FUNDACIÓN YPF

Módulo V | Aprendizaje supervisado

Clase 13

Algoritmos más comunes parte 2

FUNDACIÓN YPF

¿Ponemos a grabar el taller?

FUNDACIÓN YPF

¿Qué vamos a ver hoy?

- Repasamos Árboles de Decisión
- Random Forest
- K-Nearest Neighbors
- Support Vector Machine

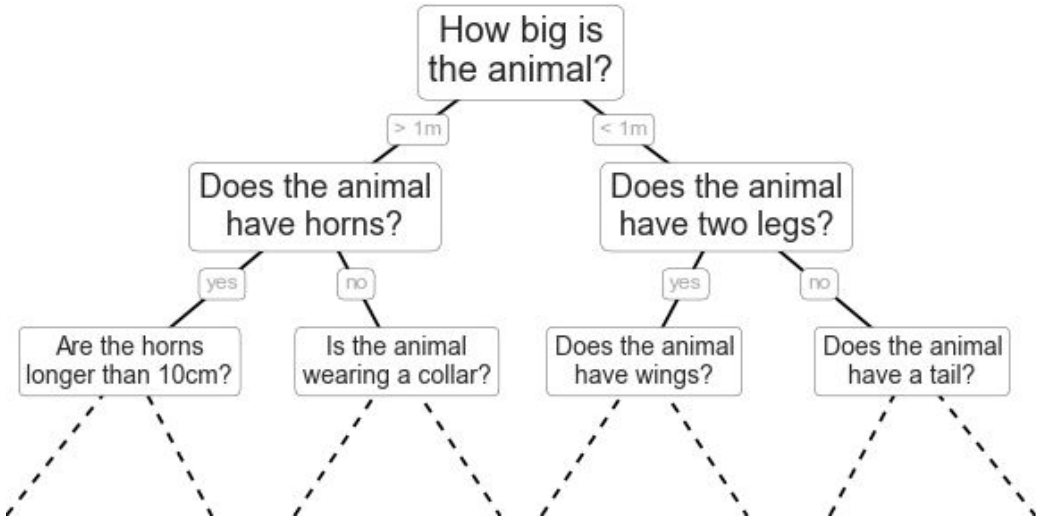
FUNDACIÓN YPF

Repasamos

FUNDACIÓN YPF

Árboles de decisión

Un árbol de decisión **hace preguntas** y va clasificando de acuerdo a las respuestas.



FUNDACIÓN YPF

Árboles de decisión

- Clasificación y Regresión
- Simple de entender, interpretar y visualizar
- Modelo base para modelos más complejos (Random Forest, etc)
- **Parámetros para nodos**:
  - Impureza de Gini
  - Ganancia de información
- Solo toman atributos numéricos
- Tendencia a ajustarse a los datos de entrenamiento

