Tiempo Total: 3 horas

Nota: Por favor, para facilitar la corrección, entregad los ejercicios 1 y 2 por separado del resto.

Ejercicio 1 (2 puntos)

Queremos encontrar un modelo polinómico para predecir el precio de segunda mano de un cierto modelo de coche. Hemos tomado 100 muestras de una página web de compra-venta, con datos del precio y la antigüedad. De ellos, separamos 20 muestras como datos de test. Probamos un modelo polinómico de grado *p*, lo ajustamos mediante regresión regularizada para un cierto valor de λ, y calculamos el error RMSE con los datos de entrenamiento y los datos de test.

Interpreta los resultados obtenidos en 4 posibles escenarios, recogidos en la tabla que sigue. Para cada caso, diagnostica si el aprendizaje ha sido correcto, y en caso de que no lo sea, identifica cuál puede ser el problema y explica en detalle qué posibles soluciones implementarías, por orden de complejidad creciente, indicando sus ventajas e inconvenientes.

Escenario	RMSE entren.	RMSE test
A	1123	1145
B	3423	3512
C	1115	3456
D	2025	1350

Aprendizaje Automático 2-Julio-2013

Ejercicio 2 (3 puntos)

Queremos diseñar un sistema de reconocimiento de objetos con visión 2D, basado en parámetros. Durante la fase de aprendizaje se han tomado 5 imágenes de los dos objetos a reconocer, y se ha calculado en cada una de ellas el área y el perímetro del objeto:

Triángulo

Área	Perímetro
2093	201,1
2188	198,8
2151	201,6
2203	199,9
2112	201,3

Círculo

Área	Perímetro
1879	154,7
1956	157,4
1909	154,7
1910	155,5
1953	156,6

Durante la fase de reconocimiento, tomamos una imagen, y en ella se detectan 4 objetos que queremos clasificar, cuyos parámetros son:

	Área	Perímetro
Objeto1	1922	154,9
Objeto2	2687	194,4
Objeto3	5756	369,8
Objeto4	2161	201,5

Se pide:

a) Elige **razonadamente** un método de clasificación para este problema, discutiendo brevemente las ventajas e inconvenientes que tiene en comparación con otros métodos alternativos.

b) Aplica el método elegido a los cuatro objetos detectados, indicando el resultado obtenido por el clasificador y comentando el resultado.

c) Explica razonadamente qué atributo elegirías si tuvieras que distinguir el triángulo y el círculo utilizando solamente uno de ellos.

Ejercicio 3 **(2 puntos)**

Un fabricante de coches te pide que analices los datos de una encuesta en la que 10,000 personas han respondido a tres preguntas (edad, género y modelo de coche). En un primer paso decides utilizar PCA sobre estos datos de tres dimensiones (cada respuesta es un vector de dimensión tres con el orden anterior).

Interpreta los resultados de dos posibles escenarios:

a) Un único valor propio explica el 99% de la varianza de los datos con vector propio [0 0 1].

b) Un valor propio explica el 60% de la varianza y su vector propio es [0 1 1]' y otro 39% por el vector [1 0 1]'.

Para cada caso, explica qué le podrías decir al fabricante de coches. ¿En qué caso la muestra de la encuesta está bien escogida?

Aprendizaje Automático 2-Julio-2013

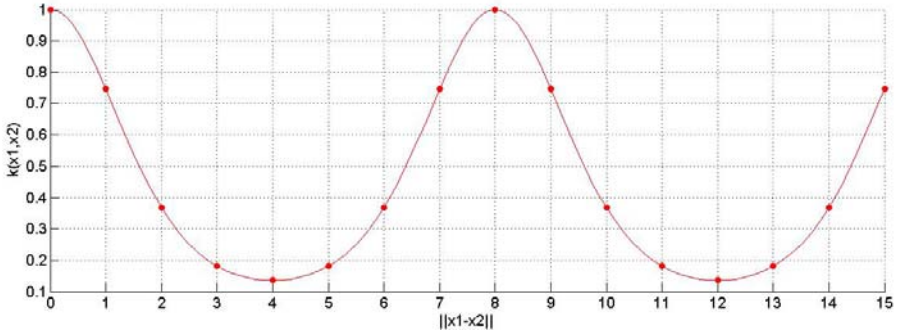
Ejercicio 4 **(3 puntos)**

Considera un problema de regresión de una sola dimensión con datos de entrada *x* y salidas *y*. Dados dos puntos *x1* y *x2* , definimos el siguiente kernel:

$$k(x_1, x_2) = \sigma^2 exp\left(\frac{-2sin^2\left(\frac{\pi||x_1 - x_2||}{p}\right)}{\lambda^2}\right)$$

donde $||x_1 - x_2||$ representa la distancia entre x1 y x2.

