

# Reconocimiento de Patrones

Departamento de Computación  
Facultad de Ciencias Exactas y Naturales  
Universidad de Buenos Aires

... de ... de 2016

## Trabajos Prácticos

Integrante	LU	Correo electrónico
Arrondo, Brian Ariel		
Benzo, Mariano		
Maddonni, Axel Ezequiel	200/14	axel.maddonni@gmail.com

# Índice

<b>1. TP1</b>	<b>3</b>
<b>2. TP2</b>	<b>4</b>
<b>3. TP3: Métodos no-paramétricos</b>	<b>5</b>
3.1. Esquema general . . . . .	5
3.2. Ventanas de Parzen . . . . .	5
3.2.1. Pseudocódigo: Ventanas de Parzen . . . . .	6
3.3. $k_n$ vecinos más cercanos . . . . .	6
3.3.1. Pseudocódigo: $k_N$ vecinos más cercanos . . . . .	7
3.4. Ejercicio 1 . . . . .	7
3.4.1. Resultados e Imágenes . . . . .	7
3.5. Ejercicio 2 . . . . .	8
3.5.1. Resultados e Imágenes . . . . .	9
<b>4. TP4: Discriminante de Fisher</b>	<b>10</b>
<b>5. TP5: Expectation Maximization</b>	<b>11</b>

## 1. TP1

## 2. TP2

### 3. TP3: Métodos no-paramétricos

Los métodos no paramétricos se utilizan para estimar las funciones de densidad de probabilidad (f.d.p.) cuando no se conoce la expresión de la misma ni los parámetros.

#### 3.1. Esquema general

Sea una cierta región  $R$  del espacio de características. La probabilidad  $P_R$  de que un cierto  $x$  pertenezca a  $R$  viene dada por:

$$P_R = p(x \in R) = \int_R p(x) dx \approx p(x) \int_R dx = p(x) * V$$

donde  $V$  es el volumen ocupado por la región  $R$ .

Supongamos que disponemos de  $N$  muestras independientes  $x_1, x_2, \dots, x_N$  correspondientes a la f.d.p. que queremos estimar, y dados:

- $k_N$  = cuántas de las  $N$  muestras pertenecen a la región  $R$
- $\frac{k_N}{N}$  es una estimación de  $P_R$  para cada  $x$ , lo cual es a su vez una estimación de  $p(x)$ .

Usando:

$$p(x) * V \approx \frac{k_N}{N}$$

Podemos estimar  $p(x)$  como :

$$p(x) = \frac{\frac{k_N}{N}}{V}$$

Es decir, se aproxima  $p(x)$  definiendo una región  $R$  pequeña alrededor de  $x$  y contando cuántos de los  $x_i$  caen en  $R$ . En los ejercicios se aplicaron dos métodos no paramétricos:

- Ventanas de Parzen
- $k_n$  vecinos más cercanos

#### 3.2. Ventanas de Parzen

Tomando una región  $R$  con forma de hipercubo  $d$ -dimensional de lado  $h_N$ , entonces:

$$V_N = (h_N)^d$$

Y sea la siguiente **función de ventana**:

$$\varphi(x) = \begin{cases} 1 & |u_j| \leq \frac{1}{2}, j = 1..d \\ 0 & \text{caso contrario} \end{cases}$$

Entonces, para un cierto  $x$ :

$$\varphi\left(\frac{x - x_i}{h_N}\right) = 1 \text{ si } x_i \in \text{hipercubo de volumen } V_N \text{ centrado en } x$$

El número de muestras, entonces, se calcula:

$$k_N = \sum_{i=1}^N \varphi\left(\frac{x - x_i}{h_N}\right)$$

y finalmente podemos estimar la f.d.p. resolviendo:

$$p_N(x) = \frac{k_N}{N} = \frac{1}{N} \sum_{i=1}^N \frac{1}{h_N^d} \varphi\left(\frac{x - x_i}{h_N}\right) \quad (1)$$

### 3.2.1. Pseudocódigo: Ventanas de Parzen

---

**Algorithm 1** Pseudocódigo del método de ventanas de parzen

---

```

1: procedure PARZEN_HIPERCUBO( $x$  (punto a evaluar),  $X$  (muestra),  $h$  (tamaño de la ven-
   tana),  $d$ (dimensión))
2:    $L \leftarrow |X|$ 
3:    $k_n \leftarrow 0$ 
4:   for  $i \in (1..L)$  do
5:      $in\_inside \leftarrow 0$ 
6:     for  $j \in (1..d)$  do
7:       if  $|x(j) - X_i(j)|/h \leq 0,5$  then
8:          $is\_inside ++$ 
9:       if  $is\_inside == d$  then
10:         $k_n ++$ 
11:    $p(x) \leftarrow \frac{k_n}{L} / h^d$ 

