

Temas Curso ML

Python

En esta sección se introducirá Python como lenguaje base y las librerías necesarias para su uso en análisis de datos (pandas, numpy, matplotlib) a través de una introducción a sumariación y visualización (Data Profiling, Estadísticos, Visualización).

Análisis Supervisado

Los algoritmos supervisados se usan cuando se quiere predecir una etiqueta (regresión para números y clasificación para categorías).

Las métricas de desempeño usan las etiquetas ya existentes en los datos (e.g. transacciones etiquetadas como *fraude* o *no fraude*, o búsquedas dentro de una tienda virtual etiquetadas como *compra* o *no compra*) para evaluar si el modelo predice bien. Esto será la base para poder elegir entre los distintos modelos según su ajuste en los datos usados.

Dentro de esta sección se verán distintos algoritmos, comentando sus ventajas y desventajas. Se buscará un entendimiento sobre los hiperparámetros que se pueden usar para mejorar su ajuste.

Clasificación

- Discriminante Lineal
- Regresión Logística
- Árboles de decisión
- Bosques Aleatorios
- K-vecinos más cercanos
- Naive Bayes
- Redes Neuronales
- Support Vector Machines (método de Kernel)

Regresión

- Lineal
- Autoregresión

Además de algunos algoritmos de clasificación que también se pueden adaptar para regresión:, Árboles de decisión, Bosques Aleatorios, K-vecinos más cercanos, Redes Neuronales.

Aprendizaje No Supervisado

Este se usa para datos no etiquetados. Es decir, que no existe una respuesta “correcta” o al menos no se tiene forma de evaluarla. Dentro de este tipo de aprendizaje se verán de dos tipos:

- Agrupación. separar los datos en grupos, de manera que cada grupo comparta características comunes.
- Asociación. Consiste en encontrar relaciones entre individuos para poder predecir recomendaciones apropiadas (e.g. recomendar productos basados en búsquedas).

Agrupación

- Agrupamiento Jerárquico
- K-medias
- Mapas autoorganizados

Asociación

- Reglas a priori
- Collaborative Filtering
 - Memory-based
 - Model-based

Preprocesamiento y reducción de dimensionalidad

Dentro de análisis supervisado, se verán algunas herramientas para reducir el número de variables para evitar la *maldición de la dimensionalidad* y eliminar dependencias, así como para visualizar datos.

- PCA
- tSNE
- Análisis de factores
- Análisis de correspondencias
- Análisis de canónico

Imágenes

Las tecnologías actuales para usar imágenes como datos se suelen basar en estructuras particulares de Redes Neuronales llamadas Redes Neuronales Convolutivas.

Dentro de este tema se tendrá que introducir algún lenguaje usado para el uso de estos modelos (Pytorch, tensorflow/keras). Y se usarán modelos preentrenados para aprender cómo ajustar los pesos a nuevos datos.