El kernel tienes tres hiperparámetros sigma, lambda y p. Utilizando sigma=1, lambda=1 y p=8 la función del kernel para las distancias entre x1 y x2 se muestra en la siguiente figura:



Atención, el eje x representa distancias entre puntos, no valores de los puntos.

Dados tres puntos de entrenamiento (x,y), D= (0 1), (4,-2), (8,1) , queremos predecir el valor en el punto x*=12.

a) Escribe la matriz de covarianza de los puntos de entrenamiento sin tener en cuenta el ruido de observación.
b) Calcula la predicción y* para x* y su varianza
c) Dibuja aproximadamente cual es la estimación para las salidas y para valores de x entre [0 y 20] dados los puntos de entrenamiento anteriores. Para ello marca primero los puntos de entrenamiento en la grafica x-y. Luego, coloca x* junto con y*. Despues dibuja la predición y y su varianza sobre el rango [0,20] de x.
d) ¿Puedes explicar la función de cada hiperparámetro del kernel?

Nota: si no tienes una calculadora científica, puedes utilizar aproximaciones hasta el primer decimal.

Tiempo Total: 3 horas

Nota: Por favor, para facilitar la corrección, entregad los ejercicios 1 y 2 por separado del resto.

Ejercicio 1 **(3 puntos)**

Queremos diseñar un clasificador de fruta, utilizando dos sensores, que dan medidas de elongación y rugosidad (0: redonda, 1: alargada; 0: lisa, 1: rugosa). Las muestras de entrenamiento son:

Fruta	Elongación	Rugosidad
Manzana	0	0,1
Manzana	0,1	0
Melón	1	0,9
Melón	0,9	1
Naranja	0,1	1
Naranja	0	0,9
Pera	0,8	0
Pera	0,9	0,1

Se pide:

a) Plantea el diseño de un clasificador de tipo one-vs-all para clasificar las 4 frutas, que utilice un algoritmo de entrenamiento iterativo, explicando cómo se entrenaría y como se usaría para predecir la clase de una nueva fruta.

b) Haz una iteración del algoritmo de entrenamiento del clasificador one-vs-all para los melones, partiendo de un modelo inicial que clasifique como melón toda fruta con rugosidad mayor que 0,5.

c) Evalúa con los datos de entrenamiento el clasificador obtenido como resultado del apartado anterior.

Aprendizaje Automático 12-Septiembre-2013

Ejercicio 2 **(2 puntos)**

Queremos diseñar un sistema para detectar ofertas que pudieran corresponder a fraudes, o errores en los datos de la oferta, en una web de venta de artículos de segunda mano. En particular nos interesa un cierto modelo de móvil, y encontramos las siguientes ofertas:

Euros
383
459
75
373
399
439
420
429
1025
318
339
359

Se pide:

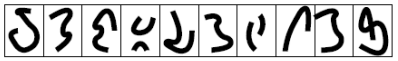
a) Diseña un método para detectar las ofertas extrañas y entrénalo

b) Calcula la salida obtenida por el detector para cada una de las ofertas de la tabla.

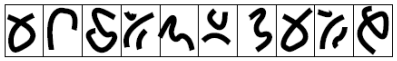
Ejercicio 3 **(2 puntos)**

Acabas de ser contratado en la T.I.A. Tu primera asignación es analizar una carpeta de llena de documentos que contienen nombres de posibles contactos de la organización *A.B.U.E.L.A.* Desafortunadamente, no conoces el alfabeto de los documentos.

Tu primera tarea fue escanear los documentos en imágenes binarias (blanco y negro) y seleccionar una ventana de 80x80 píxeles que rodea cada carácter individual. A continuación tienes un ejemplo :



Name, Last



Name, First

Country of Residence

Tu segunda tarea consiste en etiquetar cada carácter manualmente. Tienes al menos 10.000 ejemplos a etiquetar para poder entrenar un clasificador. Afortunadamente te acuerdas ligeramente de tu paso por la EINA y decides utilizar el algoritmo de clustering K-Medias para determinar el alfabeto de este lenguaje.

- ¿Que representa cada cluster? (0.5 puntos)
- ¿Utilizarías PCA? Justifica tu respuesta. (0.5 puntos)
- ¿Cual es la dimensión de tu problema? (0.5 puntos)
- ¿Cómo seleccionas el número de caracteres del lenguaje? (0.5 puntos)

Aprendizaje Automático 12-Septiembre-2013

Ejercicio 4 **(1.5 puntos)**

En un sistema de recomendación como el descrito en clase,

- ¿Cúal es el número de parámetros para n usuarios y m películas? (0.5 puntos)
- ¿Qué papel juega la regularización? ¿Qué ocurre si no regularizamos? Pon un ejemplo numérico. (1 punto)

Ejercicio 5 **(1.5 puntos)**

Dada la matriz de distancias entre los puntos A, B, C, D

	A	B	C	D
A	0	1	4	5
B		0	2	6
C			0	3
D				0

Utiliza un algoritmo de agrupamiento aglomerativo para agrupar los puntos A, B, C, D. Muestra la solución (dendrogramas) utilizando *complete link* y *average link* para calcular la distancia. Muestra en el dendrograma la distancia entre los clusters que se agrupan. (1,5 puntos)

Tiempo Total: 3 horas

Nota: Por favor, para facilitar la corrección, entregad los ejercicios 1 y 2 por separado del resto.

