

Examen de Teoría de Percepción
ETSINF, Universitat Politècnica de València, Marzo de 2015

Apellidos: Nombre:

Profesor: ☐ Carlos Martínez ☐ Roberto Paredes

Cuestiones (3 puntos, 30 minutos, sin apuntes)

- ☐ C Indicar la afirmación correcta respecto a la clasificación estadística:
- A) La regla óptima de clasificación es $c(x) = \arg \max_{c=1 \dots C} p(\mathbf{x}|c)$
 - B) Se puede aplicar que $P(c|\mathbf{x}) = P(c)p(\mathbf{x}|c)$
 - C) El clasificador basado en distancias puede verse como uno estadístico asumiendo $P(c|\mathbf{x}) = \frac{k_c}{K}$
 - D) Sólo es aplicable cuando las probabilidades *a priori* son las mismas (equiprobables)
- ☐ D Se tienen las funciones discriminantes (para puntos en \mathbb{R}^2) $g_1(\mathbf{x}) = 2x_1 - 2x_2 + 3$ y $g_2(\mathbf{x}) = 3x_1 + x_2 - 2$; el punto (2,1) se clasificaría en:
- A) La clase 1
 - B) La clase 2
 - C) Ninguna de las clases, al no superar el umbral mínimo
 - D) En una clase al azar, al situarse en la frontera de decisión
- ☐ A Se tiene un problema de reconocimiento de imágenes, donde las imágenes son de 20×20 píxeles a 256 niveles de gris, y se plantea su representación local usando ventanas de 7×7 . ¿Qué afirmación es correcta?
- A) Con rejilla de un píxel, habrá un total de 196 ventanas
 - B) La representación local ocupará siempre más que la global
 - C) Con rejilla de un píxel, la representación local requiere más de 10000 bytes
 - D) Cada ventana requerirá más bytes que la representación global
- ☐ B Indicar cuál de las siguientes características *no* es propia del banco de filtros de Mel:
- A) Generalmente emplea filtros triangulares
 - B) Transforma la señal del dominio temporal al frecuencial
 - C) Imita la percepción humana, afinando para frecuencias bajas y siendo menos fino para las altas
 - D) Permite una reducción de dimensión de la representación frecuencial
- ☐ D Dada una colección de 50 documentos, de longitud máxima 65535 *tokens* sobre un conjunto de 220 *tokens* distintos, indicar cuánto ocuparía (aproximadamente) una representación (densa) de cuentas de bigramas de *tokens* para toda la colección:
- A) 21 Kbytes
 - B) 0.8 Mbytes
 - C) 2.3 Mbytes
 - D) 4.6 Mbytes
- ☐ A Dado un espacio de representación de d dimensiones se desea reducir a k dimensiones mediante PCA. Para ello se dispone de una matriz $B_{n \times d}$ compuesta por los n vectores de entrenamiento, menos la media de dichos vectores, dispuestos en filas. Entonces:
- A) Escogeremos los k mayores eigenvectores (mayor eigenvalor asociado) de la matriz: $\frac{1}{n} B^t B$
 - B) Escogeremos los k mayores eigenvectores (mayor eigenvalor asociado) de la matriz: $\frac{1}{n} B B^t$
 - C) Escogeremos los k mayores eigenvectores (mayor eigenvalor asociado) de la matriz: $\frac{1}{n} B$
 - D) Escogeremos los k mayores eigenvectores (mayor eigenvalor asociado) de la matriz: $\frac{1}{n} B^t$

D Dado un problema de clasificación en 5 clases donde los objetos se representan en un espacio de 10 dimensiones. Se desea obtener una representación en un espacio reducido de 2 dimensiones. En general, ¿cuál de las siguientes reducciones es la menos aconsejable?

- A) Proyectar primero con PCA a 10 dimensiones y luego con LDA a 2
- B) Proyectar primero con PCA a 9 dimensiones y luego con LDA a 2
- C) Proyectar con LDA a 2 dimensiones y luego con PCA a 2
- D) Proyectar primero con PCA a 4 dimensiones y luego con LDA a 2

D Dado un problema de clasificación en C clases donde los objetos se representan en un espacio de d dimensiones, se desea obtener una representación en un espacio reducido de k dimensiones. Mediante PCA se obtiene W como matriz de proyección a d' dimensiones y a partir de los datos una vez proyectados mediante PCA se obtiene V como la matriz de proyección mediante LDA. Se debe de cumplir que:

- A) $k \leq d'$ y $d' \leq C - 1 \leq d$
- B) $k \leq C - 1 \leq d' \leq d$
- C) $k \leq d'$ y $d' \leq d$
- D) $k \leq C - 1$ y $k \leq d' \leq d$

B Indica cuál de las siguientes afirmaciones es *falsa*

- A) El algoritmo Kernel Perceptron incrementa la importancia (peso) de las muestras incorrectamente clasificadas
- B) El uso de kernels es adecuado cuando los objetos son linealmente separables en el espacio de representación original
- C) El uso de kernels es adecuado cuando los objetos no son linealmente separables en el espacio de representación original
- D) Las funciones kernel modelan el producto escalar de dos vectores en un espacio de representación alternativo

C Sean $K_1(\mathbf{x}, \mathbf{y})$ y $K_2(\mathbf{x}, \mathbf{y})$ dos kernels, indica cuál de las siguientes expresiones *no* es un kernel

