Colección de Ejercicios Propuestos y sus resoluciones

Sebastian Elizalde
Padrón Nro. 96092
sebi.elizalde@gmail.com

Julian Ferres
Padrón Nro. 101483
julianferres@gmail.com

2do. Cuatrimestre de 2018

Aprendizaje Estadístico - Sebastian Grymberg

Facultad de Ingeniería, Universidad de Buenos Aires

Contents

1 Introducción 4

2 Lista de enunciados 5

2.1 Clase 1 5
2.1.1 Ejercicio 1 5
2.1.2 Ejercicio 2: Esperanza condicional 5
2.1.3 Ejercicio 3 (SIMULACIÓN) 5
2.1.4 Ejercicio 4 (no entregable) 5
2.2 Clase 2 6
2.2.1 Ejercicio 5 6
2.2.2 Ejercicio 6 6
2.3 Clase 3 6
2.3.1 Ejercicio 7 6
2.3.2 Ejercicio 8: PARTICIONES (SIMULACIÓN) 6
2.4 Clase 4 7
2.4.1 Ejercicio 9: KNN (SIMULACIÓN) 7
2.4.2 Ejercicio 9 bis 8
2.4.3 Ejercicio 10 8
2.5 Clase 5 8
2.5.1 Ejercicio 11 8
2.5.2 Ejercicio 12 8
2.5.3 Ejercicio 13 8
2.5.4 Ejercicio 14: PERDIDA (SIMULACIÓN) 9
2.6 Clase 6 9
2.6.1 Ejercicio 15 9
2.6.2 Ejercicio 16 9

3 Soluciones 10

3.1 Clase 1 10
3.1.1 Ejercicio 1 10
3.1.2 Ejercicio 2: Esperanza condicional 10
3.1.3 Ejercicio 4 (no entregable) 11
3.2 Clase 2 11
3.2.1 Ejercicio 5 11
3.2.2 Ejercicio 6 11
3.3 Clase 3 12
3.3.1 Ejercicio 7 12
3.4 Clase 4 12
3.4.1 Ejercicio 9 bis 12
3.4.2 Ejercicio 10 13
3.5 Clase 5 13
3.5.1 Ejercicio 11 13
3.5.2 Ejercicio 12 14
3.5.3 Ejercicio 13 14
3.6 Clase 6 15
3.6.1 Ejercicio 15 15
3.6.2 Ejercicio 16 15

4 Soluciones de problemas con Simulación **16**
4.1 Ejercicio 3 (SIMULACIÓN) 16
4.2 Ejercicio 8: PARTICIONES (SIMULACIÓN) 17
4.3 Ejercicio 9: KNN (SIMULACIÓN) 20
4.3.1 Costos de KNN 23
4.4 Ejercicio 14: PERDIDA (SIMULACIÓN) 23

1 Introducción

En este informe se busca recopilar los diferentes ejercicios propuestos en la asignatura *Aprendizaje Estadístico* en el 2^{do} cuatrimestre de 2018 por el profesor Sebastian Grymberg, junto con las soluciones propuestas por los integrantes del grupo.

Serán excibidos en primer lugar aquellos ejercicios que no hayan requerido simulaciones, para luego dar paso a los mismos.

Toda la información, el codigo utilizado y la versión digital del informe se encuentran en:

https://github.com/julianferres/Aprendizaje-Estadistico

2 Lista de enunciados

2.1 Clase 1

2.1.1 Ejercicio 1

Sea una función convexa $f(x)$ no negativa, x_0, x_1, x_2 tales que:

- $x_1 \leq x_0 \leq x_2$
- $x_0 = p_1x_1 + p_2x_2$
- $p_1 \geq 0, p_2 \geq 0, p_1 + p_2 = 1$

Y sea una recta g tal que:

- $g(x_1) = f(x_1)$
- $g(x_2) = f(x_2)$

Comprobar la siguiente desigualdad:

$$f(p_1x_1 + p_1x_1) \leq p_1 f(x_1) + p_2 f(x_2)$$

2.1.2 Ejercicio 2: Esperanza condicional

Utilizando que $\forall f(X) \in \mathcal{H} : E[m(x)f(x)] = E[Yf(x)]$ y $m(x) = E[Y|X = x]$ sii $Y - m(x) \perp \mathcal{H}$, con $\mathcal{H}$ el espacio de todas las funciones de X.

Demostrar que vale: $E[\phi(x)Y|X] = \phi(x)E[Y|X]$

2.1.3 Ejercicio 3 (SIMULACIÓN)

Sea $X \sim \mathcal{N}(0,\,1)$ truncada al intervalo $[-1, 1]$
Imagine $m(x) = E[Y|X = x]$ como:

$$m(x) := \begin{cases} \frac{(x+2)^2}{2} & \text{si } si - 1 \leq x < -0.5 \\ \frac{x}{2} + 0.875 & \text{si } -0.5 \leq x \leq 0 \\ -5(x-0.2)^2 + 1.075 & \text{si } 0 < x \leq 0.5 \\ x + 0.125 & \text{si } 0.5 \leq x < 1 \end{cases}$$

Dado un x, la distribución condicional de $Y - m(x)$ es $\mathcal{N}(0,\,\sigma^2(x))$
Con $\sigma(x) = 0.2 - 0.1\cos(2x)$

Se pide simular 200 puntos (X, Y), y graficarlos en un plano.
Además, se piden los 200 pares ordenados en cuestión, para hacer análisis posteriores.