```

---

### 3.3. $k_n$ vecinos más cercanos

A diferencia del método de ventanas de parzen, en este método no se fija el  $V_N$  sino que se buscan las  $k_N$  muestras más próximas a  $x$  y se determina el volumen que las contiene. Es decir,  $V_N$  se calcula en función de la muestra. Se estima  $p_N$  calculando:

$$p_N(x) = \frac{\frac{k_N}{N}}{V_{k,N}} \quad (2)$$

### 3.3.1. Pseudocódigo: $k_N$ vecinos más cercanos

---

**Algorithm 2** Pseudocódigo del método de  $k_N$  vecinos con dimensión 2
 

---

```

1: procedure KNN_2D( $x$  (punto a evaluar),  $X$  (muestra))
2:    $L \leftarrow |X|$ 
3:    $k \leftarrow \sqrt{L}$ 
4:   for  $i \in (1..L)$  do
5:      $dist(i) \leftarrow ||x - X_i||$ 
6:    $sort(dist)$ 
7:    $V \leftarrow \pi * dist(min(L, k))^2$ 
8:    $p(x) \leftarrow \frac{k}{L}/V$ 

```

---

## 3.4. Ejercicio 1

Se realizó un estudio de la convergencia de la f.d.p. de una distribución Gaussiana univariada utilizando el método de Ventanas de Parzen y  $k_n$  vecinos más cercanos. Para dicho estudio se implementaron funciones para calcular  $N$  iteraciones de la sucesión  $p(x)$  evaluada en  $x = 0, 1$  y  $2$  usando ambos métodos, con  $N = 10, 100, 1000$  y  $10000$ . Para cada valor se usaron muestras de distintos tamaños. (ver código de funciones `knn_1D` y `parzen_cubo_1D`).

### 3.4.1. Resultados e Imágenes

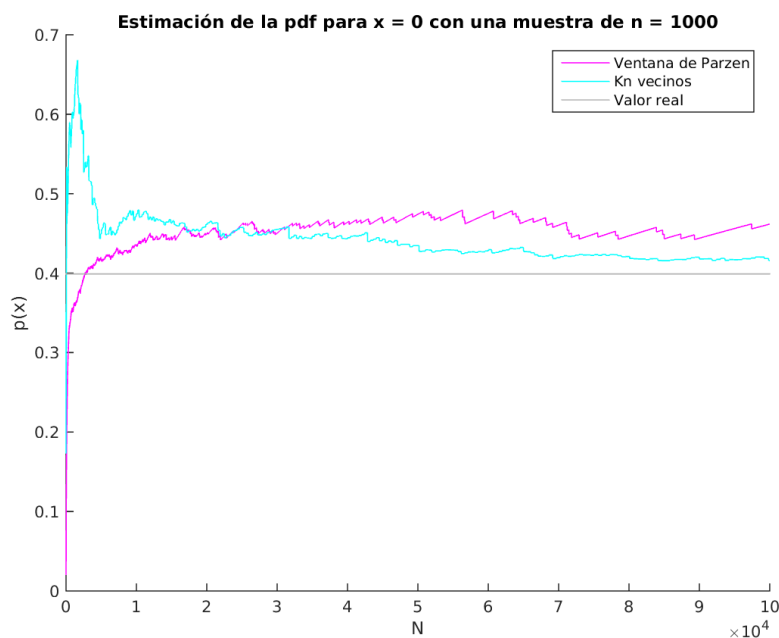


Figura 1: Convergencia de la f.d.p. de una dist. Normal (1,0).