- A) $(\mathbf{x}^t \mathbf{x}) \cdot K_1(\mathbf{x}, \mathbf{y}) \cdot (\mathbf{y}^t \mathbf{y})$
- B) $\exp(K_1(\mathbf{x}, \mathbf{y}) + 1)$
- C) $(c + K_1(\mathbf{x}, \mathbf{y}))^d \quad c, d \geq 0$
- D) $K_1(\mathbf{x}, \mathbf{y}) + 2K_2(\mathbf{x}, \mathbf{y})$

Examen de Teoría de Percepción
ETSINF, Universitat Politècnica de València, Abril de 2016

Apellidos:

Nombre:

Profesor: ☐ Carlos Martínez ☐ Roberto Paredes

Cuestiones (3 puntos, 30 minutos, sin apuntes)

- ☐ C ¿En qué situación el clasificador por máxima verosimilitud puede expresarse por $c(x) = \arg \max p(x|c)$?
- A) En cualquier situación
 - B) En ninguna situación
 - C) Cuando las probabilidades *a priori* $P(c)$ son iguales para todas las clases
 - D) Cuando hay independencia de $p(x)$
- ☐ B Supongamos una imagen en escala de grises de tamaño 10×15 centímetros, escaneada con una resolución de 250 puntos por centímetro. ¿Qué tamaño ocuparía suponiendo que se cuantifican los niveles de gris en un único byte?
- A) Menos de 5 Mbytes
 - B) Entre 5 y 10 Mbytes
 - C) Entre 10 y 15 Mbytes
 - D) Más de 15 Mbytes
- ☐ C Tenemos un sistema de reconocimiento de imágenes de gran tamaño con restricciones temporales críticas. ¿Cuál de las siguientes técnicas de detección de puntos de interés para extracción de características locales es más apropiada?
- A) Detector de contorno
 - B) Detector de esquinas complementado con extracción por rejilla
 - C) Extracción por rejilla complementada por extracción aleatoria
 - D) Cualquier técnica basada en información e invarianza
- ☐ D Ante una señal de audio que recoge un rango de frecuencias hasta 16KHz, pero de las cuáles sólo son informativas las frecuencias hasta 8KHz, ¿qué procesos de adquisición son los más apropiados?
- A) Muestreo directo a 8KHz
 - B) Muestreo directo a 16KHz
 - C) Filtrado paso bajo de 4KHz y muestreo a 8KHz
 - D) Filtrado paso bajo de 8KHz y muestreo a 16KHz
- ☐ A Emplear una función global en la representación de *tokens* para clasificación de documentos es conveniente porque:
- A) Se emplea información de la colección de documentos completa
 - B) Reduce el tamaño de los vectores de representación
 - C) Permite la introducción de contexto basado en secuencias de palabras
 - D) Permite una representación logarítmica de los documentos, dando mayor capacidad discriminativa

C PCA se resuelve minimizando el error de reconstrucción. Al final se llega a una solución basada en eigenvectores de la matriz de covarianzas. Al aplicar la transformación lineal con dichos eigenvectores se tiene que:

- A) El error de reconstrucción es 0
- B) El error de reconstrucción es la suma de los eigenvalores de los eigenvectores empleados
- C) La matriz de covarianzas en el espacio proyectado es diagonal
- D) El error de reconstrucción es la suma de los eigenvectores empleados

A ¿Cuál de estas afirmaciones sobre LDA 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 $d - 1$ eigenvectores siendo d la dimensionalidad de los datos originales
- C) Es una proyección lineal que resulta del análisis de eigenvectores de la matriz de covarianzas Σ_x
- D) LDA obtiene una matriz de proyección lineal optimizando una función objetivo que persigue maximizar las distancias intraclase mientras se minimizan las interclase

DCuál de las siguientes afirmaciones respecto a kernels es falsa:

- A) Las funciones kernel modelan el producto escalar en un espacio de representación alternativo
- B) El uso de kernels es adecuado cuando el espacio de representación original no es linealmente separable
- C) El algoritmo Kernel Perceptron acaba cuando todas las muestras de aprendizaje están bien clasificadas
- D) Las funciones kernel implican tener que proyectar los datos a un espacio de representación alternativo

C Dado un problema de clasificación en 10 clases donde los objetos se representan en un espacio de 1000 dimensiones. Se desea obtener una representación en un espacio reducido de 2 dimensiones. En general, ¿cuál de las siguientes reducciones es la menos aconsejable desde el punto de vista de estabilidad numérica?

- A) Proyectar primero con PCA a 100 dimensiones y luego con LDA a 2
- B) Proyectar primero con PCA a 10 dimensiones y luego con LDA a 2
- C) Proyectar con LDA a 2
- D) Proyectar primero con PCA a 50 dimensiones y luego con LDA a 2

B Dado un problema de clasificación en C clases donde los objetos se representan en un espacio de d dimensiones, se desea obtener una representación en un espacio reducido de k dimensiones. Mediante PCA se obtiene W como matriz de proyección a d' dimensiones y a partir de los datos una vez proyectados mediante PCA se obtiene V como la matriz de proyección mediante LDA. Se debe de cumplir que:

- A) $d' \leq C - 1$ y $k \leq C - 1 \leq d$
- B) $k \leq C - 1$ y $k \leq d' \leq d$
- C) $k \leq C - 1 \leq d' \leq d$
- D) $d' \leq C - 1$ y $k \leq d$

