2,2}^2}\right)\cdot\sum_{n=1}^{N}x_{n,2}-\left(\frac{o_{1,2}}{o_{2,2}^2}\right)\cdot\sum_{n=1}^{N}\mu_2$$

$$0=\sum_{n=1}^{N}x_{n,1}-N\cdot\mu_1+\left(\frac{o_{1,2}}{o_{2,2}^2}\right)\cdot\sum_{n=1}^{N}x_{n,2}-\left(\frac{o_{1,2}}{o_{2,2}^2}\right)\cdot N\cdot\mu_2$$

- Obtengo la derivada parcial en función de **μ2:**

$$\frac{\partial}{\partial\mu_2}=0-\left(\frac{1}{2\cdot\left(1-\left(\frac{o_{1,2}}{o_{1,1}\cdot o_{2,2}}\right)^2\right)}\right)\cdot\sum_{n=1}^{N}\left[0+\frac{2\cdot(x_{n,2}-\mu_2)}{o_{1,1}^2}-\frac{2\cdot o_{1,2}\cdot(x_{n,1}-\mu_1)\cdot(0-1)}{(o_{1,1}\cdot o_{2,2})^2}\right]$$

$$\frac{\partial}{\partial\mu_1}=\left(\frac{1}{2\cdot\left(1-\left(\frac{o_{1,2}}{o_{1,1}\cdot o_{2,2}}\right)^2\right)}\right)\cdot\left(\frac{2}{o_{1,1}^2}\right)\cdot\left(\sum_{n=1}^{N}(x_{n,2}-\mu_2)+\left(\frac{o_{1,2}}{o_{2,2}^2}\right)\cdot\sum_{n=1}^{N}(x_{n,1}-\mu_1)\right)$$

Igualo a 0:

$$0=\left(\frac{1}{2\cdot\left(1-\left(\frac{o_{1,2}}{o_{1,1}\cdot o_{2,2}}\right)^2\right)}\right)\cdot\left(\frac{2}{o_{1,1}^2}\right)\cdot\left(\sum_{n=1}^{N}(x_{n,2}-\mu_2)+\left(\frac{o_{1,2}}{o_{2,2}^2}\right)\cdot\sum_{n=1}^{N}(x_{n,1}-\mu_1)\right)$$

Se puede apreciar al igual que antes que ni $\left(\frac{1}{2\cdot\left(1-\left(\frac{o_{1,2}}{o_{1,1}\cdot o_{2,2}}\right)^2\right)}\right)$ ni $\left(\frac{2}{o_{1,1}^2}\right)$ valdrán 0 por lo que se eliminan de la ecuación.

$$0=\sum_{n=1}^{N}(x_{n,2}-\mu_2)+\left(\frac{o_{1,2}}{o_{2,2}^2}\right)\cdot\sum_{n=1}^{N}(x_{n,1}-\mu_1)$$

$$0=\sum_{n=1}^{N}(x_{n,2}-\mu_2)+(\frac{o_{1,2}}{o_{2,2}^2})\cdot\sum_{n=1}^{N}(x_{n,1}-\mu_1)$$

$$0=\sum_{n=1}^{N}x_{n,2}-\sum_{n=1}^{N}\mu_2+(\frac{o_{1,2}}{o_{2,2}^2})\cdot\sum_{n=1}^{N}x_{n,1}-(\frac{o_{1,2}}{o_{2,2}^2})\cdot\sum_{n=1}^{N}\mu_1$$

$$0=\sum_{n=1}^{N}x_{n,2}-N\cdot\mu_2+(\frac{o_{1,2}}{o_{2,2}^2})\cdot\sum_{n=1}^{N}x_{n,1}-(\frac{o_{1,2}}{o_{2,2}^2})\cdot N\cdot\mu_1$$

- Resuelvo el sistema de ecuaciones:

1) $$0=\sum_{n=1}^{N}x_{n,1}-N\cdot\mu_1+(\frac{o_{1,2}}{o_{2,2}^2})\cdot\sum_{n=1}^{N}x_{n,2}-(\frac{o_{1,2}}{o_{2,2}^2})\cdot N\cdot\mu_2$$

2) $$0=\sum_{n=1}^{N}x_{n,2}-N\cdot\mu_2+(\frac{o_{1,2}}{o_{2,2}^2})\cdot\sum_{n=1}^{N}x_{n,1}-(\frac{o_{1,2}}{o_{2,2}^2})\cdot N\cdot\mu_1$$