FUNDACIÓN YPF

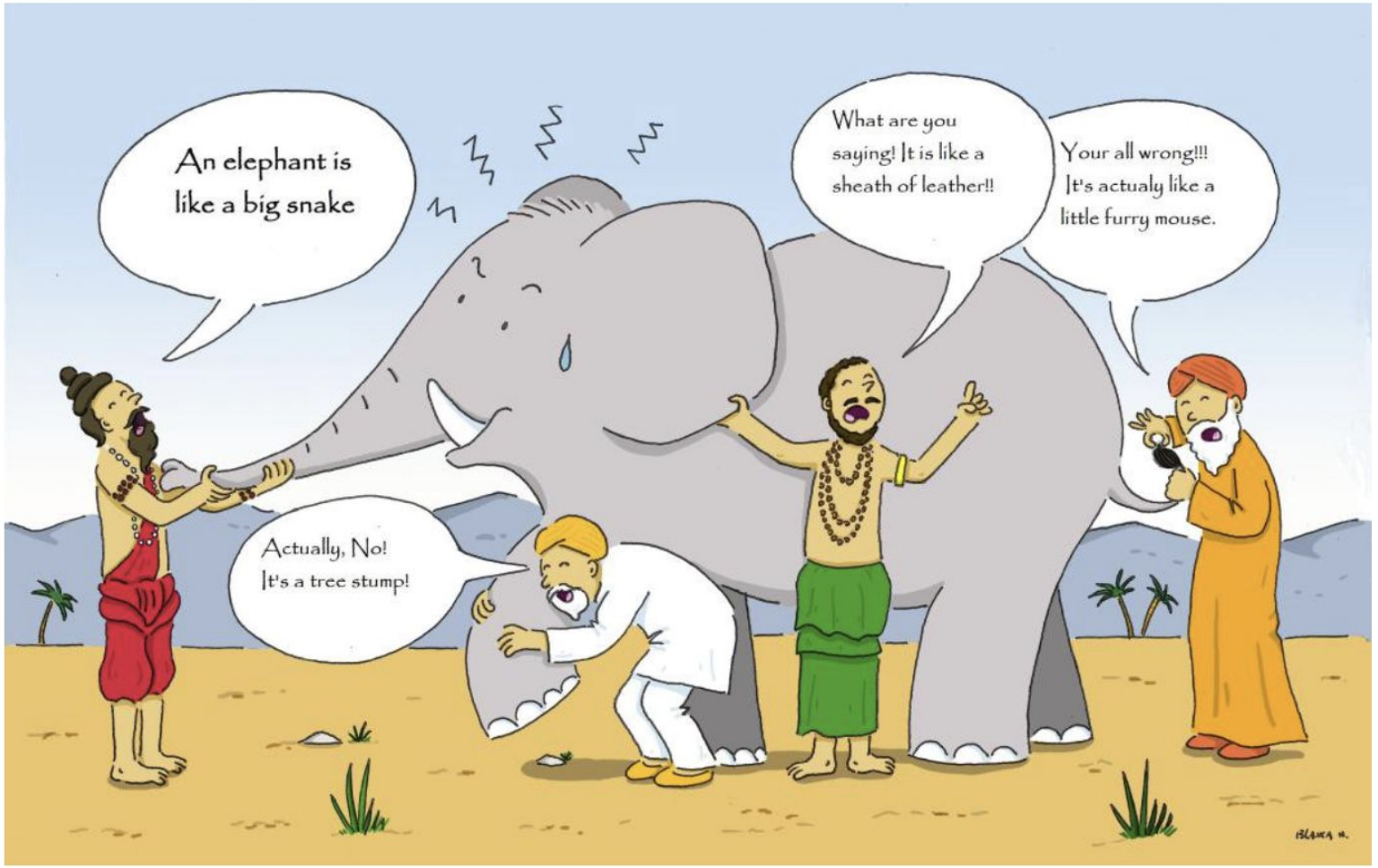
¿Alguna consulta?