Examen de Teoría de Percepción - Recuperación Primer Parcial

ETSINF, Universitat Politècnica de València, Junio de 2016

Apellidos:

Nombre:

Profesor: ☐ Carlos Martínez ☐ Roberto Paredes

Cuestiones (3 puntos, 30 minutos, sin apuntes)

- ☐ C ¿En qué situación el clasificador por máxima verosimilitud puede expresarse por $c(x) = \arg \max_c p(x)p(x|c)$?
- A) En cualquier situación
 - B) En ninguna situación
 - C) Cuando las probabilidades *a priori* $P(c)$ son iguales para todas las clases
 - D) Cuando hay independencia de $p(x)$
- ☐ D Sean las funciones discriminantes sobre \mathbb{R}^2 $g_A(x) = 2x_1^2 - x_1 + x_2 - 5$, $g_B(x) = -x_1 + 3x_2^2 + 2$, $g_C(x) = 3x_1 - 2x_2 + 4$; indicar en qué clase se clasificaría la muestra $(1, 1)$:
- A) En la clase A
 - B) En la clase B
 - C) Al azar entre B y C por dar el mismo valor sus funciones discriminantes
 - D) En la clase C
- ☐ A ¿Cuál es el objetivo del paso de escalado durante el preproceso de OCR?
- A) Uniformizar el tamaño final de las imágenes
 - B) Eliminar las partes no relevantes de la imagen
 - C) Igualar las dimensiones horizontal y vertical de la imagen
 - D) Conseguir una imagen sin transiciones abruptas
- ☐ B Tenemos una señal de audio de 5 segundos de duración a la cuál se le aplica una ventana de Hamming de anchura $W = 20ms$ y desplazamiento $S = 10ms$. ¿Qué cantidad aproximada de marcos (*frames*) se puede esperar de este proceso?
- A) 250
 - B) 500
 - C) 350
 - D) 700
- ☐ C Indicar la afirmación **incorrecta** respecto a la representación *bag-of-words* de documentos de texto:
- A) Cada documento se representa por un vector de tamaño igual al del vocabulario
 - B) Los vectores pueden ser binarios o de números naturales (*term frequency*)
 - C) El contenido de cada posición del vector de representación indica la frecuencia de una secuencia de $n > 1$ palabras
 - D) La colección de documentos se puede representar como una matriz de tamaño $V \times D$ (V tamaño del vocabulario, D número de documentos)
- ☐ A Con respecto a PCA ¿cuál de las siguientes afirmaciones es falsa?
- A) PCA preserva la separabilidad de las clases cuando se aplica después de LDA
 - B) PCA minimiza el error de reconstrucción de las muestras
 - C) Con PCA obtenemos una proyección lineal a partir de los k eigenvectores con mayor eigenvalor asociado
 - D) Con PCA deberemos restar la media de los vectores previamente a realizar la proyección

D Sea un problema de clasificación en 10 clases donde las muestras se representan en 1000 dimensiones. En general, ¿cuál de las siguientes opciones de reducción de dimensionalidad es la menos adecuada?

- A) Reducir con PCA a 10 dimensiones y con LDA a 9
- B) Reducir con PCA a 100 dimensiones y con LDA a 9
- C) Reducir con PCA a 100 dimensiones y con LDA a 2
- D) Reducir con PCA a 10 dimensiones y con LDA a 2

D Dado un problema de clasificación en C clases donde los objetos se representan en un espacio de representación de d dimensiones, se desea obtener una representación final en un espacio reducido de k dimensiones. Para ello se realizará primero una proyección mediante PCA a d' dimensiones con el fin de evitar singularidades, para posteriormente mediante LDA una proyección final a las k dimensiones. Por lo tanto se debe cumplir que, en general:

- A) $d' \leq C - 1$ y $k \leq d$
- B) $k \leq C - 1$ y $d' \leq d$
- C) $d' \leq \min(C - 1, d)$ y $k \leq d$
- D) $k \leq \min(C - 1, d')$ y $d' \leq d$

D Dado el conjunto de entrenamiento $X = \{(\mathbf{x}_1, c_1), (\mathbf{x}_2, c_2), \dots, (\mathbf{x}_n, c_n)\}$, la función discriminante asociada al problema de clasificación binaria empleando kernels es:

- A) $g(\mathbf{x}) = \sum_{i=1}^n \alpha_i c_i K(\mathbf{x}, \mathbf{x}_i) + \sum_{i=1}^n \alpha_i$
- B) $g(\mathbf{x}) = \sum_{i=1}^n \alpha_i c_i K(\mathbf{x}, \mathbf{x}_i) + \sum_{i=1}^n c_i$
- C) $g(\mathbf{x}) = \sum_{i=1}^n c_i K(\mathbf{x}, \mathbf{x}_i) + \sum_{i=1}^n \alpha_i c_i$
- D) $g(\mathbf{x}) = \sum_{i=1}^n \alpha_i c_i K(\mathbf{x}, \mathbf{x}_i) + \sum_{i=1}^n \alpha_i c_i$

B Sean $K_1(\mathbf{x}, \mathbf{y})$ y $K_2(\mathbf{x}, \mathbf{y})$ dos kernels, indica cuál de las siguientes expresiones *no* es un kernel