De la primera ecuación despejo **μ1:**

$$N\cdot\mu_1=\sum_{n=1}^{N}x_{n,1}+(\frac{o_{1,2}}{o_{2,2}^2})\cdot\sum_{n=1}^{N}x_{n,2}-(\frac{o_{1,2}}{o_{2,2}^2})\cdot N\cdot\mu_2$$

$$\mu_1=\frac{1}{N}\cdot\sum_{n=1}^{N}x_{n,1}+(\frac{o_{1,2}}{N\cdot o_{2,2}^2})\cdot\sum_{n=1}^{N}x_{n,2}-(\frac{o_{1,2}}{N\cdot o_{2,2}^2})\cdot N\cdot\mu_2$$

Y sustituyo en la segunda ecuación:

$$0=\sum_{n=1}^{N}x_{n,2}-N\cdot\mu_2+(\frac{o_{1,2}}{o_{2,2}^2})\cdot\sum_{n=1}^{N}x_{n,1}-(\frac{o_{1,2}}{o_{2,2}^2})\cdot N\cdot(\frac{1}{N}\cdot\sum_{n=1}^{N}x_{n,1}+(\frac{o_{1,2}}{N\cdot o_{2,2}^2})\cdot\sum_{n=1}^{N}x_{n,2}-(\frac{o_{1,2}}{N\cdot o_{2,2}^2})\cdot N\cdot\mu_2)$$

$$0=\sum_{n=1}^{N}x_{n,2}-N\cdot\mu_2-(\frac{o_{1,2}}{o_{2,2}^2})^2\cdot(\sum_{n=1}^{N}x_{n,2}-N\cdot\mu_2)$$

$$0=(\sum_{n=1}^{N}x_{n,2}-N\cdot\mu_2)\cdot(1-(\frac{o_{1,2}}{o_{2,2}^2})^2)$$

$$0=(\sum_{n=1}^{N}x_{n,2}-N\cdot\mu_2)$$

$$\mu_2=\frac{\sum_{n=1}^{N}x_{n,2}}{N}$$

A partir de **μ2** sustituyo en la segunda ecuación:

$$0=\sum_{n=1}^{N}x_{n,2}-N\cdot\frac{\sum_{n=1}^{N}x_{n,2}}{N}+(\frac{o_{1,2}}{o_{2,2}^2})\cdot\sum_{n=1}^{N}x_{n,1}-(\frac{o_{1,2}}{o_{2,2}^2})\cdot N\cdot\mu_1$$

$$0=\sum_{n=1}^{N}x_{n,2}-\sum_{n=1}^{N}x_{n,2}+(\frac{o_{1,2}}{o_{2,2}^2})\cdot\sum_{n=1}^{N}x_{n,1}-(\frac{o_{1,2}}{o_{2,2}^2})\cdot N\cdot\mu_1$$

$$0=(\frac{o_{1,2}}{o_{2,2}^2})\cdot\sum_{n=1}^{N}x_{n,1}-(\frac{o_{1,2}}{o_{2,2}^2})\cdot N\cdot\mu_1$$

$$0=\sum_{n=1}^{N} x_{n,1}-N\cdot\mu_1$$

$$\mu_1=\frac{\sum_{n=1}^{N} x_{n,1}}{N}$$

3.3. (p.3.10, 0.25 puntos) Desarrollar y explicar con detalle todos los pasos necesarios para obtener el estimador de máxima verosimiltud del vector media de una de una gaussiana multivariada cuya matriz de covarianza es fija y conocida, a partir de una muestra de vectores x_1,x_2,...,x_N.

$$p(x,\theta)=(2\cdot\pi)^{-\frac{D}{2}}\cdot|\Sigma|^{-\frac{1}{2}}\cdot\exp(-\frac{1}{2}\cdot(\vec{x}-\vec{\mu})^{t}\cdot\Sigma^{-1}\cdot(\vec{x}-\vec{\mu}))$$

