Examen de Teoría de Percepción - Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2015

Apellidos: ☐ **Nombre:** ☐

Profesor: ☐ **Jorge Civera** ☐ **Roberto Paredes**

Cuestiones (3 puntos, 30 minutos, sin apuntes)

$\boxed{\text{C}}$ El suavizado de una distribución de Bernoulli, ¿qué valores de los parámetros suele modificar?

A) Sólo valores cercanos a cero, pues el logaritmo de esos valores tiende a menos infinito.
B) Todos los valores en el rango de cero a uno.
C) Tanto valores cercanos a cero, como valores cercanos a uno.
D) No modifica ningún valor.

$\boxed{\text{B}}$ Dado el siguiente conjunto de vectores binarios bidimensionales:

n	1	2	3	4	5	6	7	8	9	10	11	12
x_{n1}	1	1	0	1	1	1	0	0	1	0	1	0
x_{n2}	1	0	0	0	1	0	0	1	0	1	0	1
c_n	1	1	1	1	2	2	2	2	2	2	2	2

¿Cuál es la estimación de los parámetros del clasificador Bernoulli más probable?

A) $\hat{p}(1) = \frac{1}{2}$, $\hat{p}(2) = \frac{1}{2}$, $\hat{\mathbf{p}}_1 = \left(\frac{3}{4}, \frac{1}{4}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{2}, \frac{1}{2}\right)^t$
B) $\hat{p}(1) = \frac{1}{3}$, $\hat{p}(2) = \frac{2}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{3}{4}, \frac{1}{4}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{2}, \frac{1}{2}\right)^t$
C) $\hat{p}(1) = \frac{1}{2}$, $\hat{p}(2) = \frac{1}{2}$, $\hat{\mathbf{p}}_1 = \left(\frac{5}{6}, \frac{2}{6}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{3}, \frac{1}{2}\right)^t$
D) $\hat{p}(1) = \frac{1}{3}$, $\hat{p}(2) = \frac{2}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{5}{6}, \frac{2}{6}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{3}, \frac{1}{2}\right)^t$

$\boxed{\text{B}}$ ¿Qué tipo de fronteras de decisión define un clasificador basado en distribuciones multinomiales?

A) Lineal definida a trozos.
B) Lineal.
C) Cuadrática.
D) Ninguna de las anteriores.

$\boxed{\text{A}}$ Sean A y B dos clases con igual prior y f.d.p. condicionales de clase gaussianas con los siguientes parámetros $\mu_A = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\Sigma_A = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$, $\mu_B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ y $\Sigma_B = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$, ¿qué tipo de frontera de decisión definen?

A) Lineal
B) Lineal definida a trozos
C) Cuadrática
D) Ninguna de las anteriores

$\boxed{\text{A}}$ Dada la muestra $\mathbf{x} = \begin{pmatrix} 0 \\ 0.25 \end{pmatrix}$, ¿a qué clase sería asignada por un clasificador gaussiano basado en los parámetros de la cuestión anterior?

A) A la clase A
B) A cualquiera de las dos clases
C) A la clase B
D) A ninguna de las dos clases

$\boxed{\text{C}}$ El algoritmo de condensado de Hart acaba cuando:

A) ninguna muestra se clasifica correctamente
B) cuando se vacía el conjunto GARBAGE
C) ninguna muestra se clasifica incorrectamente o se vacía el conjunto GARBAGE
D) ninguna muestra se clasifica incorrectamente

$\boxed{\text{A}}$ El algoritmo de edición de prototipos de Wilson:

A) el resultado puede depender del parámetro k con el que se realiza la clasificación
B) devuelve un conjunto con menos prototipos que el de entrada
C) las muestras ruidosas permanecen en el conjunto GARBAGE
D) se eliminan los prototipos cuya clasificación es correcta

$\boxed{\text{C}}$ Cuál de las siguientes afirmaciones es correcta:

A) En general el Bias aumenta al escoger clasificadores más fuertes
B) En general el Variance aumenta al aumentar el conjunto de aprendizaje
C) En general el Bias aumenta al escoger clasificadores más débiles
D) En general el Variance se reduce empleando boosting

$\boxed{\text{D}}$ En general cuál de las siguientes combinaciones de clasificadores funcionar mejor:

A) Combinar múltiples clasificadores k-nn con diferentes valores de k
B) Combinar múltiples clasificadores lineales con diferentes valores de $\mathbf{w}$
C) Combinar múltiples clasificadores k-nn con diferentes distancias
D) Combinar multiples clasificadores, lineales, k-nn, árboles de decisión etc.

$\boxed{\text{D}}$ En reinforcement learning:

A) Se trata de escoger $T' \subset T$ del menor tamaño posible para supervisar/etiquetar
B) En cada interacción se genera una muestra más que añadir al conjunto de aprendizaje
C) Se emplean los datos corregidos por el operador humano como nuevos datos para la adaptación de los modelos
D) Se calcula en término de operaciones de exploración y explotación

Examen de Teoría de Percepción - Recuperación 1$^{\text{er}}$ parcial

ETSINF, Universitat Politècnica de València, Junio de 2015

Apellidos: ________ **Nombre:** ________

Profesor: ☐ **Carlos Martínez** ☐ **Roberto Paredes**

Cuestiones (3 puntos, 30 minutos, sin apuntes)

$\boxed{\text{D}}$ Supongamos un problema de clasificación en el que se detecta el género de una película. ¿Cómo se expresaría en términos de la clasificación estadística $c(x) = \arg\max_c P(c|x)$?

A) x sería el género y c sería la película
B) x sería la película y c sería el género
C) Tanto x como c representarían la película
D) x sería una *representación* de la película y c una *etiqueta* asociada al género

$\boxed{\text{B}}$ Dado un problema de clasificación entre dos clases A y B, con funciones discriminantes asociadas g_A y g_B, la frontera de decisión entre A y B viene dada por las representaciones x que cumplen:

A) $\text{máx}_x\, g_A(x) = \text{máx}_x\, g_B(x)$
B) $g_A(x) = g_B(x)$
C) $\text{mín}_x\, g_A(x) = \text{máx}_x\, g_B(x)$
D) $g_A(x) > g_B(x)$

$\boxed{\text{C}}$ En un problema de reconocimiento de imágenes donde el detalle discriminativo mínimo es de 2 milímetros, ¿cual es la frecuencia de muestreo mínima que se debe aplicar (entre las enumeradas) para mantener ese nivel de detalle?

