# TÉCNICAS DE APRENDIZAJE AUTOMÁTICO -- (GRADO-INDAT) --

1. [Página Principal](https://aulas.inf.uva.es/)
2. Mis cursos
3. [TÉCNICAS DE APRENDIZAJE AUTOMÁTICO 21-22](https://aulas.inf.uva.es/course/view.php?id=412)
4. [EXÁMENES DE TEORÍA](https://aulas.inf.uva.es/course/view.php?id=412#section-7)
5. [Examen TEORÍA - SEGUNDA Convocatoria (28 JUNIO 2022)](https://aulas.inf.uva.es/mod/quiz/view.php?id=33773)

| **Comenzado el** | martes, 28 de junio de 2022, 09:13 |
| --- | --- |
| **Estado** | Finalizado |
| **Finalizado en** | martes, 28 de junio de 2022, 09:43 |
| **Tiempo empleado** | 30 minutos 1 segundos |
| **Puntos** | 13,00/30,00 |
| **Calificación** | 4,33 de 10,00 (43%) |

### Pregunta **1**

Incorrecta

Se puntúa -0,20 sobre 1,00

Marcar pregunta

#### Enunciado de la pregunta

Indicar, de los siguentes algoritmos de aprendizaje, cuál se ve más afectado por al presencia de un atributo redundante:

a.

C4.5.

b.

SVM.

c.

NB.

d.

K-NN.

#### Retroalimentación

La respuesta correcta es:

NB.

### Pregunta **2**

Sin contestar

Puntúa como 1,00

Marcar pregunta

#### Enunciado de la pregunta

En una Regresión Lineal Múltiple y en una Regresión Logística:

a.

cada una de ellas requiere siempre un método de optimización específico y, por tanto, diferente

b.

En la primera nunca se puede hacer una optimización analítica, esto es, aplicando el cálculo simbólico sobre expresiones matemáticas. Por el contrario, en la segunda, es imprescindible aplicar dicho método, es decir, que no admite derivación numérica.

c.

normalmente la función de coste, pérdida u objetivo coinciden, aunque la primera opera sobre distancias y, la segunda, sobre probabilidades

d.

en la función de coste, pérdida u objetivo específica de cada regresión, se puede aplicar el mismo método de optimización, por ejemplo, el del Descenso del Gradiente

#### Retroalimentación

La respuesta correcta es:

en la función de coste, pérdida u objetivo específica de cada regresión, se puede aplicar el mismo método de optimización, por ejemplo, el del Descenso del Gradiente

### Pregunta **3**

Sin contestar

Puntúa como 1,00

Desmarcar

#### Enunciado de la pregunta

La función de activación RELU:

a.

su primera derivada es continua en toda la recta real

b.

su derivada requiere de desarrollos en serie

c.

surge para paliar el efecto de la saturación de la función sigmoide

d.

es discontinua

#### Retroalimentación

La respuesta correcta es:

surge para paliar el efecto de la saturación de la función sigmoide

### Pregunta **4**

Correcta

Se puntúa 1,00 sobre 1,00

Marcar pregunta

#### Enunciado de la pregunta

En el aprendizaje inductivo, el sobreajuste se produce cuando:

a.

La tasa de error de la hipótesis sobre el conjunto de prueba es menor que el error de resubstitución.

b.

La hipótesis no clasifica bien las instancias del conjunto de prueba.

c.

Hipótesis con mayor error de resubstitución tienen tasas de error menores en conjuntos de prueba.

d.

La tasa de error de la hipótesis sobre el conjunto de prueba es mayor que el error de resubstitución.

#### Retroalimentación

La respuesta correcta es: Hipótesis con mayor error de resubstitución tienen tasas de error menores en conjuntos de prueba.

### Pregunta **5**

Sin contestar

Puntúa como 1,00

Marcar pregunta

#### Enunciado de la pregunta

En el aprendizaje inductivo, los bías inductivos:

a.