2.1.4 Ejercicio 4 (no entregable)

Dado el problema de decisión introducido en el 'Diseño del receptor de una comunicación binaria', verificar que:
$\delta(r) = \mathbb{1}\{P(S = 1|R = r) > P(S = 0|R = r)\}$
maximiza:
$P(S = \delta(r)) = P(S = \delta(0)|R = 0)P(R = 0) + P(S = \delta(1)|R = 1)P(R = 1)$

2.2 Clase 2

2.2.1 Ejercicio 5

Terminar la cuenta:

$$P(X,Y) = \int_0^{\frac{1}{2}} (1-x)dx + \int_{\frac{1}{2}}^{\frac{3}{4}} xdx$$

2.2.2 Ejercicio 6

Sean T, F y E exponenciales de intensidad 1 y tenemos que

$$Y = \mathbb{1}\{T + F + E < 7\}$$

Se tenia que E era inobservable. Y se concluyó que:

$$g^*(T,F) = \mathbb{1}\{T + F < 7 - \ln 2\}$$

Se pide calcular el error:

$$L^* = P(g * (T,F) \neq Y)$$

2.3 Clase 3

2.3.1 Ejercicio 7

Se quería acotar:

$$E\left[|\tilde{\eta}(x) - \eta(x)|\right] < \epsilon$$

Para $\epsilon > 0$ existe una función η_ϵ uniformemente continua a valores en el intervalo $[0,1]$ sobre un conjunto compacto C que se anula fuera de C, y que tiene la siguiente propiedad: $E[|\eta_\epsilon(x) - \eta(x)|] < \epsilon$
Por la desigualdad triangular:

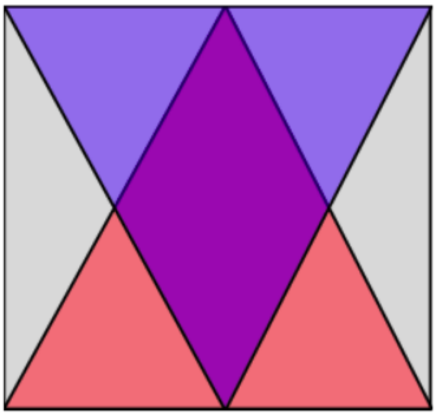
$$E\left[|\tilde{\eta}(x) - \eta(x)|\right] \leq E[\tilde{\eta}(x) - \tilde{\eta}_\epsilon(x)|] + E[|\tilde{\eta}_\epsilon(x) - \eta_\epsilon(x)|] + E[|\eta_\epsilon(x) - \eta(x)|]$$

En clase se acotaron tanto el segundo como el tercer miembro la desigualdad, y el ejercicio es entender por qué el primer término está acotado por el tercero, es decir:

$$E[\tilde{\eta}(x) - \tilde{\eta}_\epsilon(x)|] \leq E[|\eta_\epsilon(x) - \eta(x)|]$$

2.3.2 Ejercicio 8: PARTICIONES (SIMULACIÓN)

Dado el siguiente diagrama:



Sean:

X una variable aleatoria con distribución uniforme sobre el triangulo **azul**.

Y una variable aleatoria con distribución uniforme sobre el triangulo **rojo**.

Z se define como: $Z = \frac{1}{2}X + \frac{1}{2}Y$

Se pide:

1. Generar muestras de X y de Y para aprender.
2. Generar muestras a clasificar uniformes sobre el cuadrado completo.
3. Construir una regla del histograma y clasificar las muestras de 2. según lo aprendido en 1. .
4. Graficar los resultados.

2.4 Clase 4

2.4.1 Ejercicio 9: KNN (SIMULACIÓN)

Se tienen dos normales bivariadas, ambas con matriz de covarianza $\Sigma = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, pero una centrada en $\mu_0 = (-1, 0)$, y la otra en $\mu_1 = (1, 0)$.

Se pide:

1. Generar muestras de N = 100, 1000, ... , otros valores. Estas dos serían las clases en las que clasificar los puntos.
2. Generar puntos con una distribución uniforme en $[-4, 4] \times [-4, 4]$.
3. Clasificar los puntos de la uniforme usando KNN, con $K = 1, 3$ y 13, utilizando las clases del primer item.
4. Graficar los resultados.

2.4.2 Ejercicio 9 bis

"Redactar una carilla que contenga la información relevante sobre la distribución normal multivariada. Llevar ejemplos, y mostrar como cambia el gráfico al variar la matriz de covarianza."

2.4.3 Ejercicio 10

Demostrar que $\forall a, b, c \in \mathbb{R}^+$:

$$(a+b+c)^2 \leq 3(a^2+b^2+c^2)$$

Sugerencia: usar Cauchy-Schwarz / Jensen.

2.5 Clase 5

2.5.1 Ejercicio 11

Dadas las desigualdades:

$$P(S_n - E\left[S_n\right] \geq \epsilon) \leq e^{-2\epsilon^2 / \sum_{i=1}^{n}(b_i - a_i)^2}$$

$$P(S_n - E\left[S_n\right] \leq -\epsilon) \leq e^{-2\epsilon^2 / \sum_{i=1}^{n}(b_i - a_i)^2}$$

Probar que se cumple la siguiente desigualdad:

$$P(\left|S_n - E\left[S_n\right]\right| \leq -\epsilon) \leq 2e^{-2\epsilon^2 / \sum_{i=1}^{n}(b_i - a_i)^2}$$

Sugerencia: Usar que $P(A \cup B) \leq P(A) + P(B)$

2.5.2 Ejercicio 12

En el ejercicio 11, y a partir de la demostración en clase de la primera de las hipotesis, realizar la demostración de la segunda hipotesis:

$$P(S_n - E\left[S_n\right] \leq -\epsilon) \leq e^{-2\epsilon^2 / \sum_{i=1}^{n}(b_i - a_i)^2}$$

2.5.3 Ejercicio 13

Dada la función:

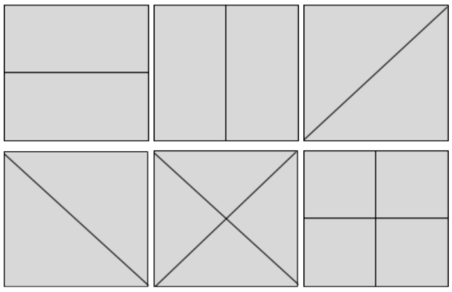
$$\Phi(u) = \ln pe^u + 1 - p - pu$$

Mostrar que:

- $\Phi(0) = 0$
- $\Phi'(0) = 0$

2.5.4 Ejercicio 14: PERDIDA (SIMULACIÓN)

Dados los siguientes clasificadores por particiones:



Elegir cuál es el mejor para las clases y los puntos generados en el ejercicio 8.