Ejercicio 1 (2 puntos)

Queremos obtener un modelo de predicción del precio de los discos duros SSD, y hemos programado un robot web que obtiene de forma automática precios de internet. La tabla muestra un ejemplo de unos pocos datos obtenidos en una prueba.

GB	Euros
32	41
64	45
128	84
240	135
480	210
500	643
512	261
1000	507

Se pide:

a) Explica razonadamente qué técnica de aprendizaje es más adecuada para este problema.

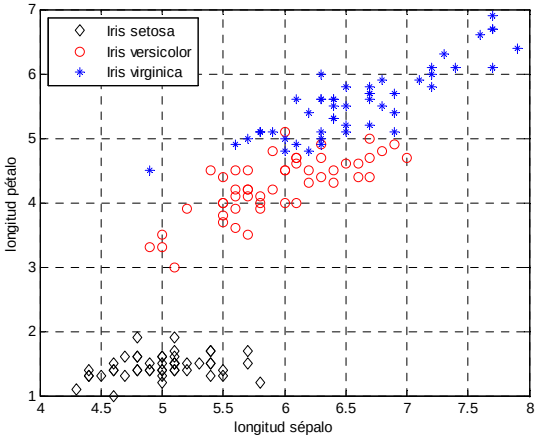
b) Entrena con los datos de la tabla el modelo de predicción escogido. Si utilizas un algoritmo con solución analítica calcula el modelo. Si utilizas uno iterativo, calcula una iteración, partiendo de un modelo inicial de 0,50 Euros por GB.

c) Con el modelo obtenido, calcula el precio predicho para un disco de 512 GB.

Aprendizaje Automático 13-Junio-2014

Ejercicio 2 (3 puntos)

Queremos diseñar un clasificador para tres especies de lirio, a partir de las longitudes del sépalo y del pétalo de la flor. La figura muestra todos los datos de entrenamiento disponibles, y la tabla recoge 5 muestras de cada clase.



Especie	Sépalo	Pétalo
I. setosa	5.1	1.5
I. setosa	5.0	1.2
I. setosa	4.6	1.4
I. setosa	4.4	1.3
I. setosa	5.7	1.5
I. versicolor	6.1	4.0
I. versicolor	5.5	4.0
I. versicolor	5.0	3.5
I. versicolor	6.4	4.3
I. versicolor	5.7	3.5
I. virginica	5.8	5.1
I. virginica	6.5	5.8
I. virginica	7.7	6.1
I. virginica	6.3	5.6
I. virginica	6.0	5.0

Se pide:

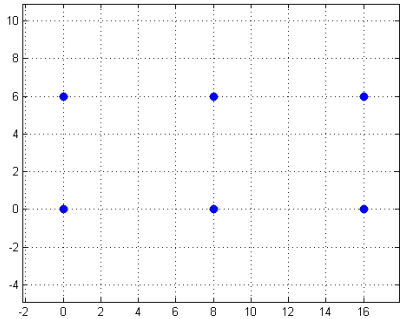
a) Explica si es adecuado diseñar el clasificador utilizando cada una de las siguientes técnicas, y cómo lo plantearías:
   1. Regresión logística
   2. Bayes ingenuo
   3. Bayes completo

b) Escoge el clasificador que te parezca más adecuado y entrénalo con los datos de la tabla. Nota: para abreviar, haz las cuentas solamente para I. virginica, y si utilizas un algoritmo iterativo, haz solamente una iteración.

c) Cuál sería el resultado del clasificador que has entrenado para un ejemplar que tuviera longitudes de sépalo = 5,5 y de pétalo = 4,7. Nota: haz las cuentas solamente para I. virginica, y deja indicado cómo se obtendría el resultado final del clasificador.

Ejercicio 3 **(2.5 puntos)**

Consideramos el algoritmo de K-medias con K=3 sobre un conjunto S de 6 puntos en el plano a=(0,0); b=(8,0); c=(16,0); d=(0,6); e=(8,6); f=(16,6)



El algoritmo usa la distancia Euclídea como métrica para asignar puntos a su centroide más cercano (en caso de empate, se da preferencia al clúster más a la izquierda y abajo por este orden).