Solo son necesarios si el conjunto de entrenamiento no se puede describir mediante una función.

b.

Tienen menor influencia si el conjunto de entrenamiento se puede describir mediante una función.

c.

Son necesarios para compensar el sesgo en la selección del conjunto de entrenamiento.

d.

Son necesarios para clasificar racionalmente instancias no vistas en el entrenamiento.

#### Retroalimentación

La respuesta correcta es:

Son necesarios para clasificar racionalmente instancias no vistas en el entrenamiento.

### Pregunta **6**

Sin contestar

Puntúa como 1,00

Marcar pregunta

#### Enunciado de la pregunta

Las operaciones OneHot y OneCold en los sistemas lineales de Aprendizaje Automático:

a.

ambas producen un vector de la misma dimensión, que el usado como entrada a dichas operaciones

b.

la primera se aplica sobre los atributos y, la segunda, sólo sobre las salidas

c.

una es la inversa de la otra

d.

si el resultado, en ambas operaciones, fuese un vector de igual dimensión, tendría la misma cantidad de ceros y de unos, tanto en un caso, como en el otro

#### Retroalimentación

La respuesta correcta es:

si el resultado, en ambas operaciones, fuese un vector de igual dimensión, tendría la misma cantidad de ceros y de unos, tanto en un caso, como en el otro

### Pregunta **7**

Correcta

Se puntúa 1,00 sobre 1,00

Marcar pregunta

#### Enunciado de la pregunta

El bucle interno del algoritmo básico de recubrimiento secuencial para la inducción de reglas, PRISM, tiene por misión:

a.

Separar los ejemplos, eliminando los ya cubiertos por las reglas.

b.

Encontrar la regla que cubra mas ejemplos de la clase seleccionada.

c.

Relizar una búsqueda de lo específico a lo general en el espacio de reglas, hasta encontrar una regla que solo cubra ejemplos de la clase seleccionada.

d.

Relizar una búsqueda de lo general a lo específico en el espacio de reglas, hasta encontrar una regla que solo cubra ejemplos de la clase seleccionada.

#### Retroalimentación

La respuesta correcta es:

Relizar una búsqueda de lo general a lo específico en el espacio de reglas, hasta encontrar una regla que solo cubra ejemplos de la clase seleccionada.

### Pregunta **8**

Correcta

Se puntúa 1,00 sobre 1,00

Desmarcar

#### Enunciado de la pregunta

En una clasificación binaria, ¿por qué no se ponen dos neuronas en la capa de salida de un MLP, sino que se usa una sola?

a.

Porque el logaritmo en base dos de dos es uno precisamente

b.

Porque poner dos, podría desembocar en clasificaciones ambiguas.

c.

Porque con la función de activación arctan(x), la salida de la neurona está en la codificación OneHot.

d.

Es debido a que empleando una función de activación sigmoide, directamente la salida es interpretable como una probabilidad de pertenencia a una clase (o la complementaria).

#### Retroalimentación

La respuesta correcta es:

Es debido a que empleando una función de activación sigmoide, directamente la salida es interpretable como una probabilidad de pertenencia a una clase (o la complementaria).

### Pregunta **9**

Sin contestar

Puntúa como 1,00

Marcar pregunta

#### Enunciado de la pregunta

En una RBF, ¿por qué no se sustituye la capa de salida por un Perceptrón Simple y sí por un Adaline?

a.

Porque el primero implicaría una cantidad de neuronas potencia de 2 del número de clases, mientras que el Adaline podría, incluso, ajustarse con una sola neurona.

b.

Porque si las salidas de la capa oculta no fueran linealmente separables, no convergería nunca el aprendizaje de la capa de salida, mientras que con el Adaline, el algoritmo se puede parar a voluntad, aunque no se alcance el 100% de aciertos.

c.

Porque el Adaline se puede ajustar muy rápidamente por el método de Mínimos Cuadrados, lo que garantiza una tasa de aciertos muy elevada.

d.

