

# Ciencia de Datos

- Módulo 3

## Tipos de problemas



## ¿Cómo escogemos qué modelo probar?

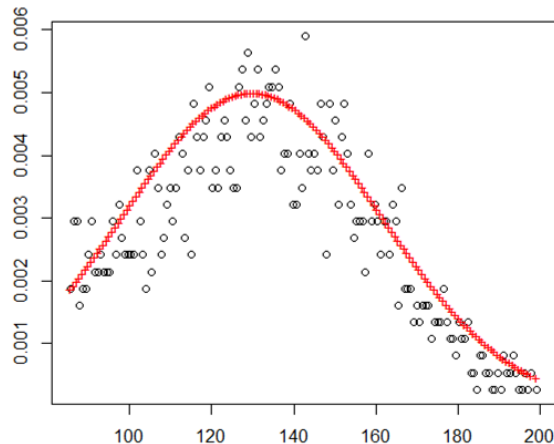
Hasta ahora vimos varios modelos y pasos para resolver una determinada pregunta, pero ¿Cómo escogemos qué modelo probar?

Determinar qué algoritmo utilizar depende de muchos factores, desde el tipo de problema en cuestión hasta el tipo de resultado que se está buscando.

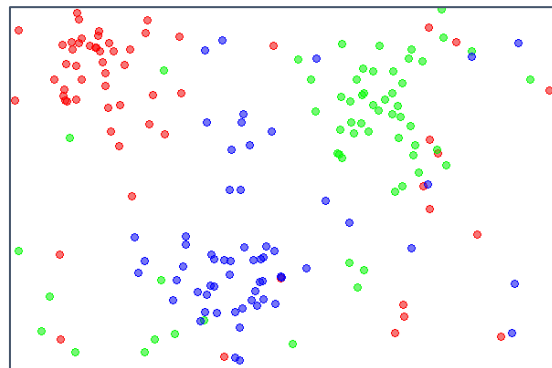


# Tipos de problemas

Lo primero que debemos estudiar es la pregunta a responder.



¿Buscamos un valor numérico?



¿Buscamos un atributo?



¿Buscamos agrupar?

## Entender los datos

Una vez que sabemos lo que queremos responder nos toca entender nuestros datos. Esta etapa es importante para obtener toda la información posible, y así escoger mejor todos nuestros pasos en la preparación de los datos y la selección de los algoritmos a probar.



<https://y26uq11r8xr1zyp0d3inciqv-wpengine.netdna-ssl.com/wp-content/uploads/2019/10/35-1-1024x597.jpg>



## Consideraciones

Finalmente, para escoger el algoritmo a usar debemos tomar en consideración:

- Si se tomara en cuenta la interpretabilidad del modelo o solo la salida del mismo.
- Qué tan complejo puede ser el modelo.
- Si se requiere una solución escalable o no.
- En cuanto tiempo se tiene para entrenar, probar e implementar.



## Tipos de problemas - Valor numérico

¿Buscamos un valor numérico?

Ejemplos:

- ¿Cuanto venderemos el próximo mes?
- ¿Cuanto nos tardamos en ir de un lugar a otro?
- ¿Qué valor tendrá una acción al finalizar la semana?
- ¿Cual es la probabilidad de que un paciente tenga determinada enfermedad?

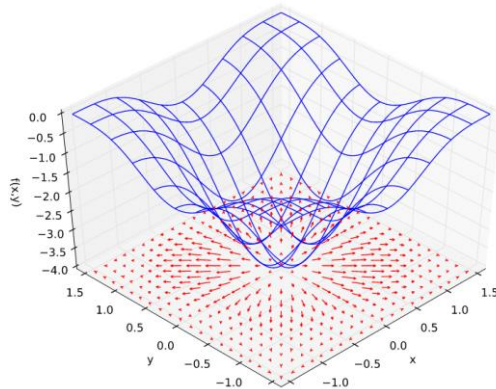


## Tipos de problemas - Valor numérico

¿Buscamos un valor numérico?

Entonces estamos en un problema de regresión, queremos hallar la línea que, dadas nuestras variables de entrada, mejor aproxime el valor que queremos predecir.

Posibles modelos: Regresión lineal, xgBoost, redes neuronales, etc.



## Tipos de problemas - Atributo

---

¿Buscamos un atributo?

Ejemplos:

- Detección de compras con TDC fraudulentas.
- Aprobar o no un crédito.
- Definir si un comentario es positivo o no.
- Clasificación de correos como spam.
- Reconocimiento de animales u objetos.



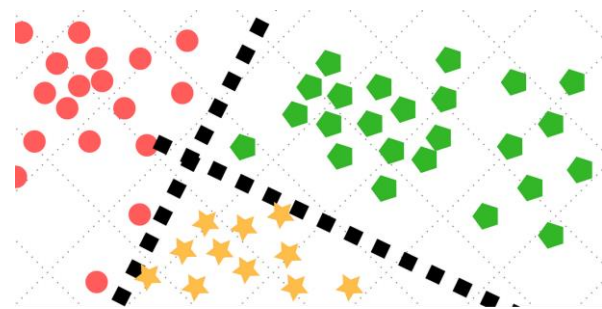


## Tipos de problemas - Atributo

¿Buscamos un atributo?

Entonces estamos en un problema de clasificación, queremos hallar las líneas que mejor separan nuestras variables tal que, cada grupo pertenezca a la clase correcta a predecir.

Posibles modelos: Regresión logística, random forest, SVM, etc.



<https://www.thedataschool.com.au/wp-content/uploads/2018/10/text-classification-1080x545.png>

## Tipos de problemas - Agrupar

---

¿Buscamos agrupar?

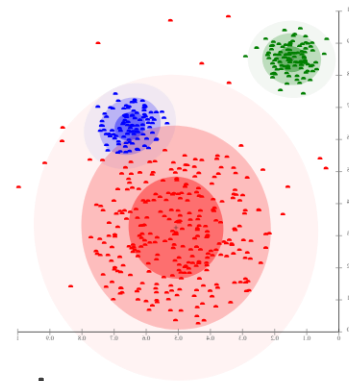
Ejemplos:

- Segmentar una base de clientes.
- Agrupar libros similares.
- Agrupar imágenes.
- Diversificar inversiones.



## Tipos de problemas - Agrupar

¿Buscamos agrupar?



Entonces estamos en un problema de agrupación o clusters, queremos hallar un número determinado de grupos tales que los elementos en cada grupo se parezcan entre ellos.

Posibles modelos: K-means, aglomeración jerárquica, maximización de la esperanza, etc.