2.6 Clase 6

2.6.1 Ejercicio 15

Demostrar que la siguiente secuencia constituye una serie convergente:

$$a_n = 8(n+1)e^{-n\epsilon^2/32}$$

Y generalizar viendo que $\forall \lambda > 0$:

$$\sum_{n=0}^{\infty} ne^{-\lambda n} < \infty$$

(*Sugerencia: Usar que* $\int_0^{\infty} xe^{-\lambda x}dx = \frac{1}{\lambda^2}$)

2.6.2 Ejercicio 16

Sea la familia de intervalos:

$$A' = \{(a, b) : a < b\}$$

Luego de calcular $N_{A'}(z_1, z_2) = |\{\emptyset, \{z_1\}, \{z_2\}, \{z_1, z_2\}\}| = 4$ en clase, se pide calcular:

1. $N_{A'}(z_1, z_2, z_3)$
2. $N_{A'}(z_1, z_2, \cdots, z_n)$

3 Soluciones

3.1 Clase 1

3.1.1 Ejercicio 1

Solución:

f(x) es convexa en el intervalo $[x_1, x_2]$ si y solo si para cualquier $x_0 = p_1x_1 + p_2x_2$ se cumple la condición a comprobar.

Para comprobar utilizaremos la recta l(x) que contiene a $(x_1, f(x_1))$ y $(x_2, f(x_2) : \forall x_0 \in [x_1, x_2] : l(x_0) \geq f(x_0)$
Pero $l(x) = mx + b = \frac{f(x_2)-f(x_1)}{x_2-x_1}x + \left(f(x_1) - \frac{f(x_2)-f(x_1)}{x_2-x_1}x_1\right)$
Por simplicidad llamo $y_0 = f(x_1)$ y $y_2 = f(x_2)$
Evaluando:

$$l(x_0) = \frac{y_2 - y_1}{x_2 - x_1}x_0 + \left(y_1 - \frac{y_2 - y_1}{x_2 - x_1}x_1\right)$$

$$= \frac{y_2 - y_1}{x_2 - x_1}(p_1x_1 + p_2x_2) + \left(y_1 - \frac{y_2 - y_1}{x_2 - x_1}x_1\right)$$

Usando que $p_2 = 1 - p_1$

$$= \frac{y_2 - y_1}{x_2 - x_1}(p_1x_1 - p_1x_2) + \frac{y_2 - y_1}{x_2 - x_1}x_2 + \left(y_1 - \frac{y_2 - y_1}{x_2 - x_1}x_1\right)$$

$$= \frac{y_2 - y_1}{x_2 - x_1}(x_1 - x_2)(-p_1) + \frac{y_2 - y_1}{x_2 - x_1}(x_2 - x_1) + y_1$$

$$= p_1y_1 - p_1y_2 + y_2$$

$$= p_1y_1 + (1 - p_1)y_2$$

$$= p_1y_1 + p_2y_2$$

Finalmente $f(p_1x_1 + p_2x_2) = f(x_0) \leq l(x_0) = p_1f(x_1) + p_2f(x_2)$

3.1.2 Ejercicio 2: Esperanza condicional

Solución:

Sea $h(X) = \frac{f(X)}{\phi(X)}$, es claro que $h(X)$ puede adoptar cualquier valor de $\mathcal{H}$, por ejemplo, si busco adoptar $a(X)$ elijo $f(X) = a(X)\phi(X) \in \mathcal{H}$.

Usando la definición de $h(X)$ en la ecuación de hipotesis, se llega a que $\forall h(X) \in \mathcal{H}$:

$$E\left[m(X)\phi(X)h(X)\right] = E\left[Y\phi(X)h(X)\right]$$

Pero entonces como $m(X)\phi(X) \in \mathcal{H}$, tiene que ser la que minimice la distancia a $Y\phi(X)$, ya que:

$$E\left[m(X)\phi(X)h(X)\right] = E\left[Y\phi(X)h(X)\right]$$

sii

$$\langle m(X)\phi(X), h(X)\rangle = \langle Y\phi(X), h(X)\rangle$$

sii $\forall h(X) \in \mathcal{H}$:

$$\langle m(X)\phi(X) - Y\phi(X), h(X) \rangle = 0$$

sii

$$m(X)\phi(X) - Y\phi(X) \perp \mathcal{H}$$

.

Por lo tanto $m(X)\phi(X)$ es la esperanza condicional de $Y\phi(X)$ en $\mathcal{H}$
Finalmente por definición:

$$m(X)\phi(X) = E\left[Y\phi(X)|X\right]$$

y con $m(X) = E\left[Y|X\right]$ se obtiene la igualdad a demostrar:

$$E\left[Y|X\right]\phi(X) = E\left[Y\phi(X)|X\right]$$

3.1.3 Ejercicio 4 (no entregable)

3.2 Clase 2

3.2.1 Ejercicio 5

Solución:

$$\begin{aligned} P(X,Y) &= \int_0^{\frac{1}{2}} (1-x)dx + \int_{\frac{1}{2}}^{\frac{3}{4}} xdx \\ &= \int_0^{\frac{1}{2}} 1dx - \int_0^{\frac{1}{2}} xdx + \int_{\frac{1}{2}}^{\frac{3}{4}} xdx \\ &= \frac{1}{2} - \frac{1}{8} + (\frac{9}{32} - \frac{1}{8}) \\ &= 17/32 \end{aligned}$$

3.2.2 Ejercicio 6

Solución:

$$L^* = P(g*(T,F) \neq Y)$$

$$= P(T+B < 7-log2, T+B+E \geq 7) + P(T+B \geq 7-log2, T+B+E < 7)$$

$$= E(e^{-(7-T-B)}\mathbb{1}\{T+B < 7-log2\}) + P((1-e^{-(7-T-B)})\mathbb{1}\{7 > T+B \geq 7-log2\})$$

$$= \int_0^{7-log2} xe^{-x}e^{-(7-x)}dx + \int_{7-log2}^{7} xe^{-x}(1-e^{-(7-x)})dx$$

(porque la densidad de T + B es ue^u en $[0, \infty)$)

$$= e^{-7}\left(\frac{(7-log2)^2}{2} + 2(8-log2) - 8 - \frac{7^2}{2} + \frac{(7-log2)^2}{2}\right)$$

(ya que $\int_x^\infty ue^{-u}du = (1+x)e^{-x}$)

$$= 0.0199611...$$

3.3 Clase 3

3.3.1 Ejercicio 7

Solución:

Se pide mostrar que:

$$E[|\tilde{\eta}(x) - \tilde{\eta}_\epsilon(x)|] \leq E[|\eta_\epsilon(x) - \eta(x)|]$$

donde:

- $\tilde{f}(x) = E\left[f(x)|X \in A(x)\right]$ para toda funcion f.