Porque, aunque el Adaline es muy simple, puede dar pie a ser sustituido por otros sistemas lineales que ofrezcan mejores resultados.

#### Retroalimentación

La respuesta correcta es:

Porque si las salidas de la capa oculta no fueran linealmente separables, no convergería nunca el aprendizaje de la capa de salida, mientras que con el Adaline, el algoritmo se puede parar a voluntad, aunque no se alcance el 100% de aciertos.

### Pregunta **10**

Sin contestar

Puntúa como 1,00

Desmarcar

#### Enunciado de la pregunta

Si todas las muestras de una misma clase pudieran estar dentro de una envolvente convexa, ¿se podría aplicar la solución de margen máximo?

a.

No, porque siempre estaría el problema de los puntos aislados, que nos sabríamos dónde englobarlos.

b.

Sí, en cualquier caso.

c.

Sí, siempre que los espacios acotados por estas envolventes convexas fuesen disjuntos.

d.

Depende de si la clasificación es binaria o no. En general, debería ser no binaria para poder aplicar el algoritmo del SVM.

#### Retroalimentación

La respuesta correcta es:

Sí, siempre que los espacios acotados por estas envolventes convexas fuesen disjuntos.

### Pregunta **11**

Incorrecta

Se puntúa -0,20 sobre 1,00

Desmarcar

#### Enunciado de la pregunta

En el aprendizaje automático se denomina conocimiento base a:

a.

el conocimiento que tiene inicialmente el método de aprendizaje sobre el problema a resolver.

b.

el criterio con el que se inicializa el algoritmo.

c.

la base de conocimiento que contiene los ejemplos.

d.

la base de conocimiento de la que aprende el algoritmo.

#### Retroalimentación

La respuesta correcta es: el conocimiento que tiene inicialmente el método de aprendizaje sobre el problema a resolver.

### Pregunta **12**

Sin contestar

Puntúa como 1,00

Marcar pregunta

#### Enunciado de la pregunta

En el modelo de neurona artificial McCulot Pitts, las sinapsis excitadoras e inhibidoras:

a.

sólo depende del valor absoluto del peso: cuanto mayor, más excitadora será la sinapsis y, al contrario, cuanto más pequeña, más inhibidora

b.

no depende de un peso concreto, sino del conjunto de ellos que posee la neurona, haciéndose una media, de tal manera que, si es positiva, s edice que todas las sinapsis son excitadoras. Y, al revés, si es negativa: son todas ellas inhibidoras

c.

depende de si en el proceso de aprendizaje, al aplicar la Regla de Hebb, el peso en cuestión tiende a cero (inhibidora) o se aleja de este valor (excitadora)

d.

se materializan a través del signo del peso asociado: positivo en el primer caso y negativo en el segundo

#### Retroalimentación

La respuesta correcta es:

se materializan a través del signo del peso asociado: positivo en el primer caso y negativo en el segundo

### Pregunta **13**

Incorrecta

Se puntúa -0,20 sobre 1,00

Desmarcar

#### Enunciado de la pregunta

Considerar una red bayesiana "lineal": C --> X --> Y --> Z, donde C es la clase y X, Y, Z los atributos.

En esta red:

a.

Los atributos son independientes dada la clase.

b.

Los atributos son independientes dada la clase y algún otro atributo.

c.

Z es independiente de la clase dado Y.

d.

No hay relaciones de independencia entre los atributos.

#### Retroalimentación

La respuesta correcta es:

Z es independiente de la clase dado Y.

### Pregunta **14**

Correcta

Se puntúa 1,00 sobre 1,00

Desmarcar

#### Enunciado de la pregunta

¿Cuál es el objetivo principal de podar un árbol?

a.

Disminuir el sobreajuste.

b.

Aumentar el error de resubstitución.

c.

Equilibrar las ramas.

d.

Disminuir el error en un conjunto de validación.

#### Retroalimentación

La respuesta correcta es:

Disminuir el sobreajuste.