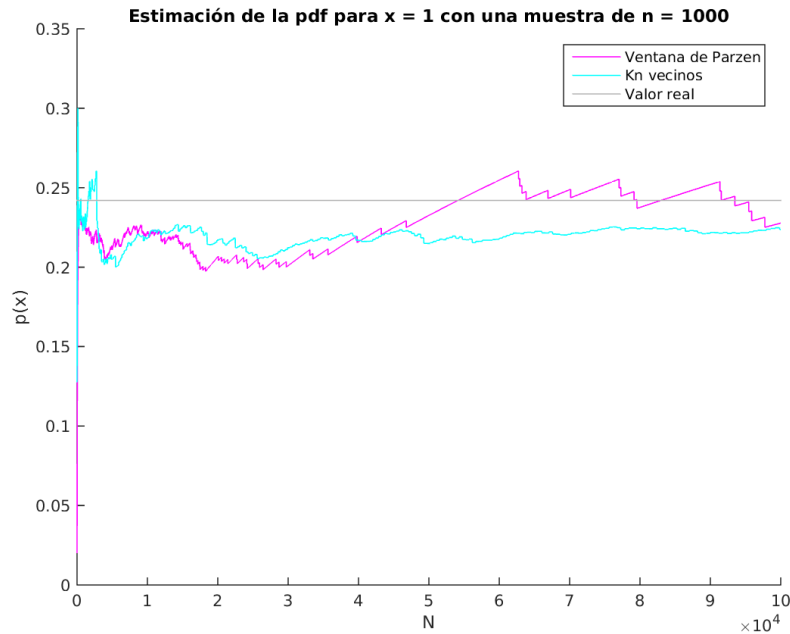


Figura 2: Convergencia de la fdp de una dist. Normal (1,0).

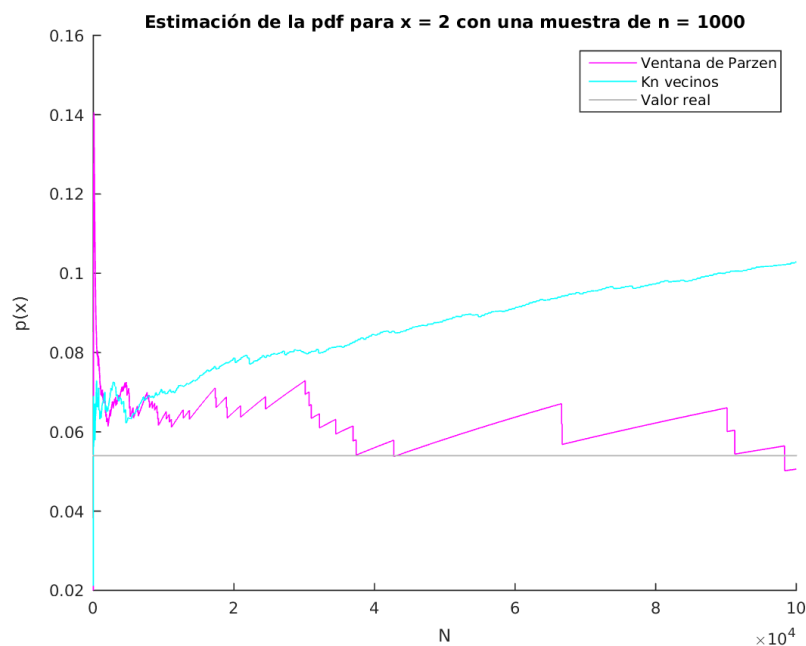


Figura 3: Convergencia de la fdp de una dist. Normal (1,0).

### 3.5. Ejercicio 2

En este ejercicio se generaron datos pertenecientes a dos clases  $w_1$  y  $w_2$ , generados a partir de dos Gaussinas bivariadas. Utilizando los métodos mencionados en el ejercicio anterior, adaptados a dos variables, se realizó la partición del espacio de características resultante (dado por las muestras generadas). Para graficar las particiones, se determinaron intervalos bidimensionales rectangulares de tamaño  $1 \times 1$  y se evaluaron los centros de cada intervalo para



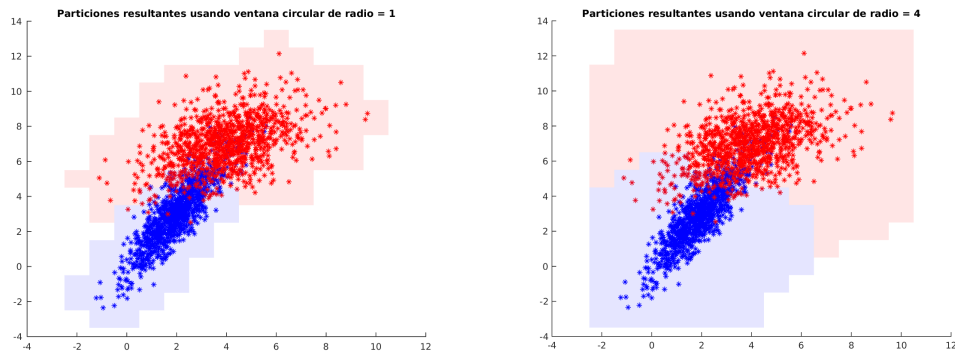


Figura 4: División del espacio de características resultante usando ventanas circulares.