A) 512 muestras por metro
B) 768 muestras por metro
C) 1024 muestras por metro
D) 2048 muestras por metro

$\boxed{\text{A}}$ Dados los *codewords* { (a,(1,1)), (m,(3,-1)), (l,(-1,2)), (o,(3,3)) }, indicar la codificación por ese *codebook* de la secuencia (2,-1), (3,0), (2,1), (1,2), (0,2), (-1,1), (0,3), (2,3), (3,2)

A) mmaallloo
B) maaallllo
C) malo
D) mmaaoollo

$\boxed{\text{B}}$ La función *ldf* empleada en la clasificación de documentos se caracteriza por:

A) Disminuir la complejidad espacial de la representación *bag-of-words*
B) Atenuar los *tokens* con presencia en muchos documentos
C) Incluir contexto en la representación del documento
D) Evitar el proceso de *stemming*

X PCA se resuelve minimizando el error de reconstrucción. Al final de se llega a un un problema de minimización con restricciones que se resuelve mediante multiplicadores de Lagrange. El problema equivalente sería este:

A) $\mathbf{w}^* = \arg\min_{\mathbf{w}\in R^d} (\mathbf{w}\Sigma_\mathbf{x}\mathbf{w}^t + (1 - \mathbf{w}\mathbf{w}^t))$
B) $\mathbf{w}^* = \arg\min_{\mathbf{w}\in R^d} (\mathbf{w}\Sigma_\mathbf{x}\mathbf{w}^t + \lambda(1 - \mathbf{w}\mathbf{w}^t))$
C) $\mathbf{w}^* = \arg\min_{\mathbf{w}\in R^d} (\mathbf{w}\Sigma_\mathbf{x}\mathbf{w}^t + (1 - \mathbf{w}^t\mathbf{w}))$
D) Ninguno de los anteriores

Esta pregunta supone que los vectores son fila (no columna), no plantea la optimización de λ y cambia maximización por minimización. Por tanto, la opción dada inicialmente por correcta (C) no sería correcta.

B Dada la diagonalización de la matriz de covarianzas $\Sigma_{3\times3}$ en valores y vectores propios $\lambda_1 = 0.7$ con $\mathbf{w}_1 = (1\ 0\ 0)$, $\lambda_2 = 5.2$ con $\mathbf{w}_2 = (0\ 1\ 0)$, y $\lambda_3 = 2.7$ con $\mathbf{w}_3 = (0\ 0\ 1)$:

A) La proyección PCA de $\mathbb{R}^3$ a $\mathbb{R}^2$ se llevará a cabo con los vectores propios $\mathbf{w}_1$ y $\mathbf{w}_2$
B) La proyección PCA de $\mathbb{R}^3$ a $\mathbb{R}^2$ se llevará a cabo con los vectores propios $\mathbf{w}_2$ y $\mathbf{w}_3$
C) La proyección PCA de $\mathbb{R}^3$ a $\mathbb{R}^1$ se llevará a cabo con el vector propio $\mathbf{w}_1$
D) Ninguna de las anteriores dado que los eigenvectores no son ortonormales

D ¿Cuál de estas afirmaciones sobre LDA **NO** es correcta?

A) LDA obtiene una matriz de proyección lineal optimizando una función objetivo que persigue maximizar las distancias interclase mientras se minimizan las intraclase
B) Es una proyección lineal donde no tiene sentido escoger más de $C - 1$ eigenvectores siendo C el número de clases
C) Es una proyección lineal que resulta del análisis de eigenvectores generalizados de dos matrices comúnmente expresadas como S_w y S_b
D) LDA obtiene una matriz de proyección lineal optimizando una función objetivo que persigue maximizar las distancias intraclase mientras se minimizan las interclase

A Se recomienda emplear funciones kernel cuando:

A) El espacio de representación original no es linealmente separable
B) El kernel escogido modela el producto escalar en un nuevo espacio de representación
C) El nuevo espacio de representación cabe en memoria
D) El espacio de representación original es linealmente separable

C Esencialmente, el algoritmo Kernel Perceptron lo que hace es:

A) Incrementar la importancia (peso) de las muestras correctamente clasificadas
B) Decrementar la importancia (peso) de las muestras correctamente clasificadas
C) Incrementar la importancia (peso) de las muestras incorrectamente clasificadas
D) Decrementar la importancia (peso) de las muestras incorrectamente clasificadas

Examen de Teoría de Percepción - Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2016

Apellidos: ☐ **Nombre:** ☐

Profesor: ☐ **Jorge Civera** ☐ **Roberto Paredes**

Cuestiones (3 puntos, 30 minutos, sin apuntes)

D ¿Cuál de los siguientes valores del parámetro $\mathbf{p}$ no define una distribución Bernoulli?

A) $\mathbf{p} = \left(\frac{3}{4}, \frac{1}{4}\right)^t$
B) $\mathbf{p} = \left(\frac{3}{4}, \frac{3}{4}\right)^t$
C) $\mathbf{p} = \left(\frac{1}{2}, \frac{1}{2}\right)^t$
D) Todos los valores anteriores del parámetro $\mathbf{p}$ definen una distribución Bernoulli.

B ¿Qué tipo de fronteras de decisión define un clasificador basado en la distribución Bernoulli?

A) Lineal definida a trozos.
B) Lineal.
C) Cuadrática.
D) Ninguna de las anteriores.

D Dado el siguiente conjunto de vectores de contadores bidimensionales:

n	1	2	3	4	5	6	7	8	9	10	11	12
x_{n1}	2	1	0	2	1	2	0	2	3	0	1	1
x_{n2}	3	0	4	0	3	0	3	2	2	3	1	4
c_n	1	1	1	1	1	1	1	1	2	2	2	2

¿Cuál es la estimación de los parámetros del clasificador multinomial más probable?

A) $\hat{p}(1) = \frac{1}{3}$, $\hat{p}(2) = \frac{2}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{2}{5}, \frac{3}{5}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{3}, \frac{2}{3}\right)^t$
B) $\hat{p}(1) = \frac{2}{3}$, $\hat{p}(2) = \frac{1}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{2}{3}, \frac{3}{5}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{3}, \frac{2}{5}\right)^t$
C) $\hat{p}(1) = \frac{1}{3}$, $\hat{p}(2) = \frac{2}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{2}{3}, \frac{3}{5}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{3}, \frac{2}{5}\right)^t$
D) $\hat{p}(1) = \frac{2}{3}$, $\hat{p}(2) = \frac{1}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{2}{5}, \frac{3}{5}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{1}{3}, \frac{2}{3}\right)^t$

B Una técnica de suavizado para la distribución multinomial es el descuento absoluto. En este suavizado se utiliza una distribución generalizada para repartir la masa de probabilidad descontada. ¿Qué tipo de distribución generalizada se podría utilizar en cualquier caso?