- A) $(\mathbf{x}^t \mathbf{x}) \cdot K_1(\mathbf{x}, \mathbf{y}) \cdot (\mathbf{y}^t \mathbf{y})$
- B) $aK_1(\mathbf{x}, \mathbf{y})^2 + bK_2(\mathbf{x}, \mathbf{y})$
- C) $(c + K_1(\mathbf{x}, \mathbf{y}))^d$ $c, d > 0$
- D) $\exp(K_1(\mathbf{x}, \mathbf{y}) + 1)$

Examen de Teoría de Percepción - Primer Parcial

ETSINF, Universitat Politècnica de València, Abril de 2017

Apellidos:

Nombre:

Profesor: ☐ Jorge Civera ☐ Carlos Martínez

Cuestiones (2 puntos, 30 minutos, sin apuntes)

- ☐ C Dado un clasificador definido por $c(x) = \arg \max_{c=1,\dots,C} g_c(x)$. Indica cuál de las siguientes definiciones de $g_c(x)$ hace que se corresponda a un clasificador de mínimo error:
- A) $g_c(x) = P(c) P(c|x)$
 - B) $g_c(x) = w_c x + w_{c0}$, con w_c y w_{c0} aprendidos por el algoritmo Perceptron
 - C) $g_c(x) = P(c|x) p(x)$
 - D) $g_c(x) = P(x|c)$
- ☐ B Indica la afirmación correcta para un sistema de reconocimiento interactivo:
- A) Se evalúan según el error final cometido
 - B) Emplean la realimentación del usuario
 - C) Siempre requieren una evaluación manual
 - D) Usan entrenamiento *on-line* en todas las interacciones del usuario
- ☐ C Se está resolviendo un problema de clasificación de imágenes de tamaño 100×100 píxeles. Se propone implementar sistemas usando características globales o características locales con ventanas de 5×5 píxeles. Indicar qué afirmación de las siguientes se dará siempre en estos sistemas:
- A) La clasificación por características locales será mejor que la clasificación por representación global
 - B) El vector de características de la representación global ocupará más memoria que el conjunto de características locales
 - C) La dimensión del vector de representación global será 10000 y la de representación local 25
 - D) El número de características locales extraídas será superior a 9000
- ☐ D Disponemos de una señal acústica muestreada a 8kHz y 16 bits. ¿Cuánto espacio en memoria requiere la representación en *frames* (marcos) de 2 segundos de señal con una tamaño de ventana (W) de 50ms y un desplazamiento (S) de 5ms? Asume que cada componente de un *frame* se representa mediante un número en coma flotante (4 bytes).
- A) Menos de 150 Kbytes
 - B) Entre 150 Kbytes y 300 Kbytes
 - C) Entre 300 Kbytes y 600 Kbytes
 - D) Más de 600 Kbytes.

A ¿Cuál de las siguientes afirmaciones **no** es cierta sobre la representación de n-gramas?

- A) Dado su grado de dispersión, no suele ser utilizada en la práctica
- B) Captura el contexto en el que ocurre un token
- C) Se suelen utilizar representaciones en memoria dispersas para compensar su dispersión
- D) Su requerimiento de espacio en memoria es exponencial con el orden del n-grama

A Sea un problema de clasificación en dos clases sobre vectores de \mathbb{R}^3 , con las muestras $\{(0, 0, 1), (1, 0, 1), (0, 1, 1)\}$ para la clase 1 y las muestras $\{(1, 1, 1), (2, 0, 1), (0, 1, 2)\}$ para la clase 2. Indicar qué matriz de proyección para \mathbb{R}^2 es la más apropiada para estas muestras.

- A) $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$
- B) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
- C) $\begin{pmatrix} 1 & 0 & 1 \\ 2 & 0 & 0 \end{pmatrix}$
- D) $\begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$

D Se quiere combinar PCA y LDA para la reducción de dimensión de una representación vectorial de \mathbb{R}^D a \mathbb{R}^k ; suponiendo que PCA reduce inicialmente a $\mathbb{R}^{k'}$, ¿qué condición de las siguientes es cierta con respecto a las tres dimensiones suponiendo que tenemos C clases?

- A) $D \geq k'$ y $k' \geq k$ y $k' > C$
- B) $D > k + k'$ y $k' \geq k$ y $C \geq k$
- C) $D \geq k$ y $D \geq k'$ y $D \geq C$
- D) $D \geq k'$ y $k' \geq k$ y $C > k$

B ¿Cuál de las siguientes afirmaciones sobre PCA y LDA es cierta?

- A) Tanto PCA como LDA minimizan el error de clasificación
- B) Ni PCA, ni LDA minimizan el error de clasificación
- C) PCA no minimiza el error de clasificación, pero LDA sí al ser una técnica supervisada
- D) LDA no minimiza el error de clasificación, pero PCA sí

Examen de Teoría de Percepción - Recuperación Primer Parcial

ETSINF, Universitat Politècnica de València, Junio de 2017

Apellidos:

Nombre:

Profesor: ☐ Jorge Civera ☐ Carlos Martínez

Cuestiones (2 puntos, 30 minutos, sin apuntes)

