

Aprendizaje NO supervisado



Diplomatura en Ciencia de Datos,
Aprendizaje Automático y sus Aplicaciones
FaMAF-UNC
agosto 2018

Qué es aprendizaje no supervisado?

Análisis exploratorio de datos (*Exploratory Data Analysis*)

- Detección de anomalías
- Prevención de fallas

Data mining

- Reglas de asociación
- Segmentación de perfiles

Acercarse a las **causas latentes** de los **fenómenos observables**

Para qué?



- Cuando no sabemos lo que queremos
- Cuando sospechamos de los datos
- Para refinar las clases que queremos

Tecnologías relacionadas

- ANOVA, testeo de hipótesis
- Proyecciones (*embeddings*)
- Reglas de asociación
- Vecinos más cercanos (recomendación)
- Clustering
- Detección de Anomalías
- Propiedades de Grafos

Problemas metodológicos

- No hay evaluación intrínseca
 - Evaluación indirecta, por impacto en otras aplicaciones
 - Evaluación interpretativa subjetiva
- Evaluación anecdótica, nunca exhaustiva
- Medidas de calidad de utilidad cuestionable

La clave está en hacer buenas preguntas, y expresar los datos buscando respuestas

- Un espacio de búsqueda muy grande → mínimos locales

Aplicaciones clásicas

- Análisis del carrito de la compra
- Segmentación de mercado (clientes)
- Caracterización de comportamiento de usuarios
- Detección de fallos en líneas de producción
- Detección de fraude
- Detección de temas en documentos
- Detección de tipos de objetos en imágenes
- Detección de comunidades

Mapa de ruta

1. Embeddings
2. Clustering
3. Reglas de Asociación
4. K-nn y recomendación
5. Grafos
6. Aprendizaje Semi-supervisado

Entregables:

- Clustering
- Recomendación