A) Una distribución Bernoulli.
B) Una distribución multinomial.
C) Una distribución Gaussiana con matriz de covarianzas diagonal.
D) Una distribución Gaussiana con matriz de covarianzas completa.

B Un clasificador Gaussiano con matriz de covarianzas común es lineal, porque ...

A) el término cuadrático se hace cero.
B) el término cuadrático es constante.
C) el factor cuadrático W_c se convierte en la matriz identidad.
D) el factor cuadrático W_c tiene determinante nulo.

$\boxed{\text{C}}$ En general, para todos dos puntos $\mathbf{x}$,$\mathbf{y}$ en un espacio vectorial se cumple la siguiente relación de distancias

A) $L_1(\mathbf{x}, \mathbf{y}) \leq L_2(\mathbf{x}, \mathbf{y})$
B) $L_1(\mathbf{x}, \mathbf{y}) \leq L_0(\mathbf{x}, \mathbf{y})$
C) $L_2(\mathbf{x}, \mathbf{y}) \leq L_1(\mathbf{x}, \mathbf{y})$
D) $L_2(\mathbf{x}, \mathbf{y}) \leq L_0(\mathbf{x}, \mathbf{y})$

$\boxed{\text{B}}$ La frontera de decisión que se obtiene con el vecino más cercano es:

A) Lineal
B) Lineal a trozos
C) Lineal cuando hay más de dos clases
D) Ninguna de las anteriores

$\boxed{\text{D}}$ Sea X un conjunto de n muestras de aprendizaje, C el número de clases y k el parámetro de los k-vecinos, entonces el clasificador por los k vecinos es equivalente a:

A) $c(\mathbf{x}) = c$ si $\mathbf{x}_{nn} \in X_c$ y $k > 3$
B) $c(\mathbf{x}) = c$ si $\mathbf{x}_{nn} \in X_c$ y $k > n$
C) $c(\mathbf{x}) = \arg\min\ P(c)$ si $k = n$ ó $1 - nn$ en caso de empate
D) $c(\mathbf{x}) = \arg\max\ P(c)$ si $k = n$ ó $1 - nn$ en caso de empate

$\boxed{\text{C}}$ En general, ¿cuál de las siguientes afirmaciones es correcta?

A) El Bias aumenta al escoger clasificadores más fuertes
B) El Variance aumenta al incrementar el conjunto de muestras de aprendizaje
C) El Bias aumenta al escoger clasificadores más débiles
D) El Variance se reduce empleando Boosting

$\boxed{\text{B}}$ Esencialmente en Bagging:

A) Se combinan diferentes clasificadores sobre el mismo conjunto de aprendizaje
B) Se combinan el mismo clasificador sobre diferentes conjuntos de aprendizaje
C) Se combinan diferentes clasificadores sobre diferentes conjuntos de aprendizaje
D) Ninguna de las anteriores

$\boxed{\text{A}}$ En la teoría interactiva de la decisión, el criterio de decisión de una hipótesis h dada la señal x, la historia h' y la realimentación f es:

A) $\arg\max_h p(h|x, h', f)$
B) $\arg\max_h p(h, h', f|x)$
C) $\arg\max_h p(h, h', x|f)$
D) $\arg\max_h p(f|h, h', x)$

Examen de Teoría de Percepción - Recuperación Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2016

Apellidos: | **Nombre:**

Cuestiones (3 puntos, 30 minutos, sin apuntes)

$\boxed{\text{C}}$ Sean A, B y C tres clases con la misma probabilidad a priori y f.d. condicional de clase de tipo Bernoulli $\mathbf{p}_A = \begin{pmatrix} 0.1 \\ 0.9 \end{pmatrix}$, $\mathbf{p}_B = \begin{pmatrix} 0.4 \\ 0.6 \end{pmatrix}$, $\mathbf{p}_C = \begin{pmatrix} 0.8 \\ 0.5 \end{pmatrix}$. ¿En qué clase sería clasificada una muestra $\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ dados los parámetros anteriores?

A) Clase A
B) Clase B
C) Clase C
D) En cualquiera de las tres

$\boxed{\text{B}}$ Dado el siguiente conjunto de vectores binarios bidimensionales:

n	1	2	3	4	5	6	7	8	9	10	11	12
x_{n1}	1	1	0	1	1	1	0	1	1	0	1	1
x_{n2}	1	0	1	0	1	0	1	1	1	1	1	1
c_n	1	1	1	1	1	1	1	1	2	2	2	2

¿Cuál es la estimación de los parámetros del clasificador Bernoulli más probable?

A) $\hat{p}(1) = \frac{1}{3}$, $\hat{p}(2) = \frac{2}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{6}{11}, \frac{5}{11}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{3}{7}, \frac{4}{7}\right)^t$
B) $\hat{p}(1) = \frac{2}{3}$, $\hat{p}(2) = \frac{1}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{6}{8}, \frac{5}{8}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{3}{4}, \frac{4}{4}\right)^t$
C) $\hat{p}(1) = \frac{1}{3}$, $\hat{p}(2) = \frac{2}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{6}{8}, \frac{5}{8}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{3}{4}, \frac{4}{4}\right)^t$
D) $\hat{p}(1) = \frac{2}{3}$, $\hat{p}(2) = \frac{1}{3}$, $\hat{\mathbf{p}}_1 = \left(\frac{6}{11}, \frac{5}{11}\right)^t$ y $\hat{\mathbf{p}}_2 = \left(\frac{3}{7}, \frac{4}{7}\right)^t$

$\boxed{\text{B}}$ ¿Cuál de los siguientes valores del parámetro $\mathbf{p}$ no define una distribución multinomial?

A) $\mathbf{p} = \left(\frac{3}{4}, \frac{1}{4}\right)^t$
B) $\mathbf{p} = \left(\frac{3}{4}, \frac{3}{4}\right)^t$
C) $\mathbf{p} = \left(\frac{1}{2}, \frac{1}{2}\right)^t$
D) Todos los valores anteriores del parámetro $\mathbf{p}$ definen una distribución multinomial.