- ☐ D) Dado un clasificador definido por $c(x) = \arg \max_{c=1,\dots,C} g_c(x)$. Indica cuál de las siguientes definiciones de $g_c(x)$ hace que **no** se corresponda a un clasificador de mínimo error:
- A) $g_c(x) = e^{P(x,c)}$
 - B) $g_c(x) = P(x, c)^3$
 - C) $g_c(x) = \sqrt{P(x, c)}$
 - D) $g_c(x) = \cos P(x, c)$
- ☐ C) En Active Learning se dispone de dos conjuntos de muestras, uno de ellos etiquetado T y otro no etiquetado U . ¿Cuál de las siguientes afirmaciones **no** es correcta?
- A) El conjunto de muestras no etiquetado U es mayor que el etiquetado T .
 - B) Se selecciona un subconjunto de U para ser etiquetado.
 - C) La composición del subconjunto de U seleccionado para ser etiquetado es indiferente.
 - D) El error del modelo entrenado con T y el subconjunto de U etiquetado debe ser menor que solamente con T .
- ☐ B) Se está resolviendo un problema de clasificación de imágenes de tamaño 50×50 píxeles. Se propone implementar sistemas usando características globales o características locales sobre ventanas de 10×10 píxeles. Indicar qué afirmación de las siguientes se dará siempre en los sistemas:
- A) El número de vectores de características con la representación global será igual al número de puntos de interés determinado en la imagen
 - B) Las dimensiones de un vector de representación global será 2500 y las de representación local 100, suponiendo que el valor de cada píxel se toma como una característica
 - C) La clasificación por características locales será peor que la clasificación por representación global
 - D) Un vector de características de la representación global ocupará menos memoria que el conjunto de los de la representación local
- ☐ C) Disponemos de una señal acústica muestreada a 44kHz y 16 bits. ¿Cuánto espacio en memoria requiere la representación en frames (marcos) de 1 segundo de señal con un tamaño de ventana (W) de 100ms y un desplazamiento (S) de 50ms? Asume que cada componente de un frame se representa mediante un número en coma flotante (4 bytes).
- A) Menos de 150 Kbytes.
 - B) Entre 150 Kbytes y 300 Kbytes.
 - C) Entre 300 Kbytes y 600 Kbytes.
 - D) Más 600 Kbytes.

D ¿Cuál de las siguientes afirmaciones **no** es cierta sobre la representación de n-gramas?

- A) Es ampliamente utilizada para el modelado de lenguaje.
- B) Principalmente captura dependencias entre tokens consecutivos.
- C) La representación bag-of-words se puede entender como una representación mediante unigramas.
- D) Su requerimiento de espacio en memoria es lineal con el orden del n-grama.

A Sea un problema de clasificación en dos clases sobre vectores de \mathbb{R}^3 , con las muestras $\{(0, 0, 1), (1, 0, 1), (0, 1, 1)\}$ para la clase 1 y las muestras $\{(1, 1, 1), (2, 0, 1), (0, 1, 2)\}$ para la clase 2. Indicar qué matriz de proyección para \mathbb{R}^2 es la más apropiada para estas muestras.

- A) $\begin{pmatrix} 1 & 0 & 2 \\ 1 & 2 & 0 \end{pmatrix}$
- B) $\begin{pmatrix} 1 & 2 & 0 \\ 1 & 1 & 0 \end{pmatrix}$
- C) $\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$
- D) $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$

B Dada la matriz de covarianzas de los datos originales $\Sigma_x \in \mathbb{R}^{D \times D}$, la matriz de proyección PCA $W \in \mathbb{R}^{k \times D}$ donde \mathbf{w}_j es el j -ésimo vector de proyección (de mayor a menor valor propio asociado) y la matriz de covarianzas diagonalizada de los datos originales $\Delta \in \mathbb{R}^{D \times D}$, ¿Cuál de las siguientes expresiones caracteriza la varianza residual (eliminada) al proyectar un conjunto de datos en \mathbb{R}^D a \mathbb{R}^k ?

- A) $\sum_{j=1}^D \mathbf{w}_j \Sigma_x \mathbf{w}_j^t - \sum_{j=1}^D \Delta_{jj}$
- B) $\sum_{j=1}^D \Delta_{jj} - \sum_{j=1}^k \mathbf{w}_j \Sigma_x \mathbf{w}_j^t$
- C) $\sum_{j=1}^k \Delta_{jj} - \sum_{i=1}^n (x_i - \hat{x}_i)^2$ con $x_i \in \mathbb{R}^D$ y $\hat{x}_i \in \mathbb{R}^k$
- D) $\sum_{i=1}^n (x_i - \hat{x}_i)^2 - \sum_{j=1}^k \mathbf{w}_j \Sigma_x \mathbf{w}_j^t$ con $x_i \in \mathbb{R}^D$ y $\hat{x}_i \in \mathbb{R}^k$

A Cuando aplicamos LDA, no tiene sentido proyectar a más de $C - 1$ de dimensiones, ¿a qué se debe esta limitación?