<u>Aplico logaritmo de la verosimilitud:</u>

$$Ls(\theta)=\log(\prod_{n=1}^{N}((2\cdot\pi)^{-\frac{D}{2}}\cdot|\Sigma|^{-\frac{1}{2}}\cdot\exp(-\frac{1}{2}\cdot(\vec{x}_n-\vec{\mu})^{t}\cdot\Sigma^{-1}\cdot(\vec{x}_n-\vec{\mu}))))$$

$$\rightarrow\log((2\cdot\pi)^{-\frac{D\cdot N}{2}}\cdot|\Sigma|^{-\frac{1\cdot N}{2}}\cdot\exp(\sum_{n=1}^{N}(-\frac{1}{2}\cdot(\vec{x}_n-\vec{\mu})^{t}\cdot\Sigma^{-1}\cdot(\vec{x}_n-\vec{\mu}))))$$

$$\rightarrow-\frac{D\cdot N}{2}\cdot\log(2\cdot\pi)-\frac{1\cdot N}{2}\cdot\log(|\Sigma|)+\sum_{n=1}^{N}(-\frac{1}{2}\cdot(\vec{x}_n-\vec{\mu})^{t}\cdot\Sigma^{-1}\cdot(\vec{x}_n-\vec{\mu}))$$

$$\rightarrow-\frac{D\cdot N}{2}\cdot\log(2\cdot\pi)-\frac{N}{2}\cdot\log(|\Sigma|)-\frac{1}{2}\cdot\sum_{n=1}^{N}((\vec{x}_n-\vec{\mu})^{t}\cdot\Sigma^{-1}\cdot(\vec{x}_n-\vec{\mu}))$$

<u>Aplico la derivación en función de **μ**:</u>

$$\frac{\partial}{\partial\mu}=-\frac{1}{2}\cdot\sum_{n=1}^{N}(2\cdot\Sigma^{-1}\cdot(\vec{x}_n-\vec{\mu}))$$

Al ser Σ simétrica su inversa también lo es y al (x-μ) no depender de dicha omversa se puede obtener el gradiente de forma cuadrática.

$$\rightarrow-\sum_{n=1}^{N}(\Sigma^{-1}\cdot(\vec{x}_n-\vec{\mu}))$$

<u>A continuación igualo a 0:</u>

$$0=-\sum_{n=1}^{N}(\Sigma^{-1}\cdot(\vec{x}_n-\vec{\mu}))$$

$$0=\sum_{n=1}^{N}(\Sigma^{-1}\cdot(\vec{x}_n-\vec{\mu}))$$

$$0=\Sigma^{-1}\cdot\sum_{n=1}^{N}(\vec{x}_n-\vec{\mu})$$

Debido a que Σ es fija:

$$0=\sum_{n=1}^{N}(\vec{x}_n-\vec{\mu})$$

$$0=\sum_{n=1}^{N}\vec{x}_n-\sum_{n=1}^{N}\vec{\mu}$$

$$0=\sum_{n=1}^{N}\vec{x}_n-N\cdot\vec{\mu}$$

$$N\cdot\vec{\mu}=\sum_{n=1}^{N}\vec{x}_n$$

$$N\cdot\vec{\mu}=\sum_{n=1}^{N}\vec{x}_n$$

$$\vec{\mu}=\frac{\sum_{n=1}^{N}\vec{x}_n}{N}$$

3.4. (p.3.16, 0.25 puntos) Ejercicio b) de la página indicada: minimizar una función con una condición de igualdad.

Función a minimzar: $q(\theta)=1+(\theta-2)^2$

Restricción (**v(θ)**): $q(\theta)+\theta=4 \rightarrow q(\theta)+\theta-4=0$

- **Primera Alternativa:**

Calculo el valor de **θ**:

$q(\theta)+\theta-4=0 \rightarrow (\theta-2)^2+\theta-3=0 \rightarrow \theta^2-4\cdot\theta+4+\theta-3=0 \rightarrow \theta^2-3\cdot\theta+1=0$

De tal forma que la restricción se cumple para:

$\theta=2.61$ y $\theta=0.38$

$q(2.61)=1+(2.61-2)^2=1.37$
$q(0.38)=1+(0.38-2)^2=3.62$

Pero aquel que minimiza la función es $\theta=2.61$

- **Aplicando Multiplicadores de Lagrange:**

$$\Lambda(\theta,\beta)=q(\theta)+\beta\cdot v(\theta)$$
$$\rightarrow q(\theta)+\beta\cdot(q(\theta)+\theta-4)$$
$$\rightarrow(1+(\theta-2)^2)+\beta\cdot((1+(\theta-2)^2)+\theta-4)$$

$$\rightarrow 1+(\theta-2)^2+\beta\cdot((\theta-2)^2+\theta-3)$$

Obtengo la deriva parcial en función de **θ** `e igualo a 0`:

$$\frac{\partial}{\partial\theta}=0+2\cdot(\theta-2)+\beta\cdot(2\cdot(\theta-2)+1-0)$$