$\boxed{\text{C}}$ Sean A y B dos clases con igual prior y f.d.p. condicionales de clase gaussianas con los siguientes parámetros $\mu_A = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\Sigma_A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$, $\mu_B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ y $\Sigma_B = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$, ¿qué tipo de frontera de decisión definen?

A) Lineal
B) Lineal definida a trozos
C) Cuadrática
D) Ninguna de las anteriores

$\boxed{\text{A}}$ La frontera de decisión entre dos clases que se obtiene con el vecino más cercano cuando las clases tienen un único prototipo es:

A) Lineal
B) Lineal a trozos
C) Cuadrática
D) Ninguna de las anteriores

$\boxed{\text{D}}$ Sea la distancia Euclídea ponderada L_w con pesos postivos y no nulos:

A) $L_w(\mathbf{x}, \mathbf{y}) \leq L_2(\mathbf{x}, \mathbf{y})$
B) $L_1(\mathbf{x}, \mathbf{y}) \leq L_w(\mathbf{x}, \mathbf{y})$
C) $L_w(\mathbf{x}, \mathbf{y}) \geq L_0(\mathbf{x}, \mathbf{y})$
D) Ninguna de las anteriores

$\boxed{\text{D}}$ Sea P el error del clasificador de Bayes. El error del vecino más cercano $\hat{P}$ tiene la siguiente propiedad asintótica (cuando $n \to \infty$):

A) $\hat{P} = P$
B) $\hat{P} = 2P$
C) $\hat{P} \leq P$
D) $P \leq \hat{P} \leq 2P$

$\boxed{\text{B}}$ En general, ¿cuál de las siguientes afirmaciones es falsa?

A) El Variance aumenta al escoger clasificadores más fuertes
B) El Variance aumenta al escoger clasificadores más débiles
C) El Bias es menor en clasificadores más fuertes
D) El Bias se reduce empleando Boosting

$\boxed{\text{A}}$ Esencialmente en Boosting:

A) Se combinan diferentes clasificadores sobre el mismo conjunto de aprendizaje
B) Se combinan el mismo clasificador sobre diferentes conjuntos de aprendizaje
C) Se combinan diferentes clasificadores sobre diferentes conjuntos de aprendizaje
D) Ninguna de las anteriores

$\boxed{\text{C}}$ En la teoría interactiva de la decisión, el criterio de decisión de una hipótesis h dada la señal x, la historia h' y la realimentación f es:

A) $\arg\max_h p(h, h', f|x)$
B) $\arg\max_h p(h, h', x|f)$
C) $\arg\max_h p(h|x, h', f)$
D) $\arg\max_h p(f|h, h', x)$

Examen de Teoría de Percepción - Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2017

Apellidos: ______ **Nombre:** ______

Profesor: ☐ **Jorge Civera** ☐ **Carlos Martínez**

Cuestiones (2 puntos, 30 minutos, sin apuntes)

A La condición de Mercer para caracterizar una función Kernel $K(x, y)$ se puede formular como:

A) $\sum_{i=1}^{n} \sum_{j=1}^{n} K(\mathbf{x}_i, \mathbf{x}_j)\, c_i\, c_j \geq 0 \quad \forall c_i, c_j \in \mathbb{R}$
B) $\mathbf{z}^t \mathbf{K} \mathbf{z} \leq 0 \quad \forall \mathbf{z} \in \mathbb{R}^{n \times 1}$ con $\mathbf{z} \neq \mathbf{0}$ y matriz Gramm $\mathbf{K}$
C) $\sum_{i=1}^{n} K(\mathbf{x}_i, \mathbf{x}_i)\, c_i \geq 0 \quad \forall c_i \in \mathbb{R}$
D) $\sum_{i=1}^{n} \sum_{j=1}^{n} K(\mathbf{x}_i, \mathbf{x}_j)\, c_i \geq 0 \quad \forall c_i \in \mathbb{R}$

C Dados el prototipo Bernoulli $\hat{p} = (0.85\ 0.05\ 0.5\ 0.15\ 0.95)^t$ estimado a partir de un conjunto de vectores binarios, y su correspondiente prototipo Bernoulli suavizado mediante truncamiento simple $\tilde{p} = (0.85\ 0.1\ 0.5\ 0.15\ 0.9)^t$. ¿Qué valor de ϵ se ha utilizado para obtener dicho prototipo Bernoulli suavizado?

A) $\epsilon = 0.00$
B) $\epsilon = 0.05$
C) $\epsilon = 0.10$
D) $\epsilon = 0.15$

B Sea el conjunto de datos:

x_1	0	3	0	1	2	4	3	3	4	1
x_2	2	4	3	3	2	1	0	0	0	2
x_3	5	1	4	4	5	2	3	1	1	3
x_4	0	1	0	0	0	6	3	4	5	4
c	A	A	A	A	A	B	B	B	B	B

Indica la estimación por máxima verosimilitud de los parámetros de las distribuciones de Multinomiales para las clases A y B con ese conjunto de datos.

A) $\mathbf{p}_A = \left(\frac{6}{21}\, \frac{14}{17}\, \frac{19}{29}\, \frac{1}{23}\right), \mathbf{p}_B = \left(\frac{15}{21}\, \frac{3}{17}\, \frac{10}{29}\, \frac{22}{23}\right)$
B) $\mathbf{p}_A = \left(\frac{3}{20}\, \frac{7}{20}\, \frac{19}{40}\, \frac{1}{40}\right), \mathbf{p}_B = \left(\frac{3}{10}\, \frac{3}{50}\, \frac{1}{5}\, \frac{11}{25}\right)$
C) $\mathbf{p}_A = \left(\frac{1}{15}\, \frac{7}{45}\, \frac{19}{90}\, \frac{1}{90}\right), \mathbf{p}_B = \left(\frac{1}{6}\, \frac{1}{30}\, \frac{1}{9}\, \frac{11}{45}\right)$
D) $\mathbf{p}_A = \left(\frac{3}{20}\, \frac{7}{20}\, \frac{19}{40}\, \frac{1}{40}\right), \mathbf{p}_B = \left(\frac{7}{10}\, \frac{47}{50}\, \frac{4}{5}\, \frac{14}{25}\right)$

D Un clasificador gaussiano sobre un espacio vectorial de dimensión D:

A) Es, en general, lineal.
B) Tiene un número de parámetros igual a D.
C) No puede expresarse como una función discriminante.
D) Tiene como parámetros de la probabilidad condicional la media y la matriz de covarianzas.