Denominamos una configuración de inicio al subconjunto de 3 puntos del conjunto S que inicializan los centroides. Por otro lado, una 3-partición es una partición de S en 3 subconjuntos (e.j. {a,b,e}, {c,d}, {f}). Cualquier 3-partición induce un conjunto de tres centroides. Una 3-partición es estable si la repetición de la iteración del k-medias no cambia los centroides (es decir, el algoritmo ha convergido).

a) ¿Cuantas configuraciones de inicio hay?
b) ¿Cuantas 3-particiones son estables? Dibújalas.
c) ¿Cúal es el número máximo de iteraciones desde una configuración inicial a la 3-partición estable?

Aprendizaje Automático 13-Junio-2014

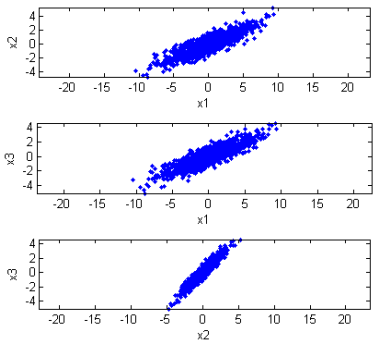
Ejercicio 4 **(1.5 punto)**

Dado un conjunto X de 1000 puntos en $R^{3,}$ su matriz de covarianza empírica es: Y su descomposición en componentes principales ha dado como resultado la base de vectores ortogonales V (vectores propios de la matriz de covarianza) y sus correspondientes valores propios S

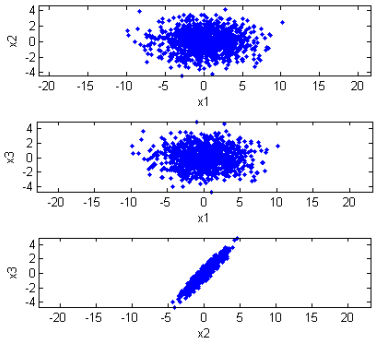
V= [-0.0030 0.0225 -0.9997; y S = [10.3 3.97 0.09];
0.7071 -0.7069 -0.0180;
-0.7071 -0.7070 -0.0137];

1) Calcula la matriz de covarianza.
2) Las siguientes figuras muestran dos distribuciones en 3D. Para cada distribución se ven las marginales por parejas (x1x2, x2x3 y x1x3). ¿Cuál de estas dos distribuciones es la correcta? Razona tu respuesta. ¿Qué estructura tiene la matriz de covarianza de la otra distribución?

Distribución A

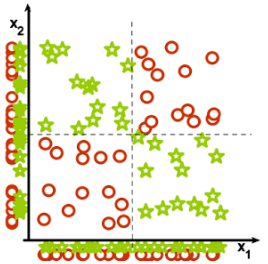


Distribución B



Ejercicio 5 **(1 punto)**

Tenemos la siguiente distribución de características x1 y x2 para un problema de clasificación de dos clases.



Responde de manera razonada a las siguientes preguntas:

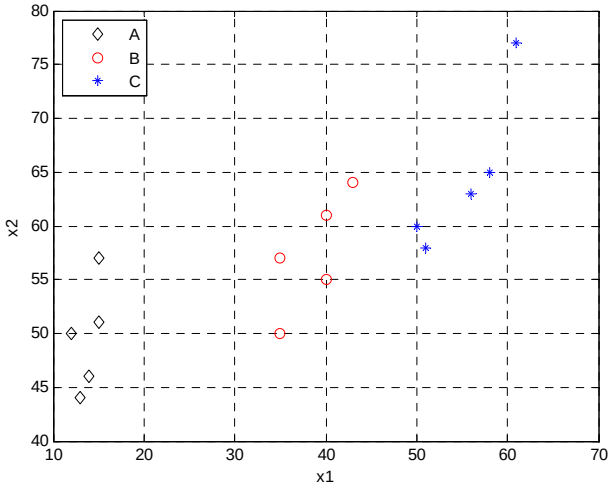
a) ¿Puedo utilizar la norma L1 para seleccionar la mejor de las dos características?
b) ¿Cómo puedo utilizar un clasificador lineal en este problema? Propón un ejemplo concreto.

Tiempo Total: 3 horas

Nota: Por favor, para facilitar la corrección, entregad los ejercicios 1 y 2 por separado del resto.

Ejercicio 1 (2 puntos)

Queremos diseñar un clasificador para 3 clases A, B y C, y disponemos de 5.000 muestras de cada clase, en las que conocemos el valor de dos atributos x1 y x2. En la figura pueden verse como ejemplo 5 muestras de cada clase.



Se pide:

a) Explica en detalle cómo diseñarías un clasificador basado en regresión logística para este problema, indicando todos los pasos que deberá seguir el algoritmo de entrenamiento. Explica también cómo evaluarías las prestaciones del clasificador obtenido.