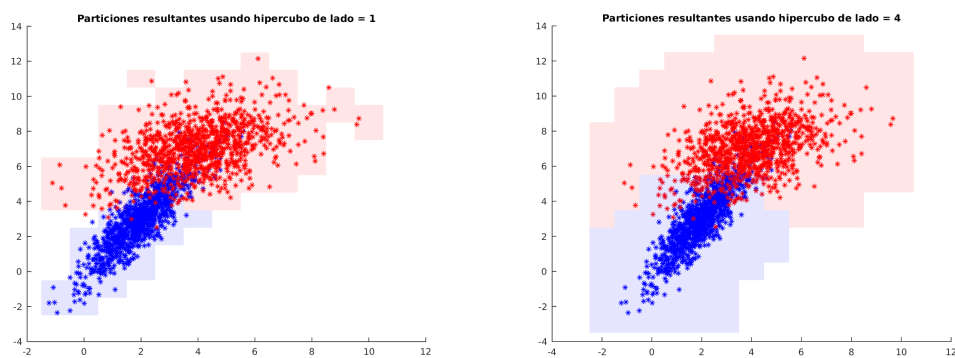


Figura 5: División del espacio de características resultante usando ventanas hipercúbicas.

aproximar la  $p(x)$  correspondiente a cada muestra. A continuación, una comparación de los resultados obtenidos entre los métodos de ventanas de parzen (con un hipercubo de dimensión 2) y  $k_n n$  vecinos más cercanos. Además incluimos una comparación usando una ventana circular centrada en el punto a evaluar. Para este último, se implementó un algoritmo de ventanas de Parzen donde el  $V$  utilizado es el área del círculo resultante centrado en el  $x$  correspondiente, con el radio del círculo ingresado por parámetro. (ver código de función `parzenr2_circulo`).

### 3.5.1. Resultados e Imágenes

Uno de los problemas característicos de este método consiste en encontrar la secuencia de volúmenes  $V_i$  o el tamaño de ventana óptimo. Por ejemplo, tomando  $V_N = V_1/N$ , los resultados son sensibles a la elección del volumen inicial  $V_1$ . Si el volumen inicial es muy pequeño, la mayoría de los volúmenes estarán vacíos, y la estimación de  $p_n$  se vuelve muy sensible a errores (ver figura ...). Por otro lado, tomando  $V_1$  grande, las variaciones en  $p(x)$  que caen dentro de la misma ventana se perderán (ver figura ...). Para solucionar este problema, se presenta el siguiente método.

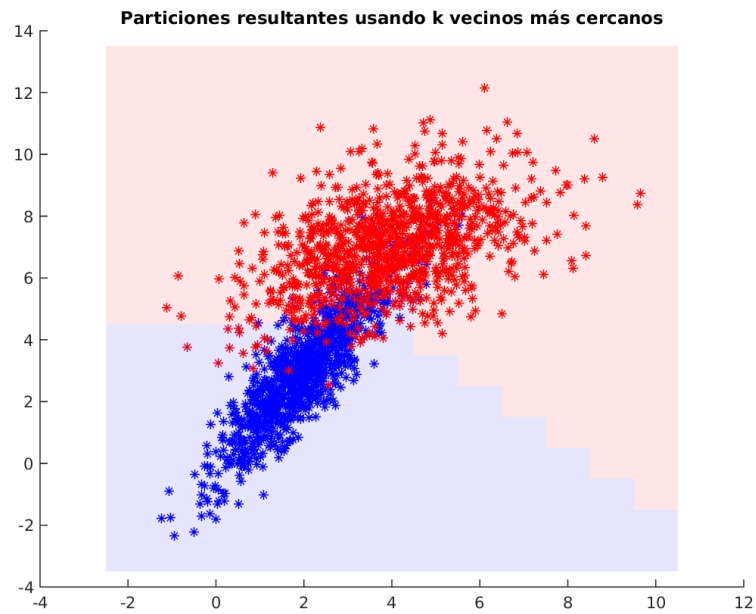


Figura 6: División del espacio de características resultante usando kvecinos más cercanos.

#### 4. TP4: Discriminante de Fisher

## 5. TP5: Expectation Maximization