

En el modelo de McCulloch y Pitts de neurona artificial, el offset:

- a. es conocido de antemano.
- b. se puede tratar como un peso más, como si controlara una entrada cuya magnitud es fija y, en general, diferente para cada problema particular.
- c. no es conocido de antemano, pero se ajusta una vez lo han hecho los pesos.
- D. se puede tratar como un peso más, como si controlara una entrada constante e igual a uno.**

Sea A un árbol de decisión inducido con ID3 y B el árbol podado obtenido a partir de A según el método de C4.5. B tiene menos nodos que A.

- a. A y B tienen el mismo error de resubstitución.
- b. En todos los conjuntos de prueba la tasa de error de B es menor que la de A.
- C. La tasa de error de B sobre el conjunto de entrenamiento siempre es mayor que la de A.**
- d. El error verdadero de B siempre es menor que el de A.

El error verdadero 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 un conjunto de prueba independiente.
- 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 cualquier conjunto de prueba.

¿A qué se denomina hipótesis HMAP?

- A. A la hipótesis que tiene la máxima probabilidad a posteriori.**
- b. A la hipótesis que maximiza la probabilidad de que una nueva instancia se clasifique correctamente.
- c. A la hipótesis que tiene la máxima probabilidad a priori.
- d. A la hipótesis que se obtiene aplicando el teorema de Bayes.

Cuando aparece un “drop-out” del 50% en una determinada capa, significa que:

- a. sólo calculará la salida de la mitad de las neuronas escogidas éstas al azar.
- B. sólo se actualizarán los pesos de la mitad de las neuronas escogidas aleatoriamente dentro de las que integran la capa.**
- c. sólo se actualizarán los pesos de la primera o segunda mitad de las neuronas, escogida dicha mitad al azar.
- d. se calculará la salida de todas las neuronas, pero sólo se transmitirán el 50% de ellas escogidas al azar.

Los algoritmos de inducción de árboles de decisión consideran las distintas particiones que cada atributo realiza sobre los datos disponibles. Las heurísticas para la selección de atributos buscan maximizar:

A. La desigualdad de la distribución de clases en la partición de los datos que induce el atributo.

b. La igualdad de la distribución de clases en la partición de los datos que induce el atributo.

c. La igualdad del tamaño de los subconjuntos en la partición de los datos que induce el atributo.

d. La desigualdad del tamaño de los subconjuntos en la partición de los datos que induce el atributo.

Las funciones núcleo:

a. no garantizan que en el espacio dual, se pueda establecer una distancia

b. complican el cálculo de las componentes dentro de una base ortonormal en el espacio transformado

C. nos dan la distancia entre dos puntos en el nuevo espacio, sin tener que calcular su transformación.

d. obligan a hallar cada muestra en el nuevo espacio

Seleccione la frase que completa la siguiente definición:

“Un programa de ordenador APRENDE ... con respecto a una clase de tareas T y medida de desempeño P si su rendimiento en tareas de T, según la medida P, mejora con la experiencia E (Mitchell, 97)”

a. “de los datos históricos H”

b. “de su comportamiento anterior B”

C. “de la experiencia E”

d. “de la información disponible I”

Disponemos de bolas blancas, negras y rojas y de cubos blancos y negros. Una urna contiene dos bolas negras, tres bolas blancas, ninguna roja, dos cubos blancos y dos negros. ¿Cuál es la estimación de Laplace de la probabilidad de obtener una bola roja de la urna seleccionando aleatoriamente uno de los objetos de la urna?

a. $1/12$

B. $1/8$

c. 0

d. $1/3$

Una red bayesiana tiene tres nodos etiquetados con las variables aleatorias discretas X, Y, Z, que toman 3, 2 y 3 valores, respectivamente. X e Y no tienen padres. Z tiene por padres a X e Y. Indicar cuantos parámetros independientes contiene al modelo paramétrico de la red.

- a. 18
- b. 23
- c. 8
- D. 15**

Un aprendizaje por etapas de una red como la RBF significa que:

- a. las capas pueden entrenarse por separado sin orden preestablecido
- B. primero se ha de entrenar la capa oculta y, con sus salidas ya ajustadas, entrenar la capa de salida**
- c. se ha de dividir las muestras de entrenamiento y secuenciarlas en el tiempo
- d. se ha de pasar cada muestra primero por la capa de entrada y, lo que dé, por la capa de salida. Y así, muestra a muestra hasta acabar con toda una época

Indicar la ventaja principal de la validación cruzada frente al Hold-out repetido como método para estimar la tasa de error:

- A. Los conjuntos de prueba no se solapan.**
- b. Los conjuntos de entrenamiento no se solapan.
- c. Menor variabilidad de la partición aleatoria.
- d. Menor coste computacional.

¿En qué consiste la suposición Naive Bayes?

- a. En suponer que los atributos son independientes.
- b. En suponer que los atributos son dependientes de la clase.
- c. En suponer que los atributos son independientes de la clase.
- D. En suponer que los atributos son independientes dada la clase.**

El uso de la función softmax implicará minimizar una función de coste:

- a. no introduce cambio alguno respecto a la clasificación binaria.
- b. igual que en anteriores ocasiones, basada literalmente en el error cuadrático
- C. basada en un función de máxima verosimilitud, ya que la salida de esta activación es interpretable como un probabilidad de pertenencia a la clase asociada a la neurona.**
- d. Simplifica los cálculos enormemente al meter explícitamente la función sigmoide

¿Qué elemento tienen en común al aprendizaje supervisado (descripción de conceptos) y el no supervisado (formación de conceptos)?

- a. Utilizan el mismo conjunto de prueba.
- b. Ambos descubren conceptos nuevos.
- C. Su objetivo es caracterizar el concepto.**
- d. La descripción de las instancias de entrenamiento.