$\boxed{\text{C}}$ Sea A y B dos clases con igual probabilidad a priori y probabilidades condicionales de clase gaussianas:

$$p(\mathbf{x} \mid A) \sim N_2(\mu_A, \Sigma_A) \qquad \text{y} \qquad p(\mathbf{x} \mid B) \sim N_2(\mu_B, \Sigma_B)$$

con

$$\mu_A = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \Sigma_A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad \mu_B = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \Sigma_B = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

¿Cuál de los siguientes pares de funciones discriminantes **no** define un clasificador equivalente al clasificador Gaussiano dado?

A) $g_A(x) = x_1 + \log \frac{1}{2} - \frac{1}{2} \qquad g_B(x) = \frac{1}{2}x_2 + \log \frac{1}{2} - \frac{1}{4}$
B) $g_A(x) = x_1 - \frac{1}{4} \qquad g_B(x) = \frac{1}{2}x_2$
C) $g_A(x) = 4x_1 + 2x_2 - 1 \qquad g_B(x) = 0$
D) Todos los anteriores pares de funciones discriminantes son equivalentes al clasificador Gaussiano dado

$\boxed{\text{B}}$ Sea X^k el conjunto de los $k \in \mathbb{N}^+$ prototipos más próximos a $\mathbf{y}$, ¿cuál de las siguientes expresiones representa un clasificador basado en los k-vecinos más cercanos?

A) $\hat{c}\,(y) = \arg\min_c |X^k \cap X_c|$
B) $\hat{c}\,(y) = \arg\max_c |X^k \cap X_c|$
C) $\hat{c}\,(y) = \arg\min_c \min_{\mathbf{x} \in X_c} d(\mathbf{x}, \mathbf{y})$
D) $\hat{c}\,(y) = \arg\min_c \max_{\mathbf{x} \in X_c} d(\mathbf{x}, \mathbf{y})$

$\boxed{\text{A}}$ Con un algoritmo de edición se consigue:

A) Eliminar prototipos que, generalmente, provocan errores de clasificación.
B) Reducir considerablemente el coste computacional de la clasificación.
C) Convertir las fronteras de decisión entre clases, usando NN, en lineales.
D) Seleccionar aquellos prototipos más cercanos a las fronteras de decisión.

$\boxed{\text{C}}$ En general, ¿cuál de las siguientes enumeraciones de clasificadores está ordenada de menor a mayor *bias* (de izquierda a derecha)?

A) k-NN, Multinomial, Gaussiano
B) Gaussiano, k-NN, Multinomial
C) k-NN, Gaussiano, Multinomial
D) Gaussiano, Multinomial, k-NN

Examen de Teoría de Percepción - Recuperación Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2017

Apellidos: ____________________ **Nombre:** ____________

Profesor: ☐ **Jorge Civera** ☐ **Carlos Martínez**

Cuestiones (2 puntos, 30 minutos, sin apuntes)

A Se define la función kernel $K(\mathbf{x},\mathbf{y}) = \frac{1}{d(\mathbf{x},\mathbf{y})+1}$ siendo $d(\mathbf{x},\mathbf{y})$ la distancia de Hamming $\left(d(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{D} 1 - \delta(x_i, y_i)\right)$, ¿cuál de las siguientes funciones **no** es un kernel?

A) $K(\mathbf{x},\mathbf{y}) = d(\mathbf{x},\mathbf{y}) + 1$
B) $K(\mathbf{x},\mathbf{y}) = \frac{1}{d(\mathbf{x},\mathbf{y})+1} + 1$
C) $K(\mathbf{x},\mathbf{y}) = e^{\frac{1}{d(\mathbf{x},\mathbf{y})+1}}$
D) $K(\mathbf{x},\mathbf{y}) = \frac{1}{2\cdot d(\mathbf{x},\mathbf{y})+2}$

B Sea el conjunto de datos:

x_1	0	1	1	0	0	1	1	1	1	0
x_2	0	0	0	1	0	1	0	0	1	1
x_3	1	1	0	0	1	0	0	0	1	0
c	A	A	A	A	A	B	B	B	B	B

Indica la estimación por máxima verosimilitud de los parámetros de las distribuciones de Bernoulli para las clases A y B con ese conjunto de datos.

A) $\mathbf{p}_A = \left(\frac{3}{5}\,\frac{4}{5}\,\frac{2}{5}\right), \mathbf{p}_B = \left(\frac{4}{5}\,\frac{3}{5}\,\frac{1}{5}\right)$
B) $\mathbf{p}_A = \left(\frac{2}{5}\,\frac{1}{5}\,\frac{3}{5}\right), \mathbf{p}_B = \left(\frac{4}{5}\,\frac{3}{5}\,\frac{1}{5}\right)$
C) $\mathbf{p}_A = \left(\frac{2}{5}\,\frac{1}{5}\,\frac{3}{5}\right), \mathbf{p}_B = \left(\frac{1}{5}\,\frac{2}{5}\,\frac{4}{5}\right)$
D) $\mathbf{p}_A = \left(\frac{7}{15}\,\frac{7}{15}\,\frac{7}{15}\right), \mathbf{p}_B = \left(\frac{8}{15}\,\frac{8}{15}\,\frac{8}{15}\right)$

C Dados el prototipo multinomial $\hat{p} = (0.3\ 0.0\ 0.4\ 0.0\ 0.5)^t$ estimado a partir de un conjunto de vectores de cuentas, y su correspondiente prototipo multinomial suavizado $\tilde{p} = (0.2\ 0.15\ 0.3\ 0.15\ 0.4)^t$. ¿Qué tipo de suavizado ha sido aplicado?

A) Laplace con $\epsilon = 0.1$
B) Laplace con $\epsilon = 0.2$
C) Descuento absoluto con $\epsilon = 0.1$ aplicando backing-off con distribución uniforme
D) Descuento absoluto con $\epsilon = 0.1$ aplicando interpolación con distribución uniforme

D ¿Cuál de las siguientes expresiones representa un clasificador basado en el vecino más cercano?

