

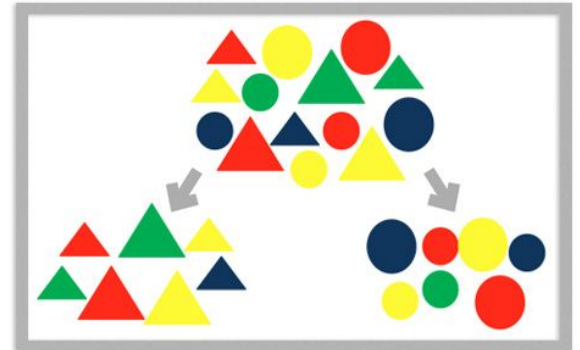
Las Tareas que podemos
resolver con ML



Clasificación

Una tarea de clasificación consiste en, dado un individuo, saber a qué clase pertenece, basándonos en lo que hemos “aprendido” de otros individuos. Por ejemplo:

¿Qué clientes de Telefónica estarán interesados en esta oferta?



Clasificación

Basándonos en la información del histórico de clientes, resumida en una serie de variables como puede ser la edad, estado civil, nivel de estudios, antigüedad como cliente etc., los algoritmos de clasificación construyen un modelo que nos permite asignar, a un nuevo cliente la etiqueta más adecuada entre estas dos: “Estará interesado” o “No estará interesado”.

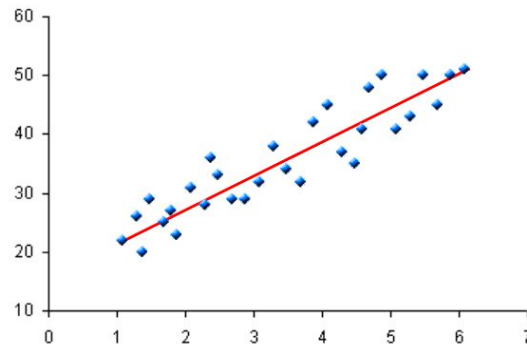
Los algoritmos de scoring son muy similares, pero más específicos. Nos dan la probabilidad de que un cliente esté interesado o no.



Regresión

Las tareas de regresión se utilizan cuando lo que se quiere averiguar es un valor numérico de una variable continua. Siguiendo con el ejemplo anterior, nos servirían para, basándonos en el histórico de consumo de los clientes, parametrizado según las variables anteriores, podamos responder a preguntas como ésta:

¿Cuál va a ser el consumo en ... (voz, datos, etc.) de este cliente en un mes?



Identificar similitudes

Se trata de identificar individuos “similares” según la información que tenemos de ellos. Es la base de los sistemas de recomendación, que te ofrecen distintos productos según los que hayas consultado o adquirido previamente.



Clustering

Las tareas de clustering tienen que ver con agrupar individuos por su similitud, pero sin un propósito específico. Suele usarse en las fases de exploración preliminar de los datos, para ver si existe algún tipo de agrupamiento natural, que puede sugerir la mejor forma de analizar los datos. Por ejemplo, estas tareas nos darían respuesta a preguntas como:

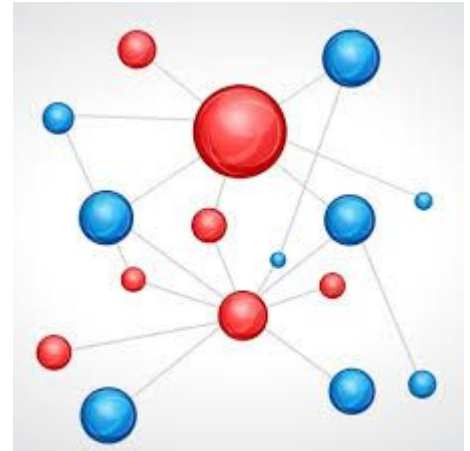
¿Se puede clasificar a nuestros clientes en grupos o segmentos naturales? ¿Qué productos deberíamos desarrollar?



Agrupar co-ocurrencias

Esta tarea busca asociaciones entre “entidades” basadas en su coincidencia en transacciones. Por ejemplo, responderían a la pregunta:

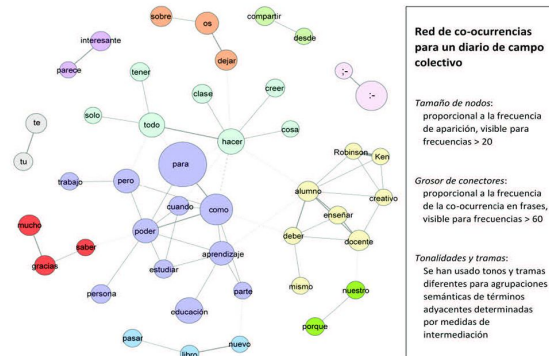
¿Qué productos se suelen comprar juntos?