$$\rightarrow 2\cdot(\theta-2)+\beta\cdot(2\cdot(\theta-2)+1)$$

$$0=2\cdot(\theta-2)+\beta\cdot(2\cdot(\theta-2)+1)$$

$$3\cdot\beta+4=\theta\cdot(2+2\cdot\beta)$$

$$\theta*(\beta)=\frac{3\cdot\beta+4}{(2+2\cdot\beta)}$$

Obtengo Λ_D(`beta`):

$$\Lambda_D(\beta)=\Lambda(\theta*(\beta),\beta)=1+((\frac{3\cdot\beta+4}{(2+2\cdot\beta)})-2)^2+\beta\cdot(((\frac{3\cdot\beta+4}{(2+2\cdot\beta)})-2)^2+(\frac{3\cdot\beta+4}{(2+2\cdot\beta)})-3)$$
$$\rightarrow 1+((\frac{3\cdot\beta+4}{(2+2\cdot\beta)})-2)^2+\beta\cdot((\frac{3\cdot\beta+4}{(2+2\cdot\beta)})-2)^2+\beta\cdot(\frac{3\cdot\beta+4}{(2+2\cdot\beta)})-3\cdot\beta$$

Obtengo la deriva parcial en función de **α** `e igualo a 0`:

$$\frac{\partial}{\partial\beta}=0+2\cdot((\frac{3\cdot\beta+4}{(2+2\cdot\beta)})-2)\cdot(\frac{-1}{2\cdot(\beta+1)^2})+((\frac{3\cdot\beta+4}{(2+2\cdot\beta)})-2)^2+\frac{\beta^2}{(2+2\cdot\beta)\cdot(\beta+1)}+(\frac{3\cdot\beta+4}{(2+2\cdot\beta)})-\frac{\beta}{2\cdot(\beta+1)^2}-3$$

$$\rightarrow\frac{\beta}{(2+2\cdot\beta)\cdot(\beta+1)^2}+\frac{\beta^3+3\cdot\beta^2}{2\cdot(2+2\cdot\beta)\cdot(\beta+1)^2}+\frac{3\cdot\beta^2+6\cdot\beta+4}{2\cdot(\beta+1)^2}-3$$
$$\rightarrow\frac{7\cdot\beta^2+14\cdot\beta+8}{4\cdot(\beta+1)^2}-3$$

$$0=\frac{7\cdot\beta^2+14\cdot\beta+8}{4\cdot(\beta+1)^2}-3$$
$$0=\frac{7\cdot\beta^2+14\cdot\beta+8-3\cdot(4\cdot(\beta+1)^2)}{4\cdot(\beta+1)^2}$$
$$0=7\cdot\beta^2+14\cdot\beta+8-12\cdot(\beta+1)^2$$

$$0=5\cdot\beta^2+10\cdot\beta+4$$

$$\beta = -1.45$$

$$\beta = -0.55$$

Sustituyo el valor de **`beta`** en **`θ*`**:

Para **`beta=-1.45:`**

$$\theta*(-1.45) = \frac{3\cdot(-1.45)+4}{(2+2\cdot(-1.45))}$$

$$\theta*(-1.45) = 0.38$$

Para **`beta=-0.55:`**

$$\theta*(-0.55) = \frac{3\cdot(-0.55)+4}{(2+2\cdot(-0.55))}$$

$$\theta*(-0.55) = 2.61$$

Verifico que valor obtenido minimiza la función:

$$q(2.61) = 1+(2.61-2)^2 = 1.37$$
$$q(0.38) = 1+(0.38-2)^2 = 3.62$$

Aquel que minimiza la función es $\theta = 2.61$

3.5. (p.3.18, 0.5 puntos) Ejercicio al pie de la página indicada: estimación por máxima verosimilitud de las probabilidades a priori de un clasificador genérico en C clases.

Modelo: $P(c=C) = p_c$

Verosimilitud: $P(S|\theta) = \prod_{c=1}^{C} p_c^{N_c}$ (Casi no se ve, pero donde pone N se refiere a N_c)

Logaritmo Verosimilitud: $q_s(\theta) = L_s(\theta) = \log(P(S|\theta)) = \sum_{c=1}^{C} N_c \cdot \log(p_c)$

Estimación de la máxima verosimilitud:

θ*= $argmax_\theta L_s(\theta) = argmax_{\sum_{c=1}^{c} pc=1} (\sum_{c=1}^{C} N_c \cdot \log(p_c))$

(No se ve pero en el argmax hay un sumatorio desde c=1 hasta C)

