

Minería de Datos

Lab. #6

Ivan Saavedra, Ph.D.

saavedrai@uninorte.edu.co

Universidad del Norte
División de Ingenierías
Dpto. Ingeniería de Sistemas



202110

1. Aprobación de Crédito

NB, DT, KNN

Aplice los diferentes clasificadores vistos en clase para clasificar si una persona seria aprobada para un prestamos en una entidad bancaria. Esta decisión se encuentra basada en atributos como el sexo, estado marital, numero de dependientes, nivel de educación, ingresos, entre otros.

Considere los conjuntos de datos para entrenamiento y testeo suministrados para entrenar y testear el modelo respectivamente.

Metodología

- Realizar una análisis descriptivo exploratorio
 - Analizar frecuencia de la respuestas a la solicitud de crédito
 - Analizar la frecuencia de aplicantes por sexo, estado marital, historia crediticia, y nivel de educación
 - Analizar por medio de un diagrama de cajas el nivel de ingresos de los aplicantes y la cantidad del crédito
 - Analizar matriz de correlación de los atributos con la variable de respuesta
- Entrenar los modelos
 - Visualizar el árbol de decisión encontrado para el modelo de Árbol de Decisiones
- Encontrar predicciones
- Evaluar precisión de los modelos
- Comparar precisiones de los modelos. Cual es el modelo que presenta mejor precisión ?

2. Clasificación de Dígitos

NB, DT, KNN

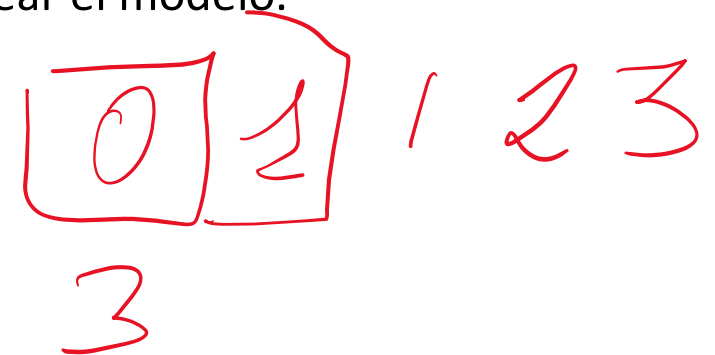
Aplice los diferentes clasificadores vistos en clase para clasificar dígitos que han sido escritos a mano. Esta decisión se encuentra basada en la codificación de las imágenes de los dígitos en una matriz de píxeles de 28x28, representando dígitos de 0 a 9.

Considere los conjuntos de datos (MNIST) suministrados para entrenar y testear el modelo.

Metodología

- Realizar una análisis descriptivo exploratorio
 - Visualice un ejemplo de imagen de cada dígito
 - Analizar por medio de un diagrama de cajas los valores de los atributos en general
 - Analizar matriz de correlación de los atributos con la variable de respuesta
- Entrenar los modelos
 - Visualizar el árbol de decisión encontrado para el modelo de Árbol de Decisiones
- Encontrar predicciones
- Evaluar precisión de los modelos
- Comparar precisiones de los modelos. Cual es el modelo que presenta mejor precisión ?

*0, 1, 2, ..., 9
Visualizar*



Modelos de Clasificación

Entregable:

- Un archivo de Jupyter Notebook con el desarrollo de los análisis
- Las conclusiones y respuestas al objetivo del análisis deben ser contestadas en el mismo notebook.
- Se sugiere que comente las secciones de manera adecuada para una mejor interpretación de su análisis.
- La fecha de entrega es el Sábado 01 de Mayo del 2021 vía catalogo web enlace de laboratorios.