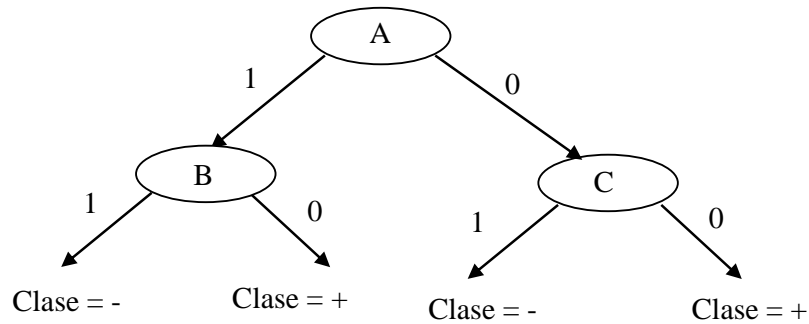


SISTEMAS INTELIGENTES

RELACIÓN DE PROBLEMAS DE APRENDIZAJE AUTOMÁTICO III

Ejercicio 1. Considérese el siguiente árbol de decisión:



Considérese el siguiente conjunto de datos de evaluación/prueba:

A	B	C	Clase
1	1	1	—
1	1	0	—
1	0	1	—
1	0	0	—
0	1	1	—
0	1	0	+
0	0	1	+
0	0	0	+

- Calcular la matriz de confusión obtenida para el árbol de decisión sobre el conjunto de prueba.
- Calcular los valores de exactitud (accuracy), fallout, recuerdo, y valor-F para el árbol de decisión y el conjunto de prueba.

Ejercicio 2. Se ha aplicado validación cruzada 10 veces (10-fold cross-validation) a un modelo dado H_1 sobre un dataset de 10 ejemplos (o items). La siguiente tabla muestra los resultados, donde y es el valor de salida de cada ejemplo en el dataset, e \hat{y}_i es la predicción dada para ese mismo ejemplo por el modelo entrenado con todos los ejemplos del dataset excepto el ejemplo i . (por ejemplo, la columna \hat{y}_7 proporciona la predicción de la función entrenada con el conjunto de ejemplos $\{1, 2, 3, 4, 5, 6, 8, 9, 10\}$)

item	y	\hat{y}_1	\hat{y}_2	\hat{y}_3	\hat{y}_4	\hat{y}_5	\hat{y}_6	\hat{y}_7	\hat{y}_8	\hat{y}_9	\hat{y}_{10}
1	1	0	1	1	0	1	1	1	0	1	0
2	1	1	1	1	1	1	1	1	1	0	1
3	0	0	0	1	0	0	0	1	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0
6	0	1	0	0	0	0	1	0	0	1	0
7	1	0	1	1	0	1	1	1	0	1	0
8	1	1	1	1	1	1	1	1	0	0	1
9	0	1	0	0	0	0	1	0	0	0	0
10	1	1	1	1	1	1	1	0	1	0	1

Se pide:

- Calcular los errores de entrenamiento y validación obtenidos por el procedimiento de validación cruzada.
- Supongamos que para otro modelo H2 el procedimiento de validación cruzada 10 veces sobre el mismo dataset proporcionó un error de entrenamiento de 0.0 y un error de validación de 0.5. ¿Qué modelo sería preferible?

Ejercicio 3. Supongamos que se aplica regresión polinómica con polinomios de grado entero D ($1 \leq D \leq 10$) sobre un dataset realizando validación cruzada 5 veces.

- Supongamos un dataset de 1000 ejemplos ¿Cuántas veces se ha ajustado el modelo? ¿Cuántos items contenía el conjunto de prueba?
- Supongamos igualmente que los resultados obtenidos para error de entrenamiento y validación son los mostrados en la figura (a). ¿Qué deberíamos hacer?
- Repetir el apartado anterior para las figuras (b), (c) y (d).