Lema previo*(desigualdad de Jensen para Esperanzas condicionales)*:

Sean W, Z variables aleatorias y $f : \mathbb{R} \to \mathbb{R}$ una función convexa:
$f(E\left[W|Z\right]) \leq E\left[f(W)|Z\right]$.

En nuestro ejericio, sean $Z = X \in A(x)$, $W = \eta_\epsilon(x) - \eta(x)$ y
$f(x) = |x| := \begin{cases} x & \text{si } x \geq 0 \\ -x & \text{si } x < 0 \end{cases}$, claramente convexa en $\mathbb{R}$
Luego:

$$\tilde{\eta}(x) - \tilde{\eta}_\epsilon(x) = E\left[\eta(x)|X \in A(x)\right] - E\left[\eta_\epsilon(x)|X \in A(x)\right] = E\left[\eta(x) - \eta_\epsilon(x)|X \in A(x)\right]$$

Pero entonces usando la notacion de W y Z:
$|\tilde{\eta}(x) - \tilde{\eta}_\epsilon(x)| = f(E\left[W|Z\right])$

Y ademas como W y Z son independientes, tambien lo son $f(W)$ y Z, luego:

$E\left[f(W)|Z\right] = E\left[f(W)\right] = E[|\eta_\epsilon(x) - \eta(x)|]$ (2)
Y el ejercicio se reduce a la demostración del ***Lema previo***.

3.4 Clase 4

3.4.1 Ejercicio 9 bis

Normal Multivariada:

Caso general:

Un vector aleatorio $X = [X_1, \ldots, X_n]^T$ sigue una distribución normal multivariante si satisface las siguientes condiciones equivalentes:

- Toda combinación lineal $Y = a_1X_1 + \cdots + a_nX_n$ está normalmente distribuida.
- Hay un vector aleatorio $Z = [Z_1, \ldots, Z_m]^T$ cuyas componentes son variables aleatorias independientes distribuidas según la normal estándar, un vector $\mu = [\mu_1, \ldots, \mu_n]^T$ y una matriz $n \times m$ A tal que $X = AZ + \mu$

- Hay un vector μ y una matriz semidefinida positiva simétrica Σ tal que la función característica de X es:

$$\phi_X(u;\mu,\Sigma) = \exp(i\mu^T u - \frac{1}{2}u^T\Sigma u)$$

Si Σ es una matriz no singular, entonces la distribución puede describirse por la siguiente función de densidad:

$$f_X(x_1,\ldots,x_n) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^\top\Sigma^{-1}(\mathbf{x}-\mu)\right)$$

donde $|\Sigma|$ es el determinante de Σ. Nótese como la ecuación de arriba se reduce a la distribución normal si Σ es un escalar (es decir, una matriz 1x1).

El vector μ en estas circunstancias es la esperanza de X y la matriz $\Sigma = AA^T$ es la matriz de covarianza de las componentes X_i.

Caso bivariante:

En el caso particular de dos dimensiones, la función de densidad (con media $(0,0)$ es:

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} - \frac{2\rho xy}{(\sigma_x\sigma_y)}\right)\right)$$

donde ρ es el coeficiente de correlación entre X e Y. En este caso,

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$

.

Interpretación geométrica:

Las curvas de equidensidad de una distribución normal multivariante son elipsoides (es decir, transformaciones lineales de hiperesferas) centrados en la media. Las direcciones de los ejes principales de los elipsoides vienen dados por los autovectores de la matriz de covarianza Σ. Las longitudes relativas de los cuadrados de los ejes principales vienen dados por los correspondientes autovectores.

3.4.2 Ejercicio 10

Solución(CS):

Por CS se ve de forma directa que:

$$3(a^2+b^2+c^2) = (1^2+1^2+1^2)(a^2+b^2+c^2) \geq (1\cdot a + 1\cdot b + 1\cdot c)^2 = (a+b+c)^2$$

3.5 Clase 5

3.5.1 Ejercicio 11

Solución:

Por la definición de valor absoluto y la propiedad $P(A \cup B) \leq P(A) + P(B)$ se puede ver directamente el resultado pedido:

$$P(|S_n - E[S_n]| \leq -\epsilon) = P((S_n - E[S_n] \geq \epsilon) \cup (S_n - E[S_n] \leq -\epsilon))$$

$$\leq P(S_n - E[S_n] \geq \epsilon) + P(S_n - E[S_n] \leq -\epsilon)$$

$$= e^{-2\epsilon^2/\sum_{i=1}^{n}(b_i-a_i)^2} + e^{-2\epsilon^2/\sum_{i=1}^{n}(b_i-a_i)^2}$$

$$= 2e^{-2\epsilon^2/\sum_{i=1}^{n}(b_i-a_i)^2}$$

3.5.2 Ejercicio 12

Solución:

Sabemos que, si $X_1, X_2, \cdots, X_n$ son variables aleatorias acotadas e independientes tales que: $P(a_i \leq X_i \leq b_i) = 1$, con $i = 1, 2, \cdots n$, se cumple:

$$P(S_n - E[S_n] \geq \epsilon) \leq e^{-2\epsilon^2/\sum_{i=1}^{n}(b_i-a_i)^2} \quad (hip)$$

Sea $X_i^* = -X_i$ para todo $i = 1, \cdots, n$, luego $P(-b_i \leq X_i^* \leq -a_i) = 1$
Además defino $S_n^* = \sum_{i=1}^{n} X_i^*$
Como las variables $X_1^*, X_2^*, \cdots, X_n^*$ cumplen con las condiciones de la hipotesis:

$$P(S_n^* - E[S_n^*] \geq \epsilon) \overset{(hip)}{\leq} e^{-2\epsilon^2/\sum_{i=1}^{n}((-a_i)-(-b_i))^2}$$

