

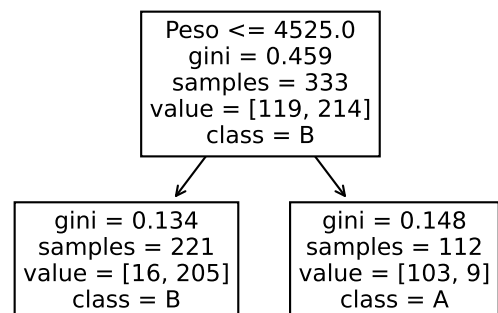
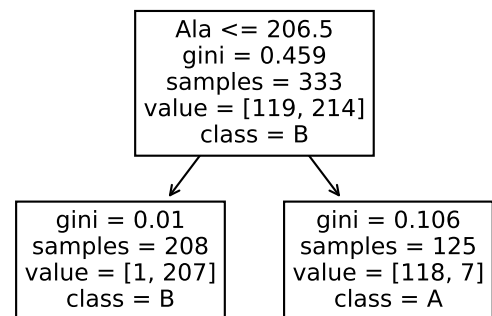
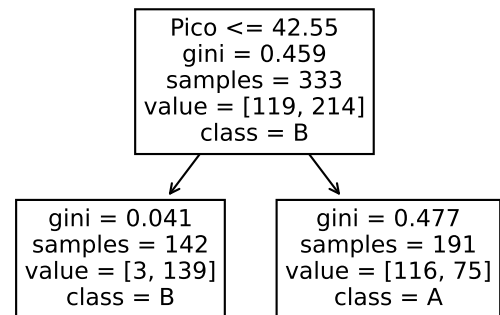
Fecha: 03/07/2024	Duración: 3 horas	Página 1 de 4
Evaluación: Parcial	Uso de Calculadora: SI	
Materia: Machine Learning para IA	Uso de Material: NO	
Turno: Nocturno	Puntaje Máximo: 30 Puntos	

Ejercicio 1

Se dispone de un conjunto de datos que contiene mediciones de la longitud del **Pico** (mm), **Ala** (mm) y el **Peso** (g) para dos especies de aves (A y B). Suponga que decide entrenar un árbol de decisión para predecir la especie en base a los tres atributos mencionados.

Preguntas:

1. Indique de qué tipo de problema de ML se trata.
2. ¿Qué funciones de pérdida (o criterios), específicos para árboles de decisión, puede usar para el entrenamiento?
3. Los diagramas a la derecha muestran las posibles divisiones que se pueden realizar en el nodo raíz. Indique cuál es la división que selecciona el algoritmo visto en clase. Justifique.
4. El algoritmo visto en clase divide los nodos de forma recursiva salvo que ciertos criterios de parada sean satisfechos. Indique cuáles son los hiperparámetros comúnmente utilizados que determinan estos criterios de parada.



Fecha: 03/07/2024

Evaluación: Parcial

Materia: Machine Learning para IA

Turno: Nocturno

Duración: 3 horas

Uso de Calculadora: SI

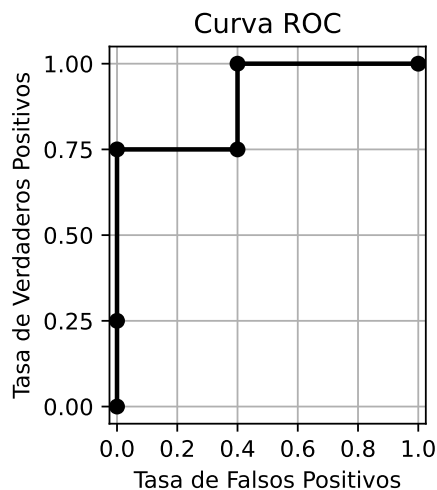
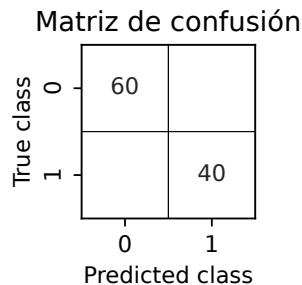
Uso de Material: NO

Puntaje Máximo: 30 Puntos

Página 2 de 4

Ejercicio 2

Se dispone de un dataset para entrenar un modelo que prediga si un cliente de una compañía telefónica renovará o no su contrato. En este contexto la clase positiva (etiqueta 1) consiste de aquellos clientes que NO renuevan.



- Suponga que luego de entrenar el modelo su matriz de confusión en un dataset de validación de $N = 140$ observaciones es la mostrada a la izquierda.
 - Calcule el Accuracy (exactitud).
 - Complete los valores faltantes sabiendo que la Precision vale 0.5 y el Recall vale 1.
- Sabiendo que este modelo va a ser usado con el fin de proponer descuentos (y por ende un costo adicional para la compañía) a aquellos clientes que probablemente NO renueven el contrato ¿cree que es un buen modelo? Justifique.
- La gráfica a la izquierda (abajo) muestra la curva ROC del modelo, ¿cuál es el punto que mejor representa a la matriz de confusión anterior (asumiendo los supuestos de 1.b))? Justifique.
- Calcule la AUC y evalúe su magnitud comparándola con su respuesta en el punto 2.