### Pregunta **15**

Sin contestar

Puntúa como 1,00

Marcar pregunta

#### Enunciado de la pregunta

Observando la evolución de la tasa de aciertos conforme aumenta las iteraciones en un MLP, ¿cómo se detectaría una situación de infra-ajuste?

a.

Por un decrecimiento continuo del error máximo por muestra

b.

Por un comportamiento oscilatorio cuando se llevan ejecutadas bastantes épocas

c.

Por un aproximación asintótica a la tasa de aciertos final

d.

Por un rápido acercamiento a la tasa de aciertos final

#### Retroalimentación

La respuesta correcta es:

Por un comportamiento oscilatorio cuando se llevan ejecutadas bastantes épocas

### Pregunta **16**

Correcta

Se puntúa 1,00 sobre 1,00

Marcar pregunta

#### Enunciado de la pregunta

En un experimento de *hold-out* repetido K veces, podemos asumir que la tasa de error sigue una distribución normal cuando:

a.

K=20, porque la estimación está menos sesgada que la validación cruzada de 10 particiones.

b.

Si cada conjunto de pueba tiene al menos 30 elementos y la tasa de error no se acerca a 0 ni a 1.

c.

Siempre.

d.

Nunca.

#### Retroalimentación

La respuesta correcta es:

Nunca.

### Pregunta **17**

Correcta

Se puntúa 1,00 sobre 1,00

Desmarcar

#### Enunciado de la pregunta

Cuando las muestras no son linealmente separables, ¿cómo opera el SVM?

a.

Se trabaja en un nuevo espacio, donde la dimensionalidad se ha reducido al máximo posible, quedando asegurado así su separación lineal.

b.

Llevándolas a un espacio donde sí lo sean, lo que implica transformarlas a un espacio, normalmente, de dimensión mucho menor que el de partida.

c.

Muy parecido al Perceptrón Simple, sólo que las muestras deberán estar escaladas entre [-1,+1].

d.

Todo el problema se traslada a un espacio donde sí sean linealmente separables, aunque no se calculan sus nuevas coordenadas, porque sólo importa su relación en términos de distancias con otras muestras.

#### Retroalimentación

La respuesta correcta es:

Todo el problema se traslada a un espacio donde sí sean linealmente separables, aunque no se calculan sus nuevas coordenadas, porque sólo importa su relación en términos de distancias con otras muestras.

### Pregunta **18**

Correcta

Se puntúa 1,00 sobre 1,00

Desmarcar

#### Enunciado de la pregunta

En un problema de clasificación binaria, las instancias están descritas por cuatro atributos binarios. Se dispone de un conjunto de datos que contiene 32 instancias, una para cada posible asignación de valores de atributos y clase. ¿Cuál es el valor del parámetro P(x1/v1) de un clasificador Naive Bayes si utilizamos la estimación de máxima verosimilitud?

a.

8/16

b.

8/32

c.

1/16

d.

2/16

e.

1/32

#### Retroalimentación

La respuesta correcta es:

8/16

### Pregunta **19**

Sin contestar

Puntúa como 1,00

Desmarcar

#### Enunciado de la pregunta

¿En el aprendizaje del Adaline se puede incorporar un Factor Momento al del Gradiente?

a.

Perfectamente, lo que ayudará a que converja más rápidamente.

b.

Sí, siempre que la función objetivo de la optimización no incluya funciones logarítmicas o exponenciales.

c.

No, porque la adición de otra función proporcional a los pesos produce valores de dichos pesos en continuo crecimiento, lo que llevaría, en pocas iteraciones, a pesos no representables con un número finito de bytes (desbordamiento).

d.

Normalmente no, porque su cálculo conlleva un aumento de la complejidad computacional elevando el problema a los de complejidad exponencial.