Pero:

- $S_n^* = \sum_{i=1}^{n} X_i^* = \sum_{i=1}^{n}(-X_i) = -S_n$
- $E[S_n^*] = E[\sum_{i=1}^{n} X_i^*] = E[\sum_{i=1}^{n}(-X_i)] = -E[-S_n] = -E[S_n]$
- $((-a_i) - (-b_i))^2 = (b_i - a_i)^2$

Luego:

$$P(S_n - E[S_n] \leq -\epsilon) = P(S_n^* - E[S_n^*] \geq \epsilon) \leq e^{-2\epsilon^2/\sum_{i=1}^{n}((-a_i)-(-b_i))^2} = e^{-2\epsilon^2/\sum_{i=1}^{n}(b_i-a_i)^2}$$

$$\boxed{P(S_n - E[S_n] \leq -\epsilon) \leq e^{-2\epsilon^2/\sum_{i=1}^{n}(b_i-a_i)^2}}$$

3.5.3 Ejercicio 13

Solución:

$\Phi(0) = \ln(1) = 0$
$\Phi'(u) = \frac{(pe^u)}{(pe^u - p + 1)} - p$, $\Phi'(0) = \frac{(pe^0)}{(pe^0 - p + 1)} - p = p - p = 0$

3.6 Clase 6

3.6.1 Ejercicio 15

Solución:

Teorema: Si f es positiva, continua y decreciente para $x \geq 1$ y $a_n = f(n)$, entonces

$$\sum_{n=0}^{\infty} a_n \quad \text{y} \quad \int_0^{\infty} f(x)\,\mathrm{d}x$$

convergen o divergen ambas simultáneamente.
Si $T \sim \exp(\lambda)$ y $a_n = ne^{-\lambda n}$,
es inmediato que como $\int_0^{\infty} xe^{-\lambda x}dx = \frac{1}{\lambda}E\left[T\right] = \frac{1}{\lambda^2} < \infty$
entonces la serie asociada es convergente.

3.6.2 Ejercicio 16

Solución:

El problema equivale a contar la cantidad de subconjuntos con elementos consecutivos hay en un conjunto $\{z_1, z_2, ..., z_n\}$ con $z_1 < z_2 < ... < z_n$.

Los subconjuntos con mas de un elemento quedan completamente definidos simplemente eligiendo sus elementos extremos,

Luego hay $\binom{n}{2} = \frac{n(n-1)}{2}$ de estos, sumados a los n conjuntos de un elemento y al conjunto vacío, se tiene que:

$$N_{A'}(z_1, z_2, \cdots, z_n) = \frac{n(n-1)}{2} + n + 1$$

En caso $n = 3$ es inmediato habiendo deducido este, por lo tanto.

$$N_{A'}(z_1, z_2, z_3) = \frac{3(3-1)}{2} + 3 + 1 = 7.$$

Que coincide con el cardinal del conjunto (hallado de forma exhaustiva):

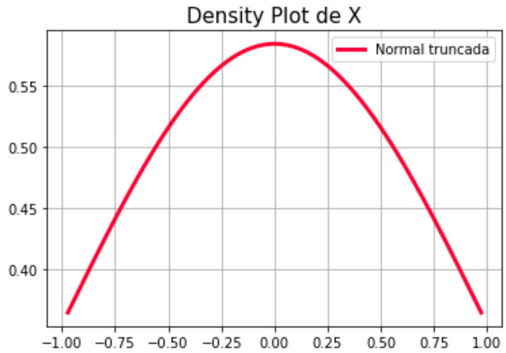
$$\{\emptyset, \{z_1\}, \{z_2\}, \{z_3\}, \{z_1, z_2\}, \{z_2, z_3\}, \{z_1, z_2, z_3\}\}$$

4 Soluciones de problemas con Simulación

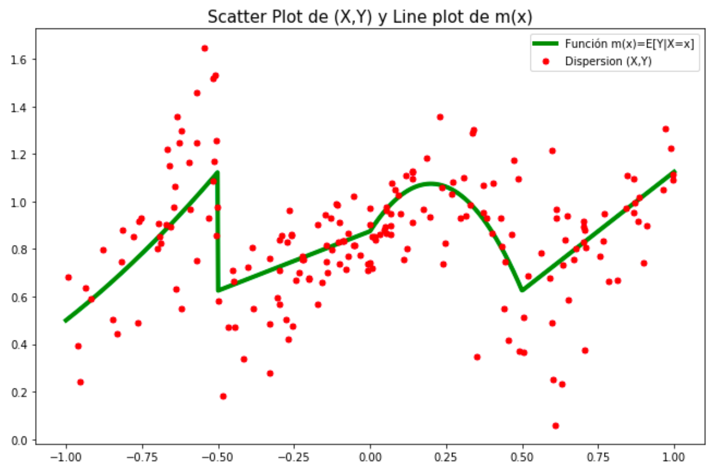
Para todos los ejercicios, se utilizaron librerias de Python (numpy, pandas, matplotlib, seaborn, sklearn) en el entorno de Jupyter Lab, cabe aclarar que el código que permitió obtener las ilustraciones y gráficos se encuentra en el repositorio antes mencionado, y habiendo en el mismo un análisis más extenso que el presentado en este informe.