Fecha: 03/07/2024

Evaluación: Parcial

Materia: Machine Learning para IA

Turno: Nocturno

Duración: 3 horas

Uso de Calculadora: SI

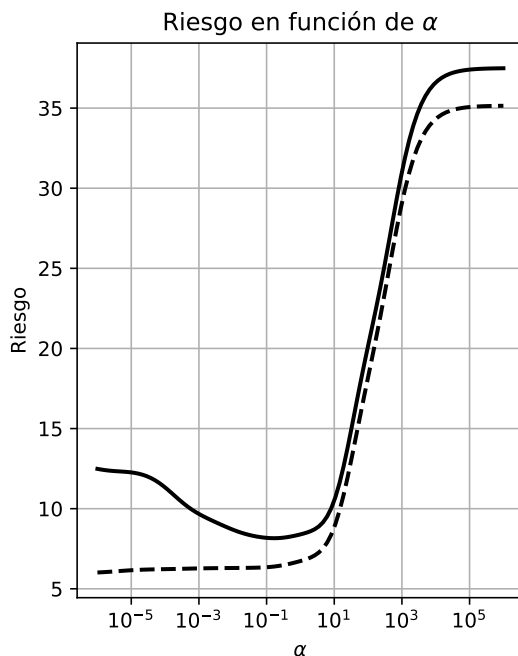
Uso de Material: NO

Puntaje Máximo: 30 Puntos

Página 3 de 4

Ejercicio 3

Suponga que está tratando de resolver un problema de regresión lineal, y observa que con un polinomio de grado 10 logra tener riesgo empírico muy pequeño en el conjunto de train. Sin embargo, observa que el riesgo empírico en el conjunto de validation es muy grande.



1. Indique cuáles son las posibles funciones de riesgo (o costo) que se utilizan comúnmente en este tipo de problemas de ML.
2. ¿Qué conclusión puede sacar sobre el modelo mencionado anteriormente?
3. Suponga que decide entrenar con la técnica de regularización Ridge. El gráfico de la izquierda representa el riesgo del modelo en función del parámetro de regularización α , tanto en train como en validation.

Identifique las curvas e indique la zona de sobreajuste. Justifique.
4. ¿Qué valor de α (aproximado) elegiría?

Fecha: 03/07/2024

Evaluación: Parcial

Materia: Machine Learning para IA

Turno: Nocturno

Duración: 3 horas

Uso de Calculadora: SI

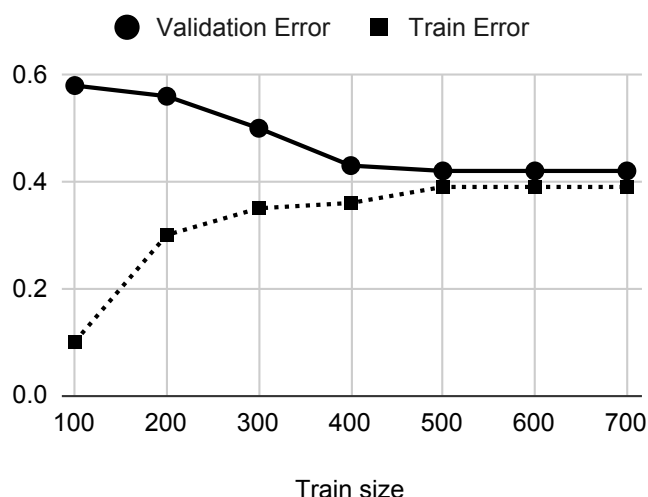
Uso de Material: NO

Puntaje Máximo: 30 Puntos

Página 4 de 4

Ejercicio 4

Para un problema de clasificación binaria se desea entrenar un modelo de ensemble con el algoritmo de Ada-boost. Para esto se dispone de un conjunto de train de $N = 700$ observaciones y un conjunto de tamaño similar para validation.



1. Si luego de la primera iteración del algoritmo el conjunto de train consiste de 500 observaciones con peso 0.001 y 200 observaciones con peso 0.0025.

¿Cuál es el error cometido por el primer modelo base?

2. A la izquierda se muestra el gráfico de la Learning Curve del modelo. ¿Qué defecto puede detectar en la misma?
3. ¿Qué puede hacer para intentar solucionarlo?
4. Describa brevemente cuáles son las principales diferencias entre los algoritmos de boosting y los de bagging.

Ejercicio 5

Las siguientes preguntas refieren al obligatorio:

1. En la competencia se utilizó la métrica F_1 -score. ¿Cree que es una métrica adecuada para el problema de clasificación de rostros? ¿Consideró otras métricas para evaluar y seleccionar modelos?
2. ¿Cuál fue la proporción de fondos a rostros que utilizó para entrenar el clasificador? ¿Cree que dicha proporción puede afectar el rendimiento del detector final?
3. Describa brevemente cuáles son en su opinión las debilidades y las fortalezas del enfoque de clasificación de rostros utilizado en el obligatorio.