Aprendizaje Automático 4-Septiembre-2014

Ejercicio 2 (3 puntos)

Hemos creado una web de fotos de gatos, en la que disponemos actualmente de 10.000 fotografías que corresponden, efectivamente, a gatos. Para evitar cachondeo, queremos diseñar un sistema para detectar fotos nuevas subidas por los usuarios, que no correspondan a gatos. Para ello utilizaremos una técnica de bolsa de palabras, que utilizando un diccionario de palabras visuales (pequeños patrones de pixels) obtiene una representación de la imagen con un vector binario que indica si se ha detectado cada una de las palabras del diccionario (w_1 a w_n) en la imagen. El diccionario visual que utilizaremos tiene 24.000 palabras.

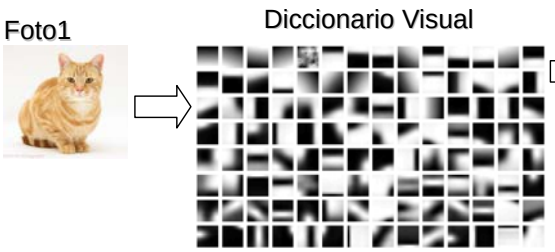


Foto	w_1	w_2	w_3	…
1	1	0	0	…
2	1	0	0	…
3	1	0	0	…
4	0	0	0	…
5	1	1	0	…
6	1	0	0	…
7	1	0	0	…
8	0	0	1	…
9	1	0	1	…
10	1	0	0	…

Se pide:

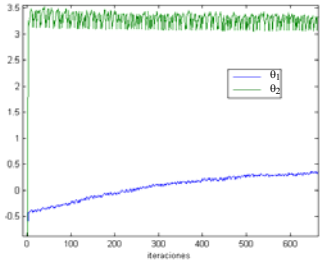
a) Explica razonadamente qué método es más adecuado para resolver este problema, indicando los detalles importantes del mismo.

b) Para demostrar cómo funcionaría, entrena el sistema con las 3 primeras palabras de las 10 fotos de la tabla, e indica cuál sería el resultado que se obtendría para una nueva foto cuya bolsa de palabras fuera W = (1, 0, 1,….)

Ejercicio 3 **(3 puntos)**

Dada un modelo de regresión lineal $y=\theta_1 + \theta_2 x$ y parejas de datos de entrenamiento (x,y): (1,4.5), (3, 11.2), (6,19.8), deseas implementar un algoritmo de gradiente estocástico para estimar θ_1 y θ_2 a partir de los datos de entrenamiento.

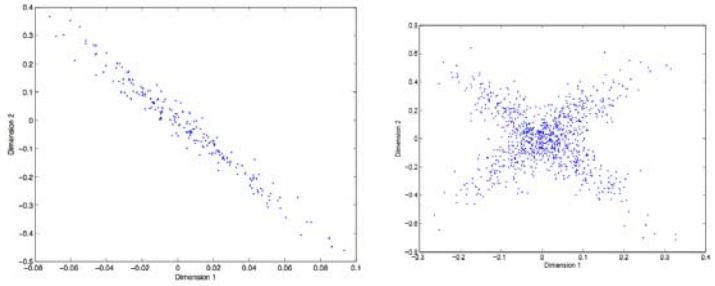
1- Describe en pseudo-código el algoritmo a implementar. Identifica los parámetros que tienes que inicializar para poder correr el algoritmo, explica el papel de cada uno de ellos y justifica los valores que usarías.
2- Calcula las tres primeras actualizaciones de θ_1 y θ_2. Utiliza como inicialización el vector (1,1).
3- ¿Llega este método siempre a la misma solución? ¿Es la misma que para un algoritmo de gradiente en batch? ¿Y que para una solución cerrada?
4- Explica el siguiente gráfico de convergencia. Comenta la calidad de la solución obtenida en la última iteración. ¿Cambiarias alguno de los parámetros del algoritmo tras ver el gráfico? Si la respuesta es sí, ¿qué comportamiento esperarías en ese caso?



Aprendizaje Automático 4-Septiembre-2014

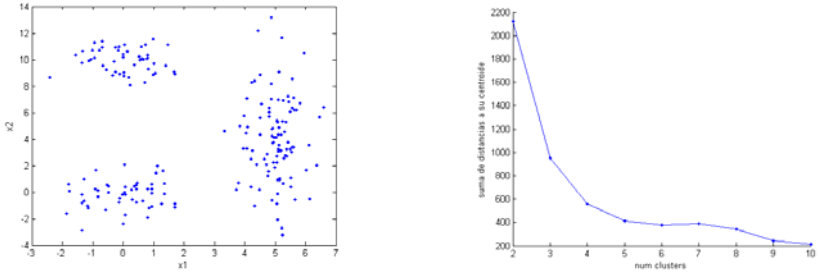
Ejercicio 4 **(1 punto)**

Dados estos dos conjuntos de puntos en 2D, dibuja las dos primeras componentes principales en cada imagen. Entrega esta hoja junto al examen. Justifica tu respuesta en base a la matriz de covarianza de los datos.



