ITESO, Universidad
Jesuita de Guadalajara

Ingeniería de características

Dr. Gaddiel Desirena López

Primavera 2023

Descripción de la asignatura

Los datos representan un activo potencialmente valioso para la sociedad, empresas y gobiernos. Sin embargo, antes de poder aprovechar los datos mediante el uso de técnicas de áreas como aprendizaje automático y minería de datos, es importante poder seleccionar las características relevantes para un problema o pregunta con el objetivo de reducir la dimensionalidad del conjunto de datos.

Descripción de la asignatura

Por lo anterior, en este curso se analizará la importancia de la ingeniería de características, así como los diversos algoritmos utilizados para la selección de características, la construcción de características compuestas y la reducción de dimensionalidad de conjunto de datos numéricos y alfanuméricos.

Propósito general

Limpiar conjuntos de datos y sustraer sus características para crear nuevas variables que puedan ser usadas como entradas por un algoritmo de aprendizaje máquina.

Temario

1. Extracción de datos de diferentes fuentes
   - 1.1 Archivos de texto, CSV y Excel
   - 1.2 Archivos JSON, XML y SHP
   - 1.3 De imágenes
2. Identificación de datos
   - 2.1 Variables numéricas y variables categóricas
   - 2.2 Valores faltantes
   - 2.3 Cardinalidad en variable categóricas
   - 2.4 Relaciones lineales
   - 2.5 Distribuciones de los datos
   - 2.6 Valores atípicos
3. Tratamiento de datos faltantes
   - 3.1 Eliminación de observaciones
   - 3.2 Sustitución por media y mediana
   - 3.3 Sustitución por moda y frecuencia
   - 3.4 Sustitución aleatoria
   - 3.5 Valores extremos

Temario

Temario

Evaluación

Actividad	Ponderación
Exámenes	30 %
Prácticas de laboratorio	15 %
Tareas	15 %
Proyecto de limpieza de conjunto de datos	40 %

Bibliografía

Título	Autor
Feature Engineering for Machine Learning and Data Analytics	Guozhu Dong, Huan Liu
Feature Engineering and Selection: A Practical Approach for Predictive Models	Max Kuhn, Kjell Johnson
Python Feature Engineering Cookbook: Over 70 recipes for creating, engineering, and transforming features to build machine learning models	Soledad Galli

Bibliografía

Título	Autor
Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists	Alice Zheng, Amanda Casari
Machine Learning Refined: Foundations, Algorithms, and Applications	Jeremy Watt, Reza Borhani, Aggelos Katsaggelos
Texture Feature Extraction Techniques for Image Recognition	Jyotismita Chaki, Nilanjan Dey

¿Qué es Ingeniería de características?

La ingeniería de características es el proceso de utilizar el conocimiento del dominio para extraer características de datos sin procesar mediante técnicas de minería de datos [1].

Motivación

- Es un paso crucial en la canalización del aprendizaje automático, porque las características adecuadas pueden aliviar la dificultad del modelado y, por lo tanto, permitir que la canalización produzca resultados de mayor calidad.
- Los profesionales están de acuerdo en que la gran mayoría del tiempo en la construcción de una canalización de aprendizaje automático se dedica a la ingeniería de características y la limpieza de datos.

Definiciones

Datos: son observaciones de fenómenos del mundo real.

Aprendizaje automático: Es un subconjunto de la Inteligencia Artificial y se refiere al estudio de algoritmos que mejoran a través de la experiencia [2].

Característica: Es una propiedad individual medible de un fenómeno que se observa. Éstas pueden ser numéricas o estructurales [3].

Conocimiento del dominio: Consiste en saber con precisión cómo se hace algo, tener una intuición de los principios subyacentes e integrarlos en la red de conocimiento existente.

Definiciones

Minería de datos: Es un conjunto de técnicas que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones que expliquen el comportamiento de estos datos.

Herramientas

https://www.python.org/

https://colab.research.google.com/

Herramientas

`https://www.anaconda.com/download/`

Biblios

- https://en.wikipedia.org/wiki/Feature_engineering
- https://en.wikipedia.org/wiki/Machine_learning
- https://en.wikipedia.org/wiki/Feature_(machine_learning)