A) $\hat{c}(y) = \arg\min_{c} \max_{\mathbf{x}\in X_c} d(\mathbf{x},\mathbf{y})$
B) $\hat{c}(y) = \arg\max_{c} \max_{\mathbf{x}\in X_c} d(\mathbf{x},\mathbf{y})$
C) $\hat{c}(y) = \arg\max_{c} \min_{\mathbf{x}\in X_c} d(\mathbf{x},\mathbf{y})$
D) $\hat{c}(y) = \arg\min_{c} \min_{\mathbf{x}\in X_c} d(\mathbf{x},\mathbf{y})$

$\boxed{\text{B}}$ A diferencia de los clasificadores basados en Bayes, el clasificador k-NN:

A) Sólo puede aplicarse a datos no vectoriales.
B) Estima directamente $\hat{P}(c|x)$.
C) Alcanza una cota de error inferior a Bayes.
D) Crea un modelo basado en inferencia sobre los prototipos.

$\boxed{\text{D}}$ Para que el clasificador k-NN alcance el error de Bayes, siendo n el número de prototipos:

A) Debe usar un valor n potencialmente infinito, independientemente del k usado.
B) Debe usar un valor k potencialmente infinito, independientemente de n.
C) Debe usar un valores k y n potencialmente infinitos, con relación constante entre ellos.
D) Debe usar un valores k y n potencialmente infinitos, pero con k de crecimiento mucho más lento que n.

$\boxed{\text{A}}$ La distancia de Mahalanobis-diagonal:

A) Equivale a prenormalizar cada componente por la desviación típica y usar distancia euclídea.
B) Asigna pesos distintos por cada prototipo considerado.
C) Usa las varianzas por clase.
D) Incrementa la medida de distancia para las componentes que presentan una mayor dispersión.

$\boxed{\text{C}}$ En general, ¿cuál de las siguientes enumeraciones de clasificadores está ordenada de mayor a menor *variance* (de izquierda a derecha)?

A) k-NN, Multinomial, Gaussiano
B) Gaussiano, k-NN, Multinomial
C) k-NN, Gaussiano, Multinomial
D) Gaussiano, Multinomial, k-NN

Examen de Teoría de Percepción - Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2018

Apellidos: | | **Nombre:** | |

Profesor: ☐ **Jorge Civera** ☐ **Carlos Martínez**

Cuestiones (2 puntos, 30 minutos, sin apuntes)

$\boxed{\text{C}}$ Ante un conjunto de muestras linealmente separable, ¿qué tipo de kernel es preferible aplicar?

A) Un kernel polinomial
B) Un kernel gaussiano
C) No es necesario aplicar kernels
D) Cualquier kernel

$\boxed{\text{D}}$ Dada una función kernel $K(\mathbf{x}, \mathbf{y})$, ¿cuál de las siguientes **no** es una función kernel?

A) $\mathbf{x}^2 K(\mathbf{x}, \mathbf{y})\mathbf{y}^{-1}$
B) $3K(\mathbf{x}, \mathbf{y})^2 + K(\mathbf{x}, \mathbf{y})$
C) $(5 + K(\mathbf{x}, \mathbf{y}))^3$
D) $-K(\mathbf{x}, \mathbf{y})$

$\boxed{\text{B}}$ Sean A y B dos clases equiprobables de objetos en $\mathbf{x} = (x_1, x_2)^t \in \{0, 1\}^2$, con f.d. condicional Bernoulli con parámetros:

$$p_A = \left(\frac{1}{4}, \frac{1}{2}\right) \qquad\qquad p_B = \left(\frac{1}{2}, \frac{1}{8}\right)$$

¿Cuál de los siguientes clasificadores realiza la misma asignación de clases que el clasificador Bernoulli dado para cada uno de los objetos binarios en $\mathbb{R}^2$?

A) $g_A(x) = x_1$, $g_B(x) = x_1 + x_2$
B) $g_A(x) = x_2$, $g_B(x) = 1 - x_2$
C) $g_A(x) = x_1$, $g_B(x) = x_2$
D) $g_A(x) = x_1$, $g_B(x) = 1 - x_1$

$\boxed{\text{D}}$ El número de parámetros de un clasificador gaussiano general es:

A) Independiente del número de clases.
B) Independiente de la dimensión de los datos.
C) Lineal con la dimensión y cuadrático con el número de clases.
D) Cuadrático con la dimensión y lineal con el número de clases.

$\boxed{\text{A}}$ Sea X_c el conjunto de prototipos de la clase c y $X^k(y)$ el conjunto de los k prototipos más cercanos a y. ¿Cuál de las siguientes reglas de clasificación representa la del clasificador por k-vecinos más cercanos?

A) $\hat{c}(y) = \arg\max_{c \in X_c} |X^k(y) \cap X_c|$
B) $\hat{c}(y) = \arg\max_{c \in X_c} |X^k(y) \cup X_c|$
C) $\hat{c}(y) = \arg\min_{c \in X_c} |X^k(y) \cap X_c|$
D) $\hat{c}(y) = \arg\min_{c \in X_c} |X^k(y) \cup X_c|$

$\boxed{\text{B}}$ El aprendizaje de distancias adaptadas a un conjunto de prototipos emplea ponderaciones que se basan en:

A) El número de prototipos disponible.
B) Las varianzas de los prototipos.
C) Las medias de los prototipos.
D) El número de vecinos elegido.

$\boxed{\text{C}}$ El algoritmo *bagging* consiste en:

A) Una combinación de clasificadores fuertes obtenida por ponderación no homogénea de las muestras del conjunto de entrenamiento.

B) Una combinación de clasificadores fuertes obtenida de un conjunto de entrenamiento no ruidoso.
C) Una combinación de clasificadores fuertes obtenida por particionado del conjunto de entrenamiento.
D) Una combinación de clasificadores débiles obtenida por particionado del conjunto de entrenamiento.

$\boxed{\text{A}}$ Al aplicar *AdaBoost*:

A) Se escogen tanto los clasificadores como el peso de clasificadores y muestras de entrenamiento.
B) Se combinan clasificadores fuertes.
C) Se consigue una reducción de *variance*
D) Se finaliza cuando el error del clasificador escogido en una iteración es lo bastante bajo.

Examen de Teoría de Percepción - Recuperación Segundo Parcial

ETSINF, Universitat Politècnica de València, Junio de 2018

Apellidos: ____________________ **Nombre:** ____________

Profesor: ☐ **Jorge Civera** ☐ **Carlos Martínez**

Cuestiones (2 puntos, 30 minutos, sin apuntes)

B Dado un conjunto de entrenamiento $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, una función K es kernel sobre el mismo si la matriz Gramm $\mathbf{K}$ asociada es:

A) Simétrica
B) Semidefinida positiva
C) Triangular superior
D) Singular

D Sea un conjunto de muestras de entrenamiento $X = \{(\mathbf{x}_1, -1), (\mathbf{x}_2, -1), (\mathbf{x}_3, +1)\}$ para la cual se define la función kernel $K(\mathbf{x}, \mathbf{y})$. Si se aplica una iteración del algoritmo Kernel Perceptron sobre X, tras la primera iteración los pesos resultantes son $\alpha = (1, 1, 0)$, ¿cuál es la función discriminante correspondiente a estos pesos?

