

INTELIGENCIA ARTIFICIAL (3398/1967)

Guía Práctica Nro. 4 - ML - Supervisado - Clasificación NO paramétrica y Ensembles

- 1) Se quiere clasificar el punto $Q = (3,2)$ con $k = 3$.
 - a) Calcular la distancia euclidiana de Q a cada punto.
 - b) Ordena los vecinos.
 - c) Decide la clase final.
 - d) Resuelva ahora con distancia Manhattan.

Punto	X	Y	Clase
P1	1	2	A
P2	2	3	A
P3	3	3	B
P4	5	1	B
P5	6	2	B

- 2) Genere utilizando la librería scikit learn un dataset sintético con diferentes ruidos y analice el comportamiento con diferentes k . Qué sucede con $k=1$, con k muy grandes. Evalúe, con las métricas conocidas, los modelos aprendidos generando datos de entrenamiento y evaluación.

from sklearn.datasets import make_moons

X, y = make_moons(n_samples=100, noise=0.35, random_state=42)

- 3) Tome los dataset ejemplo del ejercicio anterior y resuelva los mismos utilizando SVM. Realice pruebas utilizando diferentes kernels y soft margin. Compare los resultados.
- 4) Se cuenta con los siguientes datos de un estudio sobre rendimiento académico y se desea entregar un modelo predictivos basado en árboles de decisión:

ID	Horas	Asistencia	Tareas	Resultado
1	Baja	Baja	No	Reprobado
2	Baja	Alta	No	Reprobado
3	Media	Alta	No	Aprobado
4	Alta	Alta	Si	Promoción
5	Media	Baja	Si	Aprobado
6	Alta	Alta	No	Aprobado
7	Baja	Baja	Si	Reprobado
8	Alta	Baja	Si	Aprobado
9	Media	Alta	Si	Promoción
10	Alta	Alta	Si	Promoción

- a) Construya manualmente el DT utilizando el algoritmo CART. Utilice índice Gini como métrica de importancia.
- b) Implemente en python utilizando scikit learn

INTELIGENCIA ARTIFICIAL (3398/1967)

Guía Práctica Nro. 4 - ML - Supervisado - Clasificación NO paramétrica y Ensembles

- 5) Construya un modelo utilizando DT para el dataset elegido en la práctica nro 2.
 - a) Construya varios modelos variando las técnicas de pruning y analice los resultados respecto del modelo de regresión producido.
 - b) Realice una consulta por un caso y analice las condiciones que llevan a la predicción obtenida
- 6) Descargue el dataset que acompaña la práctica sobre el tiempo en pantalla vs la calidad de sueño.
 - a) Realice la partición del DataSet para entrenamiento y evaluación
 - b) Analice y previsualice la información (solo de entrenamiento)
 - c) Entrene y evalúe con las métricas abordadas los diferentes modelos.
 - d) ¿cuáles son las características comunes de las personas con baja calidad de sueño?
- 7) Retome el dataset “ruidoso” del ejercicio y compare entrenando un modelo con DT y luego aplicando bagging, particularmente, Random Forest. Analice los resultados con matrices de confusión.
- 8) Compare los resultados del ejercicio anterior aplicando la técnica de boosting entre diferentes técnicas (DT, KNN).
- 9) Construya un modelos utilizando Stacking para el dataset Iris. Utilice como modelos base DT, KNN y LogisticRegresion. Pruebe con distintos meta-modelos para analizar las mejoras.