Tanto en un perceptrón simple, como en un adaline, es preciso:

- a. procesar todas las entradas antes de efectuar ninguna actualización de pesos.
- B. se puede entrenar cada neurona de salida independientemente, ya que no hay conexión con el resto de dicha capa.**
- c. entrenar cada nodo de la capa de salida conjuntamente con los demás
- d. que la secuencia de entradas se repita en cada época.

Seleccione, entre las siguientes opciones, el número mínimo de instancias para poder estimar la varianza de un clasificador mediante la distribución normal, sabiendo que la tasa de error del clasificador sobre el conjunto de prueba es del 1%

- A. 600 instancias**
- b. 60 instancias
- c. 6000 instancias
- d. 30 instancias

Prism es el algoritmo básico de inducción de reglas proposicionales de clasificación.

Selecciona reglas que:

- a. Cubren más instancias positivas que negativas.
- b. Cubren el mínimo número posible de instancias negativas.
- C. Solo cubren instancias positivas.**
- d. Cubren el mayor número posible de instancias positivas.

En un problema de clasificación binaria, si existe una solución de separación lineal:

- a. Puede darse el caso de que dicha solución sea la única existente.
- B. No es única y, en general, existe una infinidad de ellas.**
- c. No se puede decir nada acerca de la cantidad de soluciones que puedan darse.
- d. En general, no es única, pero no se puede decir que haya infinitas.

El algoritmo K-vecinos más próximos:

- a. Es más preciso con valores más grandes de K.
- b. Es más preciso si $K=3$.
- c. K no influye en la precisión del algoritmo.
- D. Es más preciso con valores más pequeños de K.**

Cuando se está ante un problema de clasificación no binaria, una Regresión Logística opera:

a. aplicando literalmente el mismo planteamiento que la Regresión Lineal Múltiple, sólo que usando como función de coste el distanciamiento entre la probabilidad obtenida y la deseada.

B. planteando clasificadores por pares, esto es, entre cada dos clases posibles y posteriormente hallar la salida mediante un sistema de votación entre dichos clasificadores por pares.

c. construyendo un clasificador mediante la ampliación del modelo matemático, al considerar tantas probabilidades de pertenencia, como clases presente el problema.

d. codificando de manera compacta la salida deseada operando, así, igual que una Regresión Lineal Múltiple

En el algoritmo de aprendizaje del perceptrón simple:

A. no se garantiza que la solución encontrada, de existir, sea la que mejor generalice

b. se emplearán más o menos iteraciones, dependiendo de si se mantiene fijo el bias o no.

c. la Regla de Heb se aplica literalmente

d. siempre se garantiza la solución

La robustez de las máquinas de vectores soporte frente a puntos aislados ¿es total?

a. Todo lo contrario, son de enorme sensibilidad a la aparición de este tipo de puntos.

b. Sí, la solución es independiente de que haya estos puntos aislados.

C. No, es parcial: sólo en aquellos casos en que los puntos aislados no modifiquen las envolventes convexas en la zona donde aparecen los vectores soporte, la solución seguirá siendo la misma.

d. No tiene sentido hablar de esta robustez, porque el algoritmo se bloquea ante esta clase de muestras.

¿Qué relación existe entre la Regla Delta y el Descenso del Gradiente Estocástico?

a. Lo segundo impide la elección aleatoria de muestras para procesar en una iteración, lo cual no sucede en el primer caso.

b. Ninguna, ya que se aplica a sistemas de aprendizaje automático diferentes.

c. Lo primero impone siempre ir acompañada de un término momento, mientras que en el segundo no hay restricción en este sentido.

D. Lo primero constituye el núcleo de lo segundo.

Cómo se garantiza que la salida de una neurona esté acotada, si se usa el rectificador lineal como función de activación:

a. siempre, porque el rectificador lineal, por definición, es acotado

B. teniendo siempre las entradas entre cero y uno

c. aplicando un filtro posterior que acote dicha salida

d. manteniendo una estructura de capas, ya que esto impide que la salida de una neurona se dispare a valores incontrolados.

¿Es posible que haya que adaptar numéricamente los valores de la salida de la capa oculta de una RBF para ser procesada por la capa de salida?

a. De salirse fuera del intervalo $[0,1]$, se cambia de función de base radial, por ejemplo a una logística.

B. En principio, no, ya que lo que viene a continuación es un adaline.

c. Siempre, ya que puede haber valores fuera del intervalo $[0,1]$.

d. No, pues está garantizado que los valores de salida de la capa oculta están estandarizados con media en el origen y desviación estándar unidad.

El término momento o de inercia en un perceptrón multicapa:

a. es la media de la variación que han sufrido los pesos dentro de la neurona en cuestión

B. es proporcional a la variación que ha sufrido el peso en la iteración anterior

c. es la media de la variación que han sufrido los pesos de las neuronas que pertenecen a la misma capa

d. es la variación que ha sufrido el peso en la iteración anterior

En el aprendizaje inductivo, se denomina Bias de preferencia a:

A. La preferencia del algoritmo por unas hipótesis frente a otras.

b. La preferencia del algoritmo por un lenguaje de representación de hipótesis.

c. La preferencia del algoritmo por unas instancias frente a otras.

d. La preferencia del algoritmo por una clase frente a otras.

En la Regresión Lineal Múltiple, la salida ante una muestra de entrada:

a. es la suma ponderada extendida a las salidas de cada clasificador simple

b. sería la media de la que se da en cada clasificador por separado

C. es la categoría del clasificador que dé la máxima salida

d. se obtiene mediante un sistema de votación entre los clasificadores especializados