A) $g(\mathbf{x}) = K(\mathbf{x}_1, \mathbf{x}) + K(\mathbf{x}_2, \mathbf{x}) + 2$
B) $g(\mathbf{x}) = K(\mathbf{x}_1, \mathbf{x}) + K(\mathbf{x}_2, \mathbf{x}) - 2$
C) $g(\mathbf{x}) = -K(\mathbf{x}_1, \mathbf{x}) - K(\mathbf{x}_2, \mathbf{x}) + 2$
D) $g(\mathbf{x}) = -K(\mathbf{x}_1, \mathbf{x}) - K(\mathbf{x}_2, \mathbf{x}) - 2$

C Dado el siguiente prototipo multinomial $\hat{\mathbf{p}} = (0.4\ 0.2\ 0.4\ 0.0)^t$, ¿cuál sería su versión suavizada $\tilde{\mathbf{p}}$ mediante Laplace con $\epsilon = 0.2$?

A) $\tilde{\mathbf{p}} = \left(\frac{5}{10}\ \frac{2}{10}\ \frac{5}{10}\ \frac{1}{10}\right)^t$
B) $\tilde{\mathbf{p}} = \left(\frac{3}{10}\ \frac{1}{10}\ \frac{3}{10}\ \frac{3}{10}\right)^t$
C) $\tilde{\mathbf{p}} = \left(\frac{3}{9}\ \frac{2}{9}\ \frac{3}{9}\ \frac{1}{9}\right)^t$
D) $\tilde{\mathbf{p}} = \left(\frac{2}{9}\ \frac{1}{9}\ \frac{2}{9}\ \frac{4}{9}\right)^t$

A En general, ¿cuál es el principal objetivo del suavizado de la matriz de covarianzas en un clasificador gaussiano?

A) Corregir su estimación por no disponer de una muestra suficientemente representativa de la población
B) Corregir la aparición de valores excesivamente elevados, y por tanto evitar probabilidades cero
C) Evitar valores extremos, tanto cero como uno, en la matriz de covarianzas, y por tanto evitar probabilidades cero
D) Evitar valores cero en la matriz de covarianza, y por tanto evitar probabilidades cero

D Dado el siguiente conjunto de datos en Σ^3, con $\Sigma = \{a, b, c\}$:

x_n	aba	aca	aaa	abb	bba	bcb	cbc	cca	cac
c_n	A	A	A	B	B	B	C	C	C

Indicar la clasificación por vecino más cercano del objeto representado por *cbb* emplando la distancia de Hamming ($d(s,t) = \sum_{i:s_i \neq t_i} 1$)

A) Clase A
B) Clase B
C) Clase C
D) Hay un empate entre las clases B y C, se elegiría una al azar

C Al aplicar edición de prototipos, se espera que:

A) El número de prototipos se reduzca de forma drástica
B) Se den fronteras de decisión lineales
C) Queden regiones de decisión simplemente conexas
D) Desaparezcan todos los puntos más cercanos a las fronteras de decisión iniciales

A Las fuentes de error de un clasificador son *Bias*, *Variance* y *Noise*. En general, ¿qué podrías afirmar de los clasificadores que has estudiado en la asignatura?

A) Aquellos que tienen un bajo *Bias*, tienen un alto *Variance*
B) Aquellos que tienen un bajo *Bias*, tienen un bajo *Variance*
C) Todos ellos tienen un bajo *Bias*, pero diferentes *Variance*
D) Todos ellos tienen un bajo *Variance*, pero diferentes *Bias*

$\boxed{\text{B}}$ Sobre el algoritmo AdaBoost, ¿cuál de las siguientes afirmaciones es correcta?

A) Existe un peso diferente asociado a cada dimensión de los datos
B) El número de clasificadores débiles disponibles es arbitrario
C) El error máximo de un clasificador débil seleccionado puede superar el 50 %
D) No es posible seleccionar el mismo clasificador débil dos veces

Examen de Teoría de Percepción - Segundo Parcial

ETSINF, Universitat Politécnica de Valéncia, Junio de 2019

Apellidos: | | **Nombre:** | |

Profesor: ☐ **Jorge Civera** ☐ **Carlos Martínez**

Cuestiones (2 puntos, 30 minutos, sin apuntes)

[C] En el algoritmo Kernel Perceptron, cada componente del vector de salida α_i se interpreta como:

A) El peso que se le da sólo a la componente i-ésima
B) El número de veces que la muestra i-ésima se ha clasificado correctamente en el proceso
C) El número de veces que la muestra i-ésima se ha clasificado incorrectamente en el proceso
D) El peso que se le da sólo al término independiente i-ésimo

[B] ¿Qué valor de ϵ en truncamiento simple se ha aplicado al parámetro Bernoulli $\hat{\mathbf{p}} = (0 \ \ 0 \ \ 1 \ \ 1)^t$ para obtener el parámetro suavizado $\tilde{\mathbf{p}} = (0.01 \ \ 0.01 \ \ 0.99 \ \ 0.99)^t$?