Ejercicio 5 **(1 punto)**

El gráfico de la izquierda muestra una nube de puntos en 2D a la que se le ha aplicado el algoritmo Kmedias para K=2:10. La figura de la derecha muestra la suma de los errores de reconstrucción para cada valor de K.



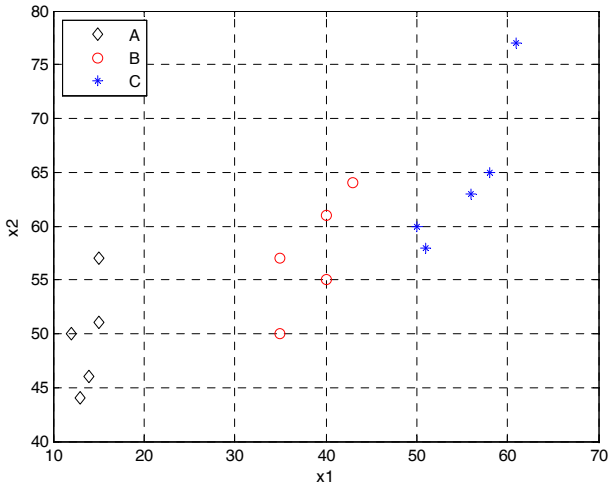
a) ¿Que número de clusters elegirías? Razona tu respuesta
b) ¿Cambiaría tu respuesta si usásemos FMM en lugar de Kmedias? Razona tu respuesta (puedes usar la figura para mostrar la solución de FMM, por ejemplo)

Tiempo Total: 3 horas

Nota: Por favor, para facilitar la corrección, entregad los ejercicios 1 y 2 por separado del resto.

Ejercicio 1 (2 puntos)

Queremos diseñar un clasificador para 3 clases A, B y C, y disponemos de 5.000 muestras de cada clase, en las que conocemos el valor de dos atributos x1 y x2. En la figura pueden verse como ejemplo 5 muestras de cada clase.



Se pide:

a) Explica en detalle cómo diseñarías un clasificador basado en regresión logística para este problema, indicando todos los pasos que deberá seguir el algoritmo de entrenamiento. Explica también cómo evaluarías las prestaciones del clasificador obtenido.

Aprendizaje Automático 4-Septiembre-2015

Ejercicio 2 (3 puntos)

Hemos recogido los siguientes datos sobre los precios de las tabletas más populares del mercado, y dos de sus características principales: el tamaño de la pantalla en pulgadas, y la velocidad del procesador según GeekBench 3 (multi-core).

Pulgadas	GeekBench3	Precio
10,5	2600	400 €
8,4	2600	300 €
9,7	4000	500 €
8,9	3200	400 €
9,7	2600	390 €
9,7	4500	490 €
7,9	2400	390 €
8,9	3200	470 €

a) Explica razonadamente qué técnica de aprendizaje es más adecuada para este problema.

b) Entrena con los datos de la tabla el modelo de predicción escogido. Si utilizas un algoritmo con solución analítica calcula el modelo. Si utilizas uno iterativo, calcula una iteración, partiendo de un modelo inicial de 40 Euros por pulgada.

c) Con el modelo obtenido, calcula el precio predicho para la nueva tableta que saldrá el próximo mes, que tiene pantalla de 8” y GeekBench3 = 4016.

Ejercicio 3 **(2 puntos)**

object	x_1	x_2
1	2	2
2	8	6
3	6	8
4	2	4

a) Agrupa estos datos en dos clusters (K=2) utilizando el algoritmo K-Medias. Inicializa los algoritmos poniendo los objetos 1 y 3 en un cluster y los objetos 2 y 4 en otro. Describe los pasos del algoritmo claramente. Calcula la función de error de reconstrucción de K-Medias en cada paso hasta convergencia.

b) Calcula el error de convergencia para K=4.

c) El algoritmo de K-medias puede interpretarse como un caso especial de agrupamiento basado en modelos Gausianos. ¿Qué restricciones impone K-Medias sobre las Gausianas?

Ejercicio 4 **(1,5 punto)**

Dados tres puntos en dos dimensiones (1, 1), (2, 2) y (3, 3),

Cual el la primera componente principal?

Si proyectamos los puntos originales en el espacio unidimensional definido por la primera componente, ¿cuál es la varianza de los datos proyectados?

Para los datos proyectados en el apartado anterior, ¿cuál es el error de reconstrucción si los volvemos a transformar al espacio original de dos dimensiones?

Ejercicio 5 **(1,5 punto)**

Acabas de implementar para Zara un sistema de recomendación de ropa utilizando el modelo visto en clase con una versión de tu código de dudosa calidad. El sistema se pone en marche el próximo lunes 7 de Septiembre. Amancio Ortega te llama a tu casa el domingo 6 por la noche porque:

a) las recomendaciones que le da el sistema no están entre 0 y 10. ¿Cuál es el problema? ¿Cómo lo solucionas?

b) los productos del nuevo catálogo tienen todos una recomendación de 0!