- A) Al rango de la matriz S_b
- B) Al rango de la matriz S_w
- C) Al rango de la matriz de covarianzas por clase Σ_c
- D) Al rango de la matriz de covarianzas de los datos Σ_x

Examen de Teoría de Percepción - Primer Parcial

ETSINF, Universitat Politècnica de València, Marzo de 2018

Apellidos:

Nombre:

Profesor: ☐ Jorge Civera ☐ Carlos Martínez

Cuestiones (2 puntos, 30 minutos, sin apuntes)

☐ A Indica cuál de los siguientes es un clasificador de mínimo error para un objeto x sobre un conjunto de clases \mathbb{C} :

- A) $\hat{c}(x) = \arg \max_{c \in \mathbb{C}} p(x, c)$
- B) $\hat{c}(x) = \arg \min_{c \in \mathbb{C}} p(x, c)$
- C) $\hat{c}(x) = \arg \max_{c \in \mathbb{C}} p(x, c) \cdot P(c)$
- D) $\hat{c}(x) = \arg \max_{c \in \mathbb{C}} \frac{p(x, c)}{P(c)}$

☐ C En el esquema de clasificación con realimentación y reentrenamiento:

- A) Nunca hay modelo inicial de clasificación.
- B) Se emplea sólo la muestra y el resultado del clasificador para cambiar el modelo de clasificación.
- C) Se debe decidir la estrategia de modificación del modelo según la realimentación recibida.
- D) Se emplea sólo la realimentación humana para cambiar el modelo de clasificación.

☐ B Ante una representación por características locales que emplea ventanas de 5×5 píxeles sobre imágenes de 64 niveles de gris, ¿qué afirmación es correcta?

- A) El tamaño en memoria de cada ventana por representación por histograma será de 25 bytes
- B) La memoria total ocupada no puede calcularse sólo con estos datos
- C) La memoria total ocupada usando representación directa por cada ventana será superior a 1000 bytes
- D) El tamaño en memoria de cada ventana por representación directa será de 12 bytes

☐ D ¿Cuántos *frames* F se extraen de un segundo de señal de 16kHz con un desplazamiento de 100 muestras?

- A) $0 < F \leq 50$ frames
- B) $50 < F \leq 100$ frames
- C) $100 < F \leq 150$ frames
- D) $150 < F \leq 200$ frames

- B** En la representación secuencia de tokens (n -gramas), al aumentar el valor de n su requerimiento de memoria crece:
- A) De forma lineal con el tamaño del vocabulario.
 - B) De forma exponencial con el tamaño del vocabulario.
 - C) De forma logarítmica con el tamaño del vocabulario.
 - D) De forma constante porque no depende de la talla del vocabulario.
- C** Sea un problema de clasificación en dos clases sobre vectores de \mathbb{R}^4 , con las muestras $\{(0, 0, 1, 1), (1, 0, 1, 1), (0, 1, 1, 0)\}$ para la clase 1 y las muestras $\{(1, 1, 1, 0), (2, 0, 0, 1), (1, 0, 1, 2)\}$ para la clase 2. Indicar qué matriz de proyección para \mathbb{R}^2 es la más apropiada para estas muestras.
- A) $\begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix}$
 - B) $\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$
 - C) $\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$
 - D) $\begin{pmatrix} 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$
- A** Indicar cuál de las siguientes es una característica de PCA.
- A) Se basa en minimizar el error de reconstrucción.
 - B) Es una técnica supervisada.
 - C) La proyección resultado garantiza el mínimo error de clasificación.
 - D) Construye la matriz de proyección con los vectores propios de la matriz entre clases.
- D** Dado el siguiente conjunto de muestras etiquetadas $\{(-2 \ 0), (-1 \ 0)\}$ de la clase A y $\{(1 \ 0), (2 \ 0)\}$ de la clase B en \mathbb{R}^2 , ¿qué vector de proyección a una única dimensión no esperarías que resultara de aplicar LDA a estas muestras?
- A) $\mathbf{w} = (1 \ 0)$
 - B) $\mathbf{w} = (1 \ 1)$
 - C) $\mathbf{w} = (-1 \ 1)$
 - D) $\mathbf{w} = (0 \ 1)$

Examen de Teoría de Percepción - Recuperación Primer 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 clasificador definido por $c(x) = \arg \max_{c=1,\dots,C} g_c(x)$. Indica cuál de las siguientes definiciones de $g_c(x)$ hace que **no** se corresponda a un clasificador de mínimo error:
- A) $g_c(x) = P(c|x)^2$
 - B) $g_c(x) = P(x|c)$
 - C) $g_c(x) = P(x, c)$
 - D) $g_c(x) = P(c|x)$
- ☐ A En *On-line Learning* se parte de un conjunto de muestras de entrenamiento iniciales T con el cual se entrena un modelo M y mediante la realimentación del usuario se genera un nuevo conjunto de entrenamiento T' que da lugar a un nuevo modelo M' que combina T y T' . Considera la tarea de clasificación de correos electrónicos en *spam* y *ham* donde T es el conjunto de correos TREC06 y T' , un conjunto de correos propios que tú mismo has etiquetado, ¿qué combinación de T y T' pondera igualitariamente todas las muestras para estimar las probabilidades a priori de cada clase en el nuevo modelo M' ?
- A) Sumar el número de correos *spam* y *ham* en T y T' , y normalizar.
 - B) Sumar las probabilidades a priori calculadas a partir de T y las calculadas a partir de T' , y no normalizar
 - C) Sumar las probabilidades a priori calculadas a partir de T y las calculadas a partir de T' , y normalizar
 - D) Sumar el número de correos *spam* y *ham* en T y T' , y no normalizar.
- ☐ C Se tiene un problema de clasificación de imágenes en niveles de gris donde se quiere emplear información global empleando la menor memoria posible. Teniendo en cuenta que se han definido 1024 niveles de gris, ¿qué tamaño mínimo de imagen de los presentados hace que la representación por histograma ocupe menos que la representación directa?
- A) 500 píxeles
 - B) 1000 píxeles
 - C) 1500 píxeles
 - D) 2000 píxeles
- ☐ D Se quiere muestrear una señal de ancho de banda 3500 Hz y se sabe que puede existir ruido de alta frecuencia en el proceso de adquisición. ¿Qué proceso debe seguirse para garantizar una adquisición fiel de la señal?
- A) Basta con muestrear a >7000 Hz
 - B) Aplicar un filtro para que pasen frecuencias ≤ 3500 Hz y muestrear a >3500 Hz
 - C) Aplicar un filtro para que pasen frecuencias en el rango $3500 \pm f_M$ Hz, donde f_M es la frecuencia del ruido
 - D) Aplicar un filtro para que pasen frecuencias ≤ 3500 Hz y muestrear a >7000 Hz
- ☐ C Dado un token que aparece con una frecuencia constante $k > 0$ en todos los documentos de una colección, ¿qué función global le asignaría el menor valor?
- A) La función Normal $G(t) = (\sum_d x_{dt}^2)^{-\frac{1}{2}}$
 - B) La función GfIdf $G(t) = \frac{\sum_d x_{dt}}{\sum_{d: x_{dt} > 0} 1}$
 - C) La función Idf $G(t) = \log \frac{D}{\sum_{d: x_{dt} > 0} 1}$
 - D) Todas le asignan el mismo valor