4.1 Ejercicio 3 (SIMULACIÓN)

Se simuló la normal truncada al intervalo $[-1, 1]$



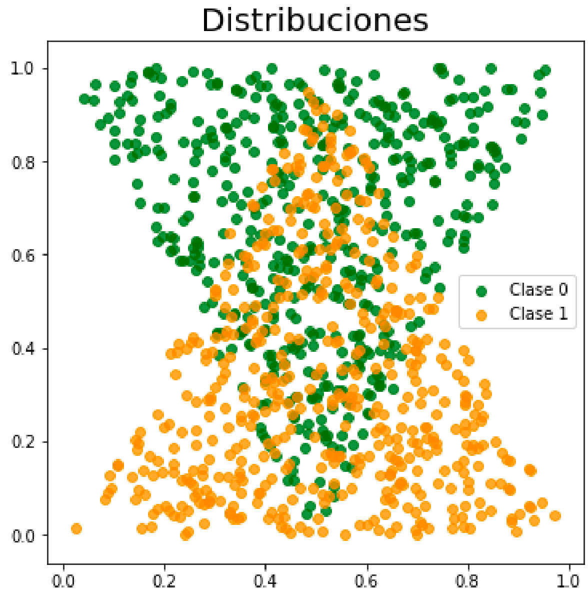
Con la función $m(x)$ y la distribución de $Y - m(x) \sim \mathcal{N}(0, \sigma^2(x))$
Con $\sigma(x) = 0.2 - 0.1\cos(2x)$



Los 200 puntos se encuentran en un archivo con la extension .csv (”Comma Separated Values”) en el repositorio de la materia.

4.2 Ejercicio 8: PARTICIONES (SIMULACIÓN)

Para comenzar, se simularon las distribuciones uniformes sobre las Regiones descriptas en el enunciado, se muestra un ejemplo con 1000 puntos de muestra para entrenar las particiones.



Simulación de las dos variables aleatrias, con $n = 1000$

Para generar la particion tengo que saber la longitud del lado de las cajas. Segun lo visto en clase, si h_n es la longitud del lado de las cajas, entonces:

$$h_n = \frac{1}{\sqrt[2d]{n}} = n^{-\frac{1}{2d}}$$

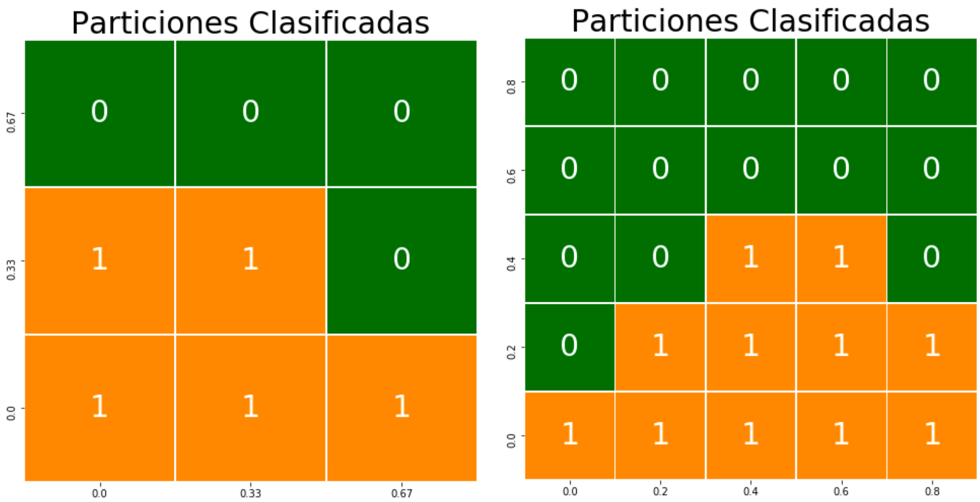
cumple las condiciones para que la regla del histograma sea universalmente consistente, estas son:

- $\lim_{n \to \infty} h_n = 0$
- $\lim_{n \to \infty} n h_n^d = \infty$

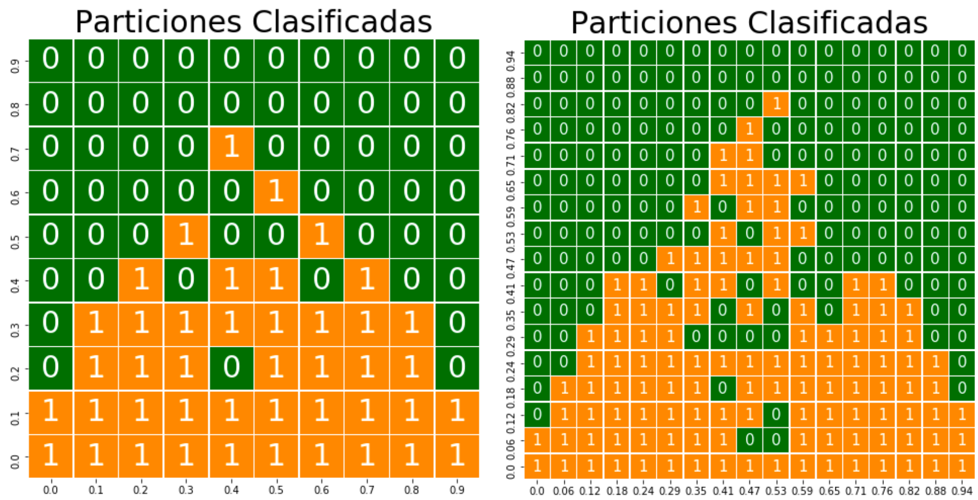
En este caso, con dos dimensiones, $d = 2$:

$$h_n = \frac{1}{\sqrt[4]{n}}$$

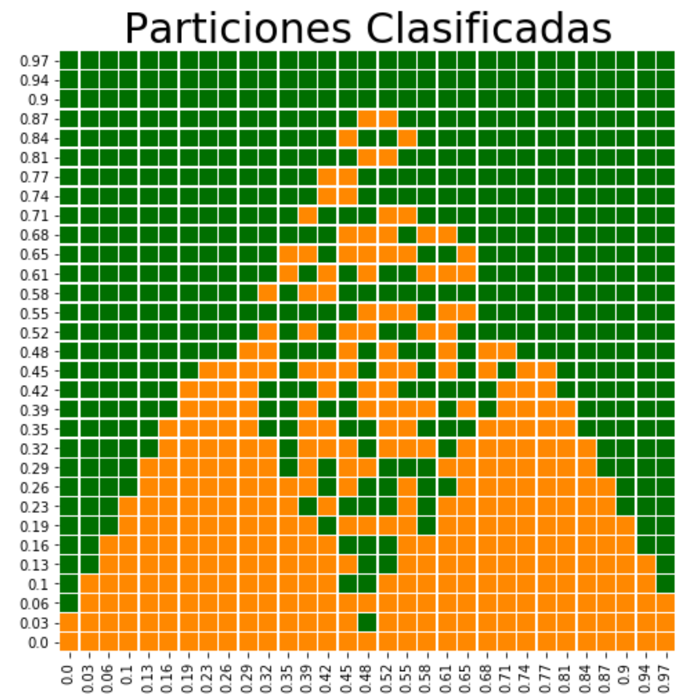
Se exciben los diferentes clasifcadores del intervalo $[0, 1] \times [0, 1]$, asignandole el empate a la clase 0.