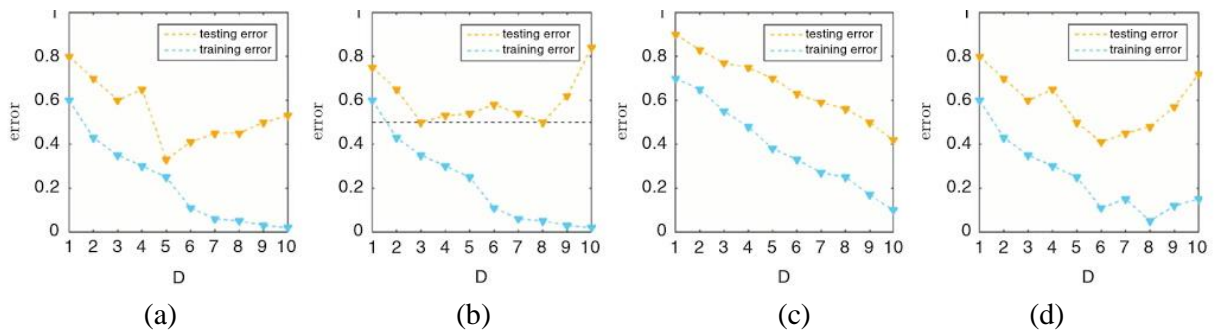
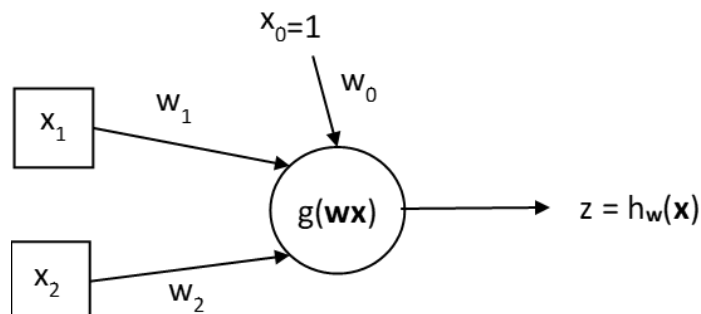


Figura 1

Ejercicio 4. Proporciona una definición clara y concisa de los siguientes conceptos en el contexto del aprendizaje automático:

Algoritmo del gradiente	Función RELU
Algoritmo del gradiente estocástico	Función sigmoide
Aprendizaje	Función tangente hiperbólica
Aprendizaje supervisado	Matriz de confusión
Árbol de decisión	Matriz de pérdida
AUC-ROC	Modelo de neurona artificial
Clasificación binaria	Perceptrón multicapa
Curva PR	Perceptrón simple
Curva ROC	Precisión
Dataset	Prevalencia
Entropía binaria	Recuerdo (recall)
Entropía cruzada binaria	Regresión lineal
Entropía de Shannon	Regresión logística
Error cuadrático medio	Regresión polinómica
Error de Bayes	Regresión y clasificación
Exactitud (accuracy)	Regularización
FPR (tasa de falsos positivos)	Sobreajuste (overfitting)
Frontera de decisión	Subajuste (underfitting)
Función logística	Tasa de aprendizaje
Función de activación	Tasa de error
Función de pérdida	TPR (tasa de auténticos positivos)
Función de verosimilitud	Umbral de decisión
Función logit	Validación cruzada k-veces
	Valor F (F-score, F-measure)

Ejercicio 5. Considérese el siguiente perceptrón simple, donde la función de activación es logística, y los pesos de la red son $(w_0, w_1, w_2) = (-1, 2, -1)$:



Considérese el siguiente conjunto de datos de evaluación/prueba:

x1	x2	Clase
0	0	-
1	0	+
1	3	-
2	4	+
-1	-1	-
2	2	+
-2	0	-
0	1	-
2	-2	+
3	0	+

- Calcular la matriz de confusión obtenida para el perceptrón sobre el conjunto de prueba.
- Calcular los valores de exactitud (accuracy), fallout, recuerdo, y valor-F para el perceptrón y el conjunto de prueba.

Ejercicio 6. Se ha aplicado validación cruzada 3 veces (3-fold cross-validation) a un modelo dado H1 sobre un dataset de 6 ejemplos (o ítems).

Para las sucesivas veces o entrenamientos (ent=1,2,3), los conjuntos de prueba/validación están formados por los ítems {1,2}, {3,4} y {5,6} respectivamente.

La siguiente tabla muestra el valor de la predicción de cada ítem para el modelo entrenado en cada una de las k veces, así como el valor real esperado y.

Ítem	y	Predicciones		
		ent=1	ent=2	ent=3
1	1	0	1	1
2	0	0	1	0
3	0	0	0	0
4	1	1	0	1
5	1	0	1	1
6	0	0	1	1

Se pide:

- Calcular los errores de entrenamiento y validación obtenidos por el procedimiento de validación cruzada.
- Supongamos que para otro modelo H2 el procedimiento de validación cruzada 3 veces sobre el mismo dataset proporcionó un error de entrenamiento de 0.3 y un error de validación de 0.6. ¿Qué modelo sería preferible?