- B** ¿Cuántos valores propios distintos de cero se obtienen de la matriz $A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ -1 & 0 & -1 & 0 \end{pmatrix}$?
- A) 1
 - B) 2
 - C) 3
 - D) 4

- D** Dada la matriz de covarianzas de los datos originales $\Sigma_{\mathcal{X}} \in \mathbb{R}^{D \times D}$, la matriz de proyección PCA $W \in \mathbb{R}^{D \times k}$ donde \mathbf{w}_j es el j -ésimo vector de proyección (de mayor a menor valor propio asociado) y la matriz de covarianzas diagonalizada de los datos originales $\Delta \in \mathbb{R}^{D \times D}$, ¿cuál de las siguientes expresiones caracteriza el error de reconstrucción de los datos de \mathcal{X} al proyectarlos de \mathbb{R}^D a \mathbb{R}^k con W ?

- A) $\sum_{j=1}^k \mathbf{w}_j^t \Sigma_{\mathcal{X}} \mathbf{w}_j - \sum_{j=1}^k \Lambda_{jj}$
- B) $\sum_{j=1}^k \mathbf{w}_j^t \Lambda_{jj} \mathbf{w}_j - \sum_{j=1}^D \Lambda_{jj}$
- C) $\sum_{j=1}^D \mathbf{w}_j^t \Lambda_{jj} \mathbf{w}_j - \sum_{j=1}^D \Lambda_{jj}$
- D) $\sum_{j=1}^D \mathbf{w}_j^t \Sigma_{\mathcal{X}} \mathbf{w}_j - \sum_{j=1}^k \Lambda_{jj}$

- A** Indica la característica de LDA que la distingue de PCA

- A) Es una técnica de reducción de dimensionalidad supervisada.
- B) Su resolución se basa en un problema de optimización.
- C) Emplea las covarianzas de los datos.
- D) Requiere el cálculo de vectores propios.

Examen de Teoría de Percepción - Primer Parcial

ETSINF, Universitat Politècnica de València, Marzo de 2019

Apellidos:

Nombre:

Profesor: ☐ Jorge Civera ☐ Carlos Martínez

Cuestiones (2 puntos, 30 minutos, sin apuntes)

- ☐ C Dado un problema de clasificación en cuatro clases A , B , C y D , con $P(A) = P(B) = 0.25$, $P(C) = 0.3$ y $P(D) = 0.2$, sobre un espacio binario unidimensional ($x \in \{0, 1\}$) y probabilidades condicionadas $P(0|A) = P(0|D) = 0.6$, $P(0|B) = 0.4$ y $P(0|C) = 0.2$, ¿en qué clase se clasificaría $x = 1$ empleando un clasificador bayesiano?
- A) Clase A
 - B) Clase B
 - C) Clase C
 - D) Clase D
- ☐ B ¿Cuál es la característica fundamental de un sistema de reconocimiento de formas interactivo?
- A) Se realiza aprendizaje incremental de los modelos de clasificación
 - B) Existe una realimentación de usuario
 - C) No es posible su evaluación automática
 - D) Necesita un entrenamiento de modelos convencionales previo a su uso
- ☐ A Tenemos una imagen en escala de grises de 256 niveles de 128×64 píxeles de la cual extraemos ventanas de 13×13 con desplazamiento horizontal de 4 píxeles y vertical de 2 píxeles. ¿Cuál es el tamaño x de su representación directa?
- A) $x < 0.5$ Mbytes
 - B) $0.5 \leq x < 1.0$ Mbytes
 - C) $1.0 \leq x < 1.5$ Mbytes
 - D) $x \geq 1.5$ Mbytes
- ☐ C ¿Qué frecuencia de corte debe aplicar un filtro de paso bajo a una señal que se pretende muestrear correctamente con una frecuencia de 8kHz?
- A) 16KHz
 - B) 8KHz
 - C) 4KHz
 - D) Es indiferente, pues ambas frecuencias no están relacionadas.

