

Taller 4

Entrega: Domingo 22 de Abril
Máximo 3 estudiantes por grupo

Prof. Fabio A. González
Introducción a los Sistemas Inteligentes - 2018-I
Ing. de Sistemas y Computación

Descargue el conjunto de datos nba.txt y lea la descripción del mismo en la cabecera del archivo. Este conjunto de datos será usado en los siguientes puntos, los cuales deben ser resueltos usando scikit-learn.

1. Entrene un modelo de clasificación basado en árboles de decisión.
 - (a) Haga una partición del conjunto de datos, usando muestreo estratificado, en 70% para entrenamiento y 30% para test (operador: Split Data).
 - (b) Entrene el modelo
 - (c) Aplique el modelo al conjunto de test
 - (d) Mida el desempeño del modelo calculando exactitud, error de clasificación, precisión, recall, curva ROC y matriz de confusión.
2. Interprete el modelo obtenido:
 - (a) Grafique el árbol obtenido
 - (b) Cuál es el atributo más discriminante? Tiene sentido? De una explicación a partir del conocimiento del problema.
 - (c) Genere 3 diferentes reglas de clasificación a partir del árbol. Explíquelas.
3. Complejidad del modelo y sobre-aprendizaje:
 - (a) Haga diferentes pruebas cambiando la profundidad máxima del árbol en el operador Decisión Tree. Pruebe los valores 1,2,...,10.
 - (b) Grafique la profundidad del árbol contra la evolución del error de entrenamiento y el error de prueba.
 - (c) De acuerdo con la gráfica, ¿Cuál sería un buen valor de profundidad para el árbol?
4. Curva de aprendizaje:
 - (a) Cambie la proporción de datos usados para entrenar usando diferentes valores: 10%, 20%,...,90%.
 - (b) Grafique la evolución del error de entrenamiento y el de prueba contra el porcentaje de ejemplos de entrenamiento.
 - (c) Explique el comportamiento de las curvas.
5. Comparación de modelos:
 - (a) Usando los datos de la NBA, entrene un modelo de clasificación Naïve Bayes y evalúelo usando validación cruzada con 10 pliegues.

- (b) Entrene un árbol de decisión y evalúelo usando validación cruzada con 10 pliegues.
- (c) Cuál de los dos modelos es mejor?

El taller debe enviarse como un Jupyter notebook a través del siguiente Dropbox file request, antes de la medianoche de la fecha límite. El archivo debe nombrarse como `iis-taller4-unalusername1-unalusername2-unalusername3.ipynb`, donde `unalusername` es el nombre de usuario asignado por la universidad (incluya los nombres de usuario de todos los miembros del grupo). En caso de que necesite incluir archivos de respaldo además del notebook, envíelo como un archivo zip que contenga todos los archivos y el notebook.