Aplico multiplicadores de Lagrange, derivo en función de **pc** e igualo a 0:

$$\Lambda(p_1 \cdots P_n, \beta) = \sum_{c=1}^{C} N_c \cdot \log(p_c) + \beta \cdot (1 - \sum_{c=1}^{C} p_c)$$

$$\frac{\partial}{\partial p_c} = (\frac{N_c}{p_c}) - \beta$$

$$0=(\frac{N_c}{p_c})-\beta$$

$$\beta=\frac{N_c}{p_c}$$

$$p_c(\beta)=\frac{N_c}{\beta}$$

Uso Función dual de Lagrange sustituyendo el resultado de **pc**, derivo sobre **beta** e igualo a 0:

$$\Lambda_D(\beta)=\sum_{c=1}^{C} N_c\cdot\log(\frac{N_c}{\beta})+\beta\cdot(1-\sum_{c=1}^{C}\frac{N_c}{\beta})$$

$$\Lambda_D(\beta)=\sum_{c=1}^{C} N_c\cdot\log(N_c)-\sum_{c=1}^{C} N_c\cdot\log(\beta)+\beta-\frac{\beta}{\beta}\cdot\sum_{c=1}^{C} N_c$$

$$\Lambda_D(\beta)=\sum_{c=1}^{C} N_c\cdot\log(N_c)-\sum_{c=1}^{C} N_c\cdot\log(\beta)+\beta-\sum_{c=1}^{C} N_c$$

$$\frac{\partial}{\partial\beta}=0-\sum_{c=1}^{C} N_c\cdot\frac{1}{\beta}+1-0$$

$$0=\frac{N}{\beta}+1$$

$$N=\beta$$

Sustituyo **beta:**

p_n*(**beta** *) = N_c / N

3.7. (p.3.26, 0.5 puntos) Mostrar la traza de tres iteraciones de descenso por gradiente para minimizar la función indicada usando paso decreciente con el número de iteraciones, k (por ejemplo \rho_k = 1/(2k)).

Función a minimizar: $q(\theta)=(\theta_1-1)^2+(\theta_2-2)^2+\theta_1\cdot\theta_2$

Consideraciones: $p(k)=\frac{1}{2\cdot k}$ $\theta_1=(-1,+1)$

Aplico las derivadas parciales:

$$\frac{\partial}{\partial\theta_1}=2\cdot\theta_1-2+\theta_2$$

$$\frac{\partial}{\partial\theta_2}=2\cdot\theta_2-4+\theta_1$$

Sustituyo en la fórmula del algoritmo:

$$\theta(k+1)=\theta(k)-pk\cdot\nabla\cdot q(\theta)|\theta=\theta(k)$$

$$\rightarrow \theta(k)-\left(\frac{1}{2\cdot k}\right)\cdot\begin{pmatrix}2\cdot\theta_1-2+\theta_2\\ 2\cdot\theta_2-4+\theta_1\end{pmatrix}$$

Primera iteración:

$$\theta(2)=\theta(1)-\left(\frac{1}{2\cdot 1}\right)\cdot\begin{pmatrix}2\cdot\theta_1-2+\theta_2\\ 2\cdot\theta_2-4+\theta_1\end{pmatrix}$$

$$\rightarrow\begin{pmatrix}-1\\ +1\end{pmatrix}-\left(\frac{1}{2}\right)\cdot\begin{pmatrix}2\cdot(-1)-2+1\\ 2\cdot 1-4-1\end{pmatrix}$$

$$\rightarrow\begin{pmatrix}-1\\ +1\end{pmatrix}+\begin{pmatrix}3/2\\ 3/2\end{pmatrix}$$

$$\theta(2)=\begin{pmatrix}1/2\\ 5/2\end{pmatrix}$$

Segunda iteración:

$$\theta(3)=\begin{pmatrix}1/2\\ 5/2\end{pmatrix}-\left(\frac{1}{2\cdot 2}\right)\cdot\begin{pmatrix}2\cdot(1/2)-2+(5/2)\\ 2\cdot(5/2)-4+(1/2)\end{pmatrix}$$

$$\theta(3)=\begin{pmatrix}1/8\\ 17/8\end{pmatrix}$$

Tercera iteración:

$$\theta(4)=\begin{pmatrix}1/8\\ 17/8\end{pmatrix}-\left(\frac{1}{2\cdot 3}\right)\cdot\begin{pmatrix}2\cdot(1/8)-2+(17/8)\\ 2\cdot(17/8)-4+(1/8)\end{pmatrix}$$

$$\theta(4)=\begin{pmatrix}1/16\\ 33/16\end{pmatrix}$$