

Introducción a los métodos de aprendizaje automático

UT02 – Trabajo de
Aplicación 1



Universidad
Católica del
Uruguay

TA1 - Ejercicio 1

- Abrir y explorar el conjunto de datos “Titanic” en data.world (buena herramienta!)
<https://data.world/nrippner/titanic-disaster-dataset>
- ... ó descargar de
<https://gist.github.com/michhar/2dfd2de0d4f8727f873422c5d959fff5>
- definir el problema de predicción
- analizar atributos y comprender / documentar su significado en el problema
- revisar atributos faltantes y estadísticas de los atributos
- revisar visualizaciones de relaciones entre atributos y variable de clasificación
- generar un POSTER con todo lo analizado

TA1 - Ejercicio 2

- Utilizando RM, aplicar los diferentes bloques para corregir los defectos encontrados en el Ej.1
- Finalmente visualizar el dataset resultante (estadísticas, correlaciones, gráficos,etc.)
- Documentar qué se hizo en cada caso y por qué

TA1 - Ejercicio 3 analizar relaciones de atributos en dataset “Titanic”

- Primero usar scatter plot matrix y visualizar
- Generar matriz de correlación usando el operador “Correlation Matrix”, estudiar los parámetros que acepta el operador
- Observar la matriz resultante y los pesos asignados a los atributos.
- Operadores útiles para gestionar los atributos altamente correlacionados:
 - Select by Weights
 - Remove correlated attributes
 - Weight by Correlation

TA3 - Ejercicio 2

Para cada bloque de detección de outliers de RM, generar un resumen:

- breve descripción
- parámetros que acepta
- característica de aplicación más importante

TA3 - Ejercicio 3

Cargar el dataset “Iris”

- analizar atributos (tipos de datos, rangos, distribuciones, etc.)
- visualizar scatter plots
- normalizar
- aplicar modelo: PCA
- agregar bloque de detección de outliers por distancia
 - $K=1$, outliers = 10 , distancia euclideana