A) $\epsilon = 10^{-1}$
B) $\epsilon = 10^{-2}$
C) $\epsilon = 10^{-3}$
D) $\epsilon = 10^{-4}$

[D] El clasificador multinomial que sigue $Mult_D(x_+, \mathbf{p}_c)$ se aplica sobre objetos que representan:

A) Valores naturales entre 1 y D
B) Proporciones de valores en una población de elementos en el rango 1-D
C) Secuencias de valores en el rango 1-D
D) Vectores de número de ocurrencias de valores en el rango 1-D en una secuencia

[B] En la estimación por máxima verosimilitud de un clasificador gaussiano con matriz de covarianza común se estiman:

A) Probabilidades a priori y matrices de covarianza de cada clase
B) Probabilidades a priori y medias de cada clase, y la matriz de covarianzas del total de datos de entrenamiento
C) Probabilidades a priori de cada clase, y la media y matriz de covarianzas del total de datos de entrenamiento
D) Medias de cada clase y la matriz de covarianzas del total de datos de entrenamiento

$\boxed{\text{A}}$ Al aplicar suavizado por umbralizado de covarianza con $\epsilon > 0$, la matriz de covarianzas que se espera se caracterizará por:

A) Tener los mismos o más ceros fuera de la diagonal que la original
B) Tener los mismos o más ceros en la diagonal que la original
C) Ser diagonal
D) Tener menos ceros que la original

$\boxed{\text{X}}$ En general, la frontera de decisión entre dos clases de un clasificador basado en el vecino más cercano es una frontera lineal definida a trozos, pero que no es lineal globalmente. ¿En cuál de los siguientes casos esto es así? Pregunta cancelada por ser tanto B como C correctas

A) Cuando únicamente se dispone de una muestra de cada clase
B) Cuando las muestras de cada clase se disponen en rectas diferentes y éstas rectas son paralelas
C) Cuando las muestras de cada clase se disponen en rectas diferentes y éstas rectas intersectan
D) En todos los casos anteriores.

$\boxed{\text{D}}$ Cuando se hace aprendizaje de distancias, generalmente se hace:

A) Una selección de los prototipos que mejor contribuyen a la distancia
B) Una proyección sobre un espacio linealmente separable
C) Un cambio del exponente al que se eleva la diferencia entre componentes
D) Un escalado de las distintas componentes de los datos

$\boxed{\text{A}}$ En general, ¿cuál de las siguientes características es propia de un clasificador fuerte?

A) Tienen una precisión alta
B) Necesitan pocos datos de entrenamiento
C) Su sesgo (*bias*) es alto
D) Permiten cancelar el ruido (*noise*)

Examen de Teoría de Percepción - Recuperación Segundo Parcial

ETSINF, Universitat Politécnica de Valéncia, Junio de 2019

Apellidos: | | **Nombre:** | |

Profesor: ☐ **Jorge Civera** ☐ **Carlos Martínez**

Cuestiones (2 puntos, 30 minutos, sin apuntes)

$\boxed{\text{D}}$ Al aplicar funciones kernel:

A) Se proyecta explícitamente a un espacio alternativo donde las muestras serán linealmente separables
B) Se proyecta explícitamente a un espacio alternativo, aunque no se garantiza siempre si las muestras serán linealmente separables
C) No se proyecta explícitamente a un espacio alternativo, aunque sí se garantiza la separabilidad lineal de las muestras en cualquier caso
D) No se proyecta explícitamente a un espacio alternativo ni se garantiza la separabilidad lineal en todos los casos

$\boxed{\text{C}}$ Dada la función kernel $K(\mathbf{x}, \mathbf{y})$, con $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{D\times 1}$, indicar cuál de las siguientes funciones derivadas sería un kernel

A) $1 + K(\mathbf{x}, \mathbf{y})^{-1}$
B) $K(\mathbf{x}, \mathbf{y})^2 - K(\mathbf{x}, \mathbf{y})$
C) $2\mathbf{x}^t K(\mathbf{x}, \mathbf{y})\mathbf{y}$
D) $-K(\mathbf{x}, \mathbf{y})$

$\boxed{\text{B}}$ En la estimación de un clasificador probabilístico, habitualmente se asume una distribución de probabilidad en uno de sus factores y se estiman sus parámetros. ¿Sobre qué término se suele hacer?:

A) Sobre la probabilidad *a posteriori* $P(c|x)$
B) Sobre la probabilidad condicionada $p(x|c)$
C) Sobre la probabilidad *a priori* $P(c)$
D) Sobre la probabilidad incondicional $p(x)$

$\boxed{\text{B}}$ ¿Cuál de las siguientes es una forma correcta del clasificador de Bernoulli de parámetros $\Theta = \{P(1), \dots, P(C), p_1, \dots, p_C\}$?

A) $c^*(x) = \arg\max_{c=1,\dots,C} \prod_{d=1}^{D} p_{cd}^{x_d}(1 - p_{cd})^{(1-x_d)}$
B) $c^*(x) = \arg\max_{c=1,\dots,C} P(c) \prod_{d=1}^{D} p_{cd}^{x_d}(1 - p_{cd})^{(1-x_d)}$
C) $c^*(x) = \arg\max_{c=1,\dots,C} \log P(c) + \sum_{d=1}^{D} \log p_{cd}^{x_d} + \log(1 - p_{cd})^{x_d}$
D) $c^*(x) = \arg\max_{c=1,\dots,C} \sum_{d=1}^{D} p_{cd}^{x_d}(1 - p_{cd})^{(1-x_d)}$

$\boxed{\text{A}}$ Sea una distribución multinomial de parámetro $\mathbf{p} = \left(\frac{1}{4}, \frac{1}{2}, 0, \frac{1}{8}, \frac{1}{8}\right)$. Al aplicar descuento absoluto de valor $\epsilon = \frac{1}{16}$ y distribuir por *back-off*, ¿cuál es el prototipo multinomial resultante?

A) $\left(\frac{3}{16}, \frac{7}{16}, \frac{4}{16}, \frac{1}{16}, \frac{1}{16}\right)$
B) $\left(\frac{19}{80}, \frac{39}{80}, \frac{4}{80}, \frac{9}{80}, \frac{9}{80}\right)$
C) $\left(\frac{5}{21}, \frac{9}{21}, \frac{1}{21}, \frac{3}{21}, \frac{3}{21}\right)$
D) $\left(\frac{4}{16}, \frac{8}{16}, \frac{1}{16}, \frac{2}{16}, \frac{2}{16}\right)$

$\boxed{\text{C}}$ Con un clasificador gaussiano general se obtienen fronteras de decisión de tipo:

A) Lineal
B) Lineal a trozos
C) Cuadráticas
D) De otro tipo

$\boxed{\text{D}}$ A diferencia del clasificador de Bayes, un clasificador k-NN:

A) Nunca puede ser óptimo
B) No puede verse con un equivalente probabilístico
C) Se puede aplicar también a datos no vectoriales
D) Realiza estimaciones directas de $P(c|x)$

$\boxed{\text{A}}$ Si se aplica condensado sobre un conjunto no editado:

A) Se pueden mantener prototipos fuera de norma (*outliers*)
B) No se reducirá en ningún caso el conjunto de prototipos
C) Se mantiene el coste computacional con respecto al que tendría con un conjunto editado
D) No se mantienen los prototipos cercanos a las fronteras de decisión