Si necesitas preguntarle algo a Amancio, hazlo y escribe su respuesta!

Tiempo Total: 3 horas

Nota: Por favor, para facilitar la corrección, entregad los ejercicios 1 y 2 por separado del resto.

Ejercicio 1 **(2.5 puntos)**

Queremos predecir el precio de objetivos para fotografía macro en cámaras reflex, en función de su longitud focal, y de si tienen o no VR (Reducción de Vibraciones). Hemos programado un robot web que extrae automáticamente información de varias tiendas de internet, pero hemos comprobado que siempre que lo ejecutamos aparecen datos erróneos. La tabla muestra un ejemplo de los datos obtenidos:

Focal(mm)	VR	Precio
40	NO	230
60	NO	330
85	SI	550
90	NO	1440
90	SI	620
100	NO	450
105	SI	805
180	SI	1250

Se pide:

a) Explica qué técnica de aprendizaje es más adecuada para este problema, detallando qué pasos serán necesarios en el algoritmo de entrenamiento.

b) Entrena con los datos de la tabla el modelo escogido. Si utilizas un algoritmo con solución analítica, calcula el modelo. Si utilizas uno iterativo, calcula una iteración del algoritmo, empezando con un modelo que corresponda a 6€ por milímetro de focal.

c) Con el modelo obtenido, calcula el precio predicho para un objetivo de 120mm, con VR.

Aprendizaje Automático 21-Junio-2016

Ejercicio 2 **(2.5 puntos)**

Hemos medido la longitud del sépalo y del pétalo de varias flores de la misma especie, y queremos diseñar un sistema que detecte ejemplares que no correspondan a la misma especie. Disponemos de los siguientes datos de entrenamiento:

Sépalo	Pétalo
7.0	4.7
6.9	4.9
5.5	4.0
6.5	4.6
4.9	3.3
6.6	4.6
5.2	3.9

Se pide:

a) Explica razonadamente qué técnica de aprendizaje es más adecuada para este problema, y entrena con los datos de la tabla el modelo escogido.

b) Con el modelo obtenido, calcula la respuesta de nuestro sistema para los siguientes casos:

Sépalo	Pétalo
6.0	3.8
7.0	5.0

Ejercicio 3 **(2.5 puntos)**

Dados los siguientes datos A1=(2,10), A2=(2,5), A3=(8,4), A4=(5,8), A5=(7,5), A6=(6,4), A7=(1,2), A8=(4,9), calcular el diagrama de Venn y el dendograma resultante de ejecutar el algoritmo de clustering jerárquico aglomerativo utilizando como medida de distancia entre clústers la distancia entre los puntos más cercanos (single linkage). La tabla de distancias euclídeas entre los datos es la siguiente:

	A1	A2	A3	A4	A5	A6	A7	A8
A1	0	$\sqrt{25}$	$\sqrt{36}$	$\sqrt{13}$	$\sqrt{50}$	$\sqrt{52}$	$\sqrt{65}$	$\sqrt{5}$
A2		0	$\sqrt{37}$	$\sqrt{18}$	$\sqrt{25}$	$\sqrt{17}$	$\sqrt{10}$	$\sqrt{20}$
A3			0	$\sqrt{25}$	$\sqrt{2}$	$\sqrt{2}$	$\sqrt{53}$	$\sqrt{41}$
A4				0	$\sqrt{13}$	$\sqrt{17}$	$\sqrt{52}$	$\sqrt{2}$
A5					0	$\sqrt{2}$	$\sqrt{45}$	$\sqrt{25}$
A6						0	$\sqrt{29}$	$\sqrt{29}$
A7							0	$\sqrt{58}$
A8								0