Particiones con n=100 y n=1000 respectivamente



Particiones con n=10000 y n=100000 respectivamente



Particiones con un millón de datos de train.

4.3 Ejercicio 9: KNN (SIMULACIÓN)

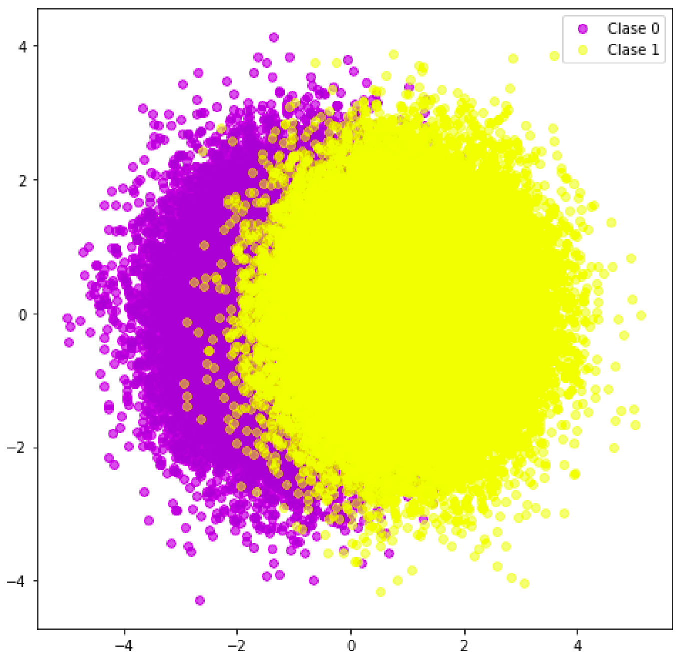
Se procedió a simular las dos normales requeridas, catalogando como de *Clase* 0 a los puntos generados con distribución Normal $\mathcal{N}(\mu_0, \Sigma_0)$ con:

- $\mu_0 = (0, -1)$
- $\Sigma_0 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

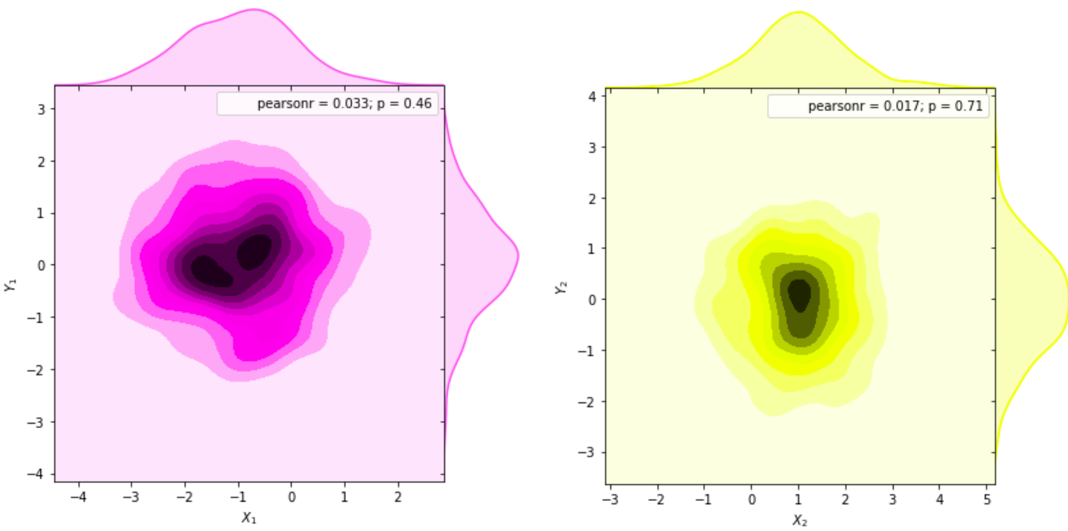
y los puntos de *Clase* 1 a los puntos generados con distribución Normal $\mathcal{N}(\mu_1, \Sigma_1)$ con:

- $\mu_1 = (0, 1)$
- $\Sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

Se muestra una dispersión de los mismos, habiendo generado 100000 puntos en total (50000 de cada clase):