Agrupar co-ocurrencias

Mientras las técnicas de clustering buscan agrupar elementos, basándose atributos de éstos, la agrupación de co-ocurrencias se basa en que dichos elementos aparezcan juntos en una transacción. Por ejemplo, es habitual que una persona que compra una cámara de fotos, compre también una funda para la cámara, o una tarjeta de memoria.

Por ello, puede ser interesante hacer promociones de ambos productos a la vez. Sin embargo, a veces no son tan “evidentes” las coincidencias y por eso resulta muy interesante analizarlas.



Profiling

Cuando hablamos de Profiling, hablamos de comportamientos típicos. Estas técnicas buscan caracterizar el comportamiento esperado de un individuo, grupo o población. Se pueden plantear preguntas como:

¿Cuál es el consumo de móvil típico de este segmento de clientes?



Profiling

La descripción de estos comportamientos “típicos” se suele usar como referencia para detectar comportamientos inusuales o anomalías. Basándonos en las compras típicas de un determinado cliente, podemos detectar si un nuevo cargo en su tarjeta de crédito se ajusta a ese patrón.

Podemos crear asignarle un “score” o grado de sospecha de fraude, y lanzar una alerta cuando se supere cierto umbral.



Predicción de vínculos

Intenta predecir conexiones entre elementos. Por ejemplo, entre miembros de una red social o profesional. Te hacen sugerencias como:

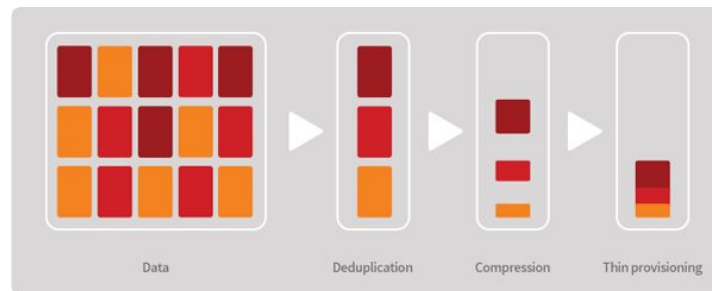
“María y tú tenéis 10 amigos en común. ¿No deberíais ser amigas?” “Personas que probablemente conozcas”



Reducción de datos

A veces es necesario reducir el volumen de datos de trabajo. Por ejemplo, en lugar de trabajar con una enorme base de datos de preferencias de consumo de películas, trabajar con una versión reducida de ellos, como sería el “género” de la película, más que la película concreta.

Siempre que se realiza una reducción de datos se pierde información. Lo importante es llegar a una solución de compromiso entre la pérdida de información y la mejora de los Insights.



Modelado Causal

Estas tareas lo que buscan es detectar la influencia de unos hechos sobre otros. Por ejemplo, si se incrementan las ventas en un grupo de clientes a los que nos hemos dirigido con una campaña de marketing:

¿Se incrementaron las ventas gracias a la campaña o simplemente el modelo predictivo detectó bien a los clientes que hubieran comprado de cualquier forma?

En este tipo de tareas es muy importante definir bien las condiciones que se tienen que dar para poder hacer esa conclusión causal.

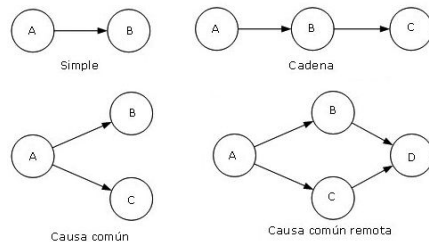


Fig. 1. Grafos causales que representan distintos tipos de causalidad.

Por tanto, cuando queremos abordar con ML un problema de negocio, como el típico ejemplo de fuga de clientes o churn , lo que queremos averiguar es qué clientes están más o menos predispuestos a dejar de serlo.

Podríamos abordarlo como un problema de clasificación, o clustering, incluso como un problema de regresión. **Según cómo definamos el problema, trabajaremos con una familia de algoritmos u otra.**