#### Retroalimentación

La respuesta correcta es:

No, porque la adición de otra función proporcional a los pesos produce valores de dichos pesos en continuo crecimiento, lo que llevaría, en pocas iteraciones, a pesos no representables con un número finito de bytes (desbordamiento).

### Pregunta **20**

Sin contestar

Puntúa como 1,00

Marcar pregunta

#### Enunciado de la pregunta

¿Cuántos parámetros son necesarios para parametrizar una red TAN que tiene 3 variables aleatorias discretas con dominio en tres valores?

a.

27

b.

39

c.

9

d.

21

#### Retroalimentación

La respuesta correcta es:

39

### Pregunta **21**

Correcta

Se puntúa 1,00 sobre 1,00

Marcar pregunta

#### Enunciado de la pregunta

En al aprendizaje no supevisado (o formación de conceptos):

a.

Se obtiene una caracterización del concepto (como en el aprendizaje supervisado)

b.

Siempre se conoce el concepto objetivo.

c.

No hay un concepto objetivo.

d.

No hay ejemplos.

#### Retroalimentación

La respuesta correcta es:

Se obtiene una caracterización del concepto (como en el aprendizaje supervisado)

### Pregunta **22**

Correcta

Se puntúa 1,00 sobre 1,00

Marcar pregunta

#### Enunciado de la pregunta

El Descenso del Gradiente frente al de Mínimos Cuadrados:

a.

es mucho más rápido

b.

funciona mucho peor ante la presencia de puntos aislados

c.

converge siempre por debajo de una cota de error dada

d.

no requiere conocer las expresiones analíticas de la función por optimizar, sino su valor numérico punto a punto

#### Retroalimentación

La respuesta correcta es:

no requiere conocer las expresiones analíticas de la función por optimizar, sino su valor numérico punto a punto

### Pregunta **23**

Correcta

Se puntúa 1,00 sobre 1,00

Desmarcar

#### Enunciado de la pregunta

La normalización o escalado de las muestras en el SVM:

a.

es critico, porque puede dar lugar a una distorsión del espacio, lo que afecta de lleno al cálculo de la métrica (distancia)

b.

Estas operaciones son necesarias, pero siempre que se hagan de manera global, no separadamente por atributo, como suele ser lo normal en otros algoritmos de Aprendizaje Automático.

c.

depende de los casos. Sólo en el supuesto de que alguno de los atributos, que por su naturaleza sea categórico, haya sido transformado a numérico, entonces, sí es imprescindible estas operaciones previas.

d.

es una circunstancia que simplifica los cálculos, pero se puede aplicar el algoritmo de aprendizaje igualmente sin variar, con ello, el resultado final

#### Retroalimentación

La respuesta correcta es:

es critico, porque puede dar lugar a una distorsión del espacio, lo que afecta de lleno al cálculo de la métrica (distancia)

### Pregunta **24**

Correcta

Se puntúa 1,00 sobre 1,00

Marcar pregunta

#### Enunciado de la pregunta

Indicar el principal inconveniente de la validación cruzada repetida como método para estimar la varianza de la tasa de error:

a.

Ninguno, pues los conjunto de entrenamiento no se solapan.

b.

Estimación pesimista de la varianza, por el exceso de repetición.

c.

Estima varianzas distintas que la validación cruzada sin repetición.

d.

Estimación optimista de la varianza, por el exceso de repetición.

#### Retroalimentación

La respuesta correcta es: Estimación optimista de la varianza, por el exceso de repetición.

### Pregunta **25**

Incorrecta

Se puntúa -0,20 sobre 1,00

Marcar pregunta

#### Enunciado de la pregunta

La tasa de error de un clasificador es:

a.

La probabilidad de cometer un error al clasificar una instancia elegida de forma aleatoria.

b.

El error que comete el clasificador sobre cualquier conjunto de prueba independiente del de entrenamiento.

c.

La probabilidad de cometer un error al clasificar una instancia no utilizada para entrenar el clasificador.

d.

El error que comete el clasificador sobre un conjunto de prueba independiente del de entrenamiento.