Ejercicio 4 **(2.5 puntos)**

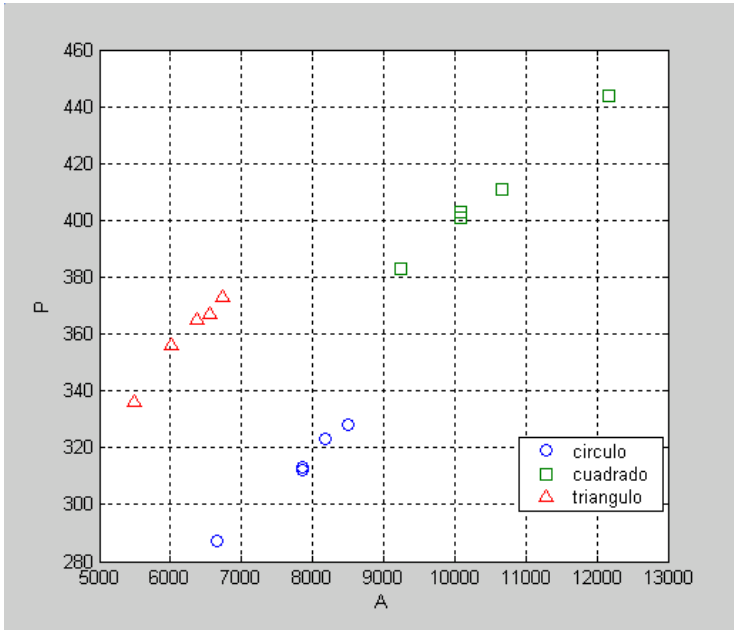
Explica con el mayor detalle que puedas, porqué es o no buena idea utilizar PCA en combinación con el clasificador de caracteres manuscritos MNIST. En todo caso, explica con el mayor detalle posible cómo se haría.

Tiempo Total: 3 horas

Nota: Por favor, para facilitar la corrección, entregad los ejercicios 1 y 2 por separado del resto.

Ejercicio 1 (2 puntos)

Se quiere diseñar un sistema de reconocimiento para tres objetos distintos con un sistema de visión 2D. Durante la fase de aprendizaje se han tomado 5000 imágenes de cada objeto se han calculado a partir de ellas el área y el perímetro. En la figura pueden verse como ejemplo 5 muestras de cada clase.



Se pide:

a) Explica en detalle cómo diseñarías un clasificador basado en regresión logística para este problema, indicando todos los pasos que deberá seguir el algoritmo de entrenamiento. Explica también cómo evaluarías las prestaciones del clasificador obtenido.

Aprendizaje Automático 22-Junio-2017

Ejercicio 2 (3 puntos)

Queremos predecir el precio del próximo modelo de móvil que saldrá al mercado, y hemos programado un robot web que extrae información sobre los modelos actuales a la venta en diversas páginas web. La tabla muestra un ejemplo de los datos obtenido, referidos a tamaño de pantalla, velocidad de procesador (según GeekBench 4 multi-core), y precio.

Pulgadas	GeekBench4	Precio
5,5	5534	909,00 €
4,7	5517	769,00 €
5,5	4054	769,00 €
4,7	3922	659,00 €
5,5	2442	993,00 €
4,7	2390	399,00 €
4	4072	489,00 €

a) Explica razonadamente qué técnica de aprendizaje es más adecuada para este problema.

b) Entrena con los datos de la tabla el modelo de predicción escogido. Si utilizas un algoritmo con solución analítica calcula el modelo. Si utilizas uno iterativo, calcula una iteración, partiendo de un modelo inicial de 150 Euros por pulgada.

c) Con el modelo obtenido, calcula el precio predicho para un nuevo móvil que saldrá el próximo mes, que tiene pantalla de 5” y GeekBench4 = 6000.

Ejercicio 3 (2.5 puntos)

Dada la matriz de cuatro muestras X:

6	**-3**	**-2**	**7**
-4	**5**	**6**	**-3**

Deberás usar PCA para reducir el número de dimensiones. Se pide:

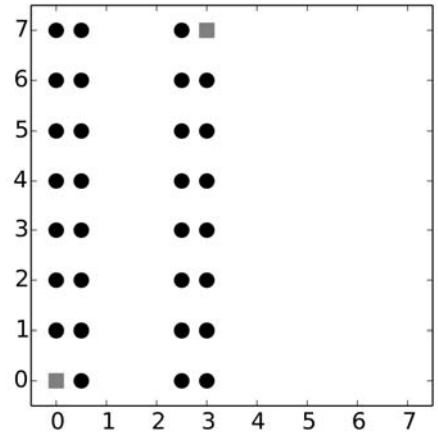
a) Calcular la matriz de covarianza de la muestras **estandarizadas con la media**.
b) Calcular los vectores propios de la matriz de covarianza e indicar cuál corresponderá al valor propio mayor y cuál al menor.
c) Calcular la proyección de las muestras al espacio mono-dimensional correspondiente al **menor** valor propio.

Ayudas: Grafica los datos. Puedes dejar indicadas las fracciones y raíces.

Aprendizaje Automático 22-Junio-2017

Ejercicio 4 **(2.5 puntos)**

Considera los datos de la siguiente figura:



Supón que los rectángulos son las semillas para el algoritmo de k-medias con k=2. Se pide:

a) Calcular la solución final que se obtiene al ejecutar k-medias y dibujarla sobre la figura. Indicar cuántas iteraciones son necesarias para llegar a esta solución.
b) En una figura aparte, dibujar una situación inicial que llevaría a obtener una solución distinta, y dibujar también cuál sería esta solución.
c) ¿Está acotado el número de iteraciones que k-medias ejecutará? Si es así, explica lo mejor que puedas cuál es esa cota.