



03/11/2023

UNIDAD 3 - ACTIVIDAD 3: ENSAMBLES Y BOSQUES ALEATORIOS.

Grupo B

Descripción breve

Se busca que el estudiante pueda probar los modelos de ensamble en un problema simple.

M.I.A. Wilfrido Cortes Orozco
wilfrido.co@morelia.tecnm.mx

Objetivos de la actividad.

Esta actividad busca cumplir los siguientes objetivos:

- Que el estudiante pueda crear un bosque aleatorio.
- Que el estudiante pueda identificar las ventajas y desventajas de los ensambles y los bosques aleatorios.

Descripción de la actividad.

El set de datos para esta actividad se llama “diabetes.csv”, es un conjunto de datos con información acerca de pacientes que tienen (o no) diabetes. Este dataset ya se ha utilizado previamente. Esta actividad requiere que se haya completado la actividad anterior.

Hay que realizar un análisis exploratorio de los datos breve:

1. Cargar el set de datos en un dataframe de pandas.
2. Mostrar los mejores resultados para ambos problemas anteriores (regresión y clasificación).
3. Importe las librerías para el VotingClassifier y el RandomForest (tanto para regresión como para clasificación).
4. Entrene un modelo de votación para regresión con 3 modelos de regresión de su preferencia (por ejemplo: regresión lineal, árbol de decisión, KNeighborRegressor), utilice las mismas muestras de “X” y “y” que en la actividad anterior.
5. Pruebe el modelo de votación con la muestra de pruebas y evalúe con MSE.
6. ¿Tuvo mejores o peores resultados?, ¿qué tanto tardó en entrenarse?
7. Entrene un modelo de votación para clasificación con 3 modelos de clasificación de su preferencia (por ejemplo: regresión logística, árbol de decisión y KNeighborClassifier), utilice las mismas muestras de “X” y “y” que la actividad anterior.
8. Pruebe el modelo de votación con la muestra de pruebas y evalúe con Accuracy_Score y F1_Score.
9. ¿Tuvo mejores o peores resultados?, ¿qué tanto tardó en entrenarse?
10. Entrene un modelo de bosque aleatorio para clasificación, puede emplear los hiperparámetros que considere adecuados (de esto dependerá la eficiencia del modelo), para el entrenamiento se debe usar la misma muestra que en los casos anteriores.
11. Se debe evaluar el modelo con F1_Score y Accuracy_Score.
12. ¿Qué tal se comportó?, ¿mejoró o empeoró con respecto a los modelos anteriores?, ¿tomó más o menos tiempo para entrenarse?
13. Entrene nuevamente el bosque aleatorio, pero cambiando los hiperparámetros de manera radical. ¿Qué tal se comportó el modelo?
14. Escriba sus conclusiones obtenidas al utilizar modelos de ensamble.