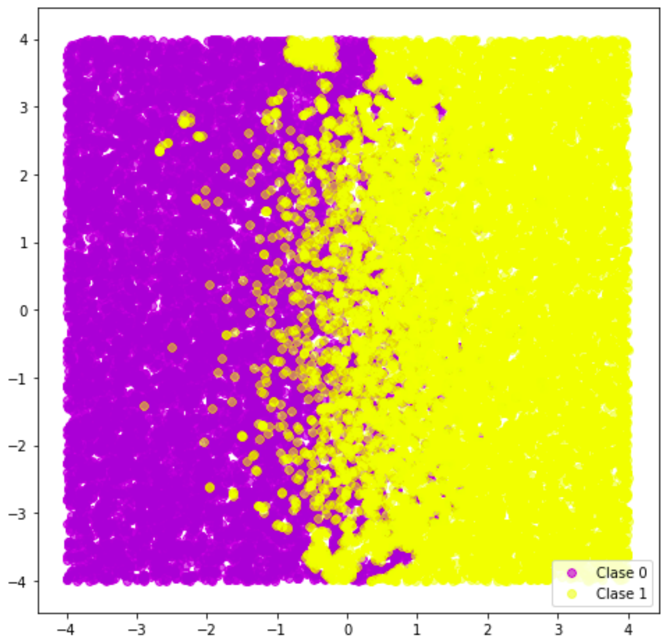
Dispersión de puntos generados



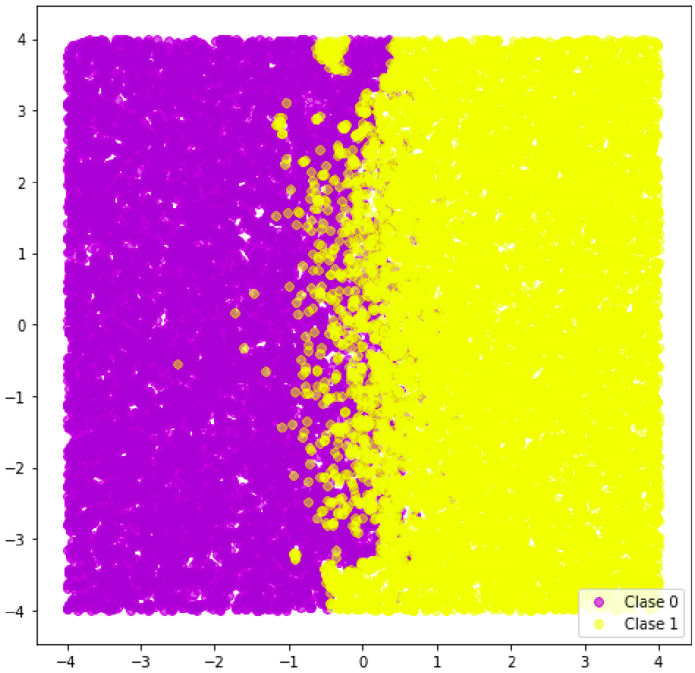
Distribuciónes de ambas Normales con sus marginales

Luego se paso a entrenar (con este set de 100000 puntos de train) al algoritmo de KNN, variando el hiperparametro K, y utilizando siempre la distancia euclideana.

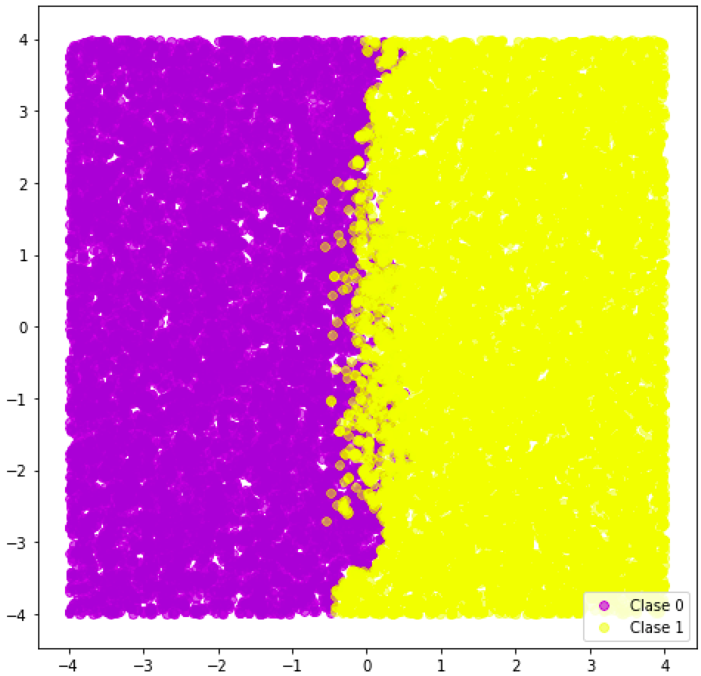
Se procedió a clasificar el set de 20000 puntos distribuidos de forma uniforme en el intervalo $[-4, 4] \times [-4, 4]$, obteniendo los siguientes resultados sobre los puntos de Testing(uniformes):



KNN Tomando el vecino más cercano (K=1)



KNN (K=3)



KNN (K=13)

4.3.1 Costos de KNN

Veremos como se comporta KNN prediciendo puntos de Test.

Para eso generamos nuevos puntos (en dos partes), identicamente distribuidos a la normal 0 en caso de error_n0 y a la normal 1 en caso de error_n1.

Se toma como estimador de la función perdida del clasificador $\Phi \in \mathcal{C}$ a:

$$\widehat{L}_n(\Phi) = \frac{1}{n} \sum_{i=1}^{n-test} 1\{\Phi(X_i) \neq Y_i\}$$

Con X_i el vector aleatorio de features, con Y_i su label correspondiente.

	$\widehat{L_n}$	
Clasificador	n0_test	n1_test
K=1	0.220	0.220
K=3	0.195	0.186
K=13	0.162	0.165

Se puede concluir que si el set de Test es i.i.d al de train, producirá menor error al aumentar K.

4.4 Ejercicio 14: PERDIDA (SIMULACIÓN)

De forma analoga al ejercicio de particiones, se construyeron los seis clasificadores propuestos, se exciben los clasificadores $1, 2$ y 6 (siguiendo el esquema del enunciado):

Clasificadores con n=10000.

Utilizando como estimador de la función perdida para el clasificador $\Phi \in \mathcal{C}$:

$$\widehat{L}_n(\Phi) = \frac{1}{n} \sum_{i=1}^{n-test} 1\{\Phi(X_i) \neq Y_i\}$$

Con X_i el vector aleatorio de features, con Y_i su label correspondiente.

Se estimaron los errores tanto del set de Training como el de Testing (de $\lfloor \frac{n}{6} \rfloor = 1666$ puntos generados con la misma distribución que los de Train)

Obteniendo los siguientes resultados:

$\widehat{L_n}$ - Train

Clasificador	error
1	0.7478
2	0.5028
3	0.6716
4	0.6573
5	0.6716
6	0.7478

Se podría sospechar que el Clasicador 3 y 5 han clasificado las regiones de la misma manera, se comprueba por simple verificación en el clasificador.

$\widehat{L_n}$ - Test

Clasificador	error
1	0.756
2	0.501
3	0.476
4	0.496
5	0.476
6	0.491

Con esto se podría decir que frente a los datos de Training, el que produce menor error estimado es el Clasificador 2 (Quizas porque se puede ajustar mejor por su geometría a las regiones)

No obstante, con los datos de Testing, los clasificadores 3 y 5 son los que minimizan el error estimado.