D El espacio de almacenamiento que requiere una representación basada en n -gramas de un vocabulario con talla $|V|$ es

- A) $n \cdot |V|$
- B) $n \cdot \log|V|$
- C) $|V|^{\log n}$
- D) $|V|^n$

A ¿Cuál de los siguientes pares de vectores **no** son vectores propios de la matriz $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$?

- A) $x_1 = (1 \quad 0)$ y $x_2 = (0 \quad 1)$
- B) $x_1 = (1 \quad 1)$ y $x_2 = (-1 \quad 1)$
- C) $x_1 = (1 \quad -1)$ y $x_2 = (-1 \quad -1)$
- D) $x_1 = (-2 \quad 2)$ y $x_2 = (-3 \quad -3)$

D ¿Qué caracteriza la matriz de covarianza de los datos proyectados mediante los vectores de proyección PCA?

- A) Es una matriz dispersa
- B) Es una matriz completa
- C) Es la matriz identidad
- D) Es una matriz diagonal

B Es usual aplicar LDA después de PCA, ¿cuál de las siguientes combinaciones tiene sentido para un problema de clasificación D -dimensional en C clases siendo $D \gg C$?

- A) Proyección PCA a $C/2$ seguida de proyección LDA a $D/2$
- B) Proyección PCA a $D/2$ seguida de proyección LDA a $C/2$
- C) Proyección PCA a $C/2$ seguida de proyección LDA a C
- D) Proyección PCA a $D/2$ seguida de proyección LDA a C

Examen de Teoría de Percepción - Recuperación Primer Parcial

ETSINF, Universitat Politècnica de València, Junio de 2019

Apellidos:

Nombre:

Profesor: ☐ Jorge Civera ☐ Carlos Martínez

Cuestiones (2 puntos, 30 minutos, sin apuntes)

☐ C Indica cuál de los siguientes es un clasificador de Bayes para un objeto x sobre un conjunto de clases \mathbb{C} :

- A) $\arg \max_{c \in \mathbb{C}} p(x|c)$
- B) $\arg \max_{c \in \mathbb{C}} P(c|x)^{-1}$
- C) $\arg \min_{c \in \mathbb{C}} -\log P(c|x)$
- D) $\arg \min_{c \in \mathbb{C}} \log P(c)p(x|c)$

☐ A La reformulación de la teoría de la decisión estadística en el caso interactivo se basa en incluir como factor condicionante, aparte del objeto a clasificar x :

- A) La realimentación del usuario f
- B) El modelo obtenido por entrenamiento *off-line* M
- C) La clasificación no interactiva $\hat{c}(x)$
- D) El objeto x en otra modalidad

☐ C Tenemos una extracción por características locales de 11×11 píxeles sobre una imagen de 64×128 píxeles. ¿Cuál es el máximo número de puntos de interés que podría haber?

- A) Menos de 4000
- B) Entre 4000 y 6000
- C) Entre 6001 y 8000
- D) Más 8000

☐ A Los modelos más usados actualmente en reconocimiento de habla continua son:

- A) Los basados en *deep learning*
- B) Los modelos ocultos de Markov (HMM) continuos con mixturas de gaussianas
- C) Los modelos ocultos de Markov (HMM) discretos
- D) Los modelos lineales que emplean segmentación de traza

B Ante una colección de documentos donde hay un token que aparece una sola vez en uno solo de los documentos, ¿qué afirmación es cierta sobre las funciones globales aplicadas sobre ese token?

- A) Las funciones globales normal, GfIdf e Idf valen 1
- B) Las funciones globales normal y GfIdf valen 1, pero no necesariamente Idf
- C) Las funciones globales normal e Idf valen 1, pero no necesariamente GdIdf
- D) Las funciones globales GfIdf e Idf valen 1, pero no necesariamente la normal

D ¿En cuál de las siguientes situaciones tiene menos sentido aplicar reducción de dimensionalidad?

- A) La cantidad de parámetros del modelo puede desbordar la memoria
- B) La dimensionalidad intrínseca es menor a la obtenida en la representación
- C) Existen valores correlados en la representación
- D) El número de muestras por clase es superior a la dimensión

B ¿Qué características presenta la reducción de dimensión por PCA?

- A) Preserva la continuidad, la discriminación y la invarianza
- B) Minimiza el error de reconstrucción
- C) Optimiza la cohesión intraclase
- D) Minimiza el error de clasificación

D ¿Cuál es el rango máximo de la matriz entre-clases S_b para un conjunto de muestras de \mathbb{R}^D pertenecientes a C clases distintas?

- A) D
- B) $D - 1$
- C) C
- D) $C - 1$

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

ETSINF, Universitat Politècnica de València, Junio de 2015

Apellidos:

Nombre:

Profesor: ☐ Carlos Martínez ☐ Roberto Paredes

Cuestiones (3 puntos, 30 minutos, sin apuntes)

- ☐ 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
- ☐ 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) $\max_x g_A(x) = \max_x g_B(x)$
 - B) $g_A(x) = g_B(x)$
 - C) $\min_x g_A(x) = \max_x g_B(x)$
 - D) $g_A(x) > g_B(x)$
- ☐ 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
- ☐ 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
- ☐ 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 $\Sigma_{3 \times 3}$ 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