#### Retroalimentación

La respuesta correcta es: El error que comete el clasificador sobre un conjunto de prueba independiente del de entrenamiento.

### Pregunta **26**

Correcta

Se puntúa 1,00 sobre 1,00

Desmarcar

#### Enunciado de la pregunta

¿Es posible que en dos experimentos con el mismo conjunto de datos, igual norma (distancia) e idéntico número de clusters, la salida de la capa oculta varíe?

a.

Sí, porque el algoritmo de las k-medias es muy sensible a los baricentros iniciales

b.

No, el algoritmo de las k-medias siempre converge a la misma solución.

c.

No porque, aunque haya un efecto aleatorio de los baricentros iniciales, se corrige conforme crece el número de iteraciones.

d.

Sí, dado que la presencia de puntos aislados distorsiona su asignación a un determinado baricentro.

#### Retroalimentación

La respuesta correcta es:

Sí, porque el algoritmo de las k-medias es muy sensible a los baricentros iniciales

### Pregunta **27**

Correcta

Se puntúa 1,00 sobre 1,00

Marcar pregunta

#### Enunciado de la pregunta

¿Cómo se puede buscar un compromiso razonable entre precisión y robustez al ruido en el algoritmo K-vecinos más próximos?

a.

Utilizando una métrica adecuada para el problema.

b.

Seleccionando K en el entrenamiento hasta que el error de resubstitución sea nulo.

c.

Seleccionando K en el entrenamiento hasta que el error de resubstitución no aumente, para evitar el sobreajuste.

d.

Seleccionando K mediante un proceso de validación cruzada sobre el conjunto de entrenamiento.

#### Retroalimentación

La respuesta correcta es:

Seleccionando K mediante un proceso de validación cruzada sobre el conjunto de entrenamiento.

### Pregunta **28**

Sin contestar

Puntúa como 1,00

Marcar pregunta

#### Enunciado de la pregunta

El método del Descenso del Gradiente en una Regresión Lineal:

a.

no admite como función de coste, pérdida u objetivo el error cuadrático medio en cualquiera de sus vertientes

b.

todas las muestras deben ocupar el mismo orden en cada época

c.

la función de coste, pérdida u objetivo coincide con la usada en el Método de Mínimos cuadrados, aunque normalmente sobre un conjunto más reducido de muestras

d.

en el momento que detecta un número de muestras inferior al de pesos (incluido el bias), aborta su ejecución

#### Retroalimentación

La respuesta correcta es:

la función de coste, pérdida u objetivo coincide con la usada en el Método de Mínimos cuadrados, aunque normalmente sobre un conjunto más reducido de muestras

### Pregunta **29**

Incorrecta

Se puntúa -0,20 sobre 1,00

Desmarcar

#### Enunciado de la pregunta

En los árboles de decisión:

a.

Todos los caminos usan todos los atributos.

b.

No se selección atributos irrelevantes.

c.

Puede haber menos nodos terminales que valores de la clase del concepto.

d.

Un ejemplo con atributos conocidos puede llegar a dos nodos hoja.

#### Retroalimentación

La respuesta correcta es:

Puede haber menos nodos terminales que valores de la clase del concepto.

### Pregunta **30**

Correcta

Se puntúa 1,00 sobre 1,00

Desmarcar

#### Enunciado de la pregunta

El aprendizaje del perceptrón simple:

a.

siempre converge

b.

se basa en la aplicación literal de la Regla de Hebb

c.

siempre que modifica los pesos en una determinada iteración, lo hace recurriendo a una adaptación de la Regla de Hebb, donde se sustituye la salida de la neurona, por la respuesta deseada

d.

se hace imprescindible que sea por épocas enteras

#### Retroalimentación

La respuesta correcta es:

siempre que modifica los pesos en una determinada iteración, lo hace recurriendo a una adaptación de la Regla de Hebb, donde se sustituye la salida de la neurona, por la respuesta deseada