FUNDACIÓN YPF

Random Forest

FUNDACIÓN YPF

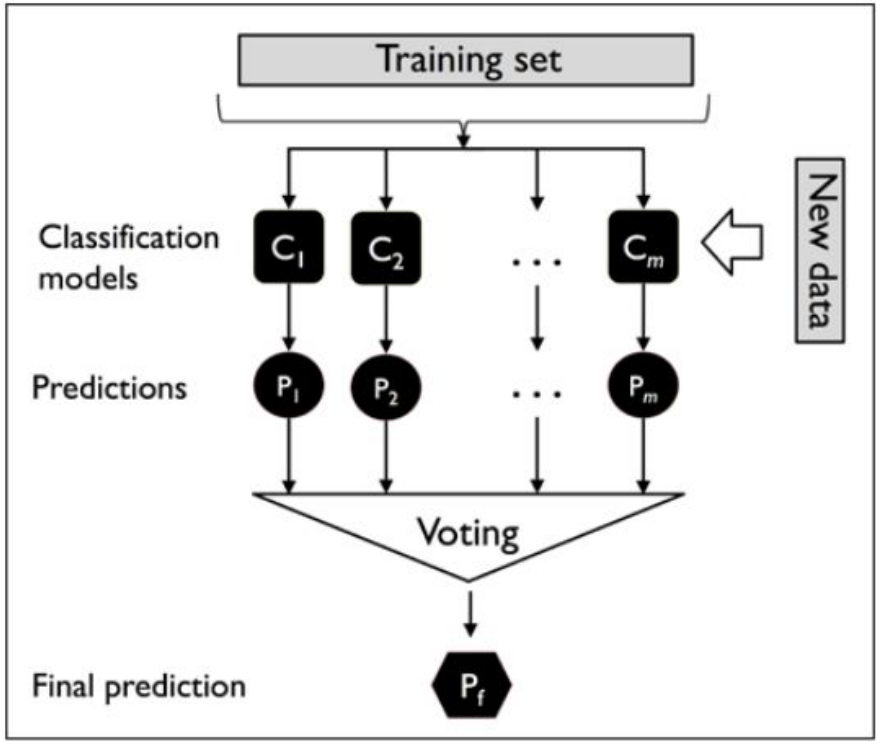
Random Forest



FUNDACIÓN YPF

Random Forest

- Entrenar muchos modelos y cada uno tendrá un resultado.
- La clasificación o valor resultante es el resultado más frecuente.



FUNDACIÓN YPF

Random Forest

- Si todos los modelos son muy parecidos, no agregan mucha información nueva en la votación final.
- Se necesitan **modelos diferentes entre sí, poco correlacionados.**

- Regresión y Clasificación
- **Ventajas**
  - Robusto frente a outliers y ruido
  - Buenos estimadores de error e importancia de variables

FUNDACIÓN YPF

K-Nearest Neighbors (kNN)

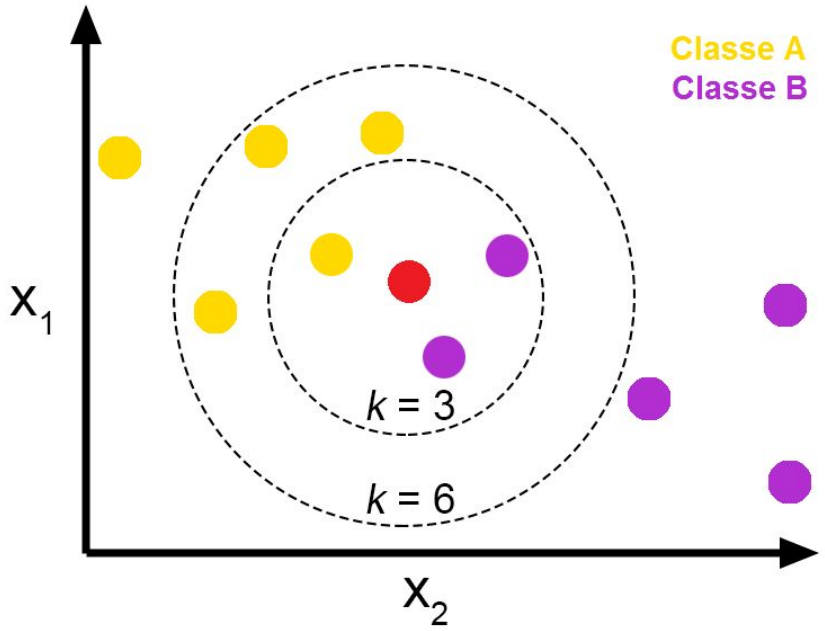
FUNDACIÓN YPF

K Nearest Neighbors (kNN)

El método de **K Nearest Neighbors (kNN)** consiste en etiquetar muestras en función de la etiqueta que tengan los **K** vecinos más cercanos

FUNDACIÓN YPF

K Nearest Neighbors (kNN)



- Calcula la **distancia** a cada uno de los puntos existentes.
- Selecciona la etiqueta que más frecuente aparece en las K clases

FUNDACIÓN YPF

Support kNN Machines

- Clasificación y Regresión
- **Ventajas**
  - Rápido porque no tiene nada que ajustar, solo almacena los datos
  - Es fácil de entender sobre todo en clasificación
- **Desventajas**
  - La elección del número de vecinos no es trivial
  - No tiene buen desempeño en datasets con desbalance de clases
  - Requiere homogeneizar features

FUNDACIÓN YPF

Descanso

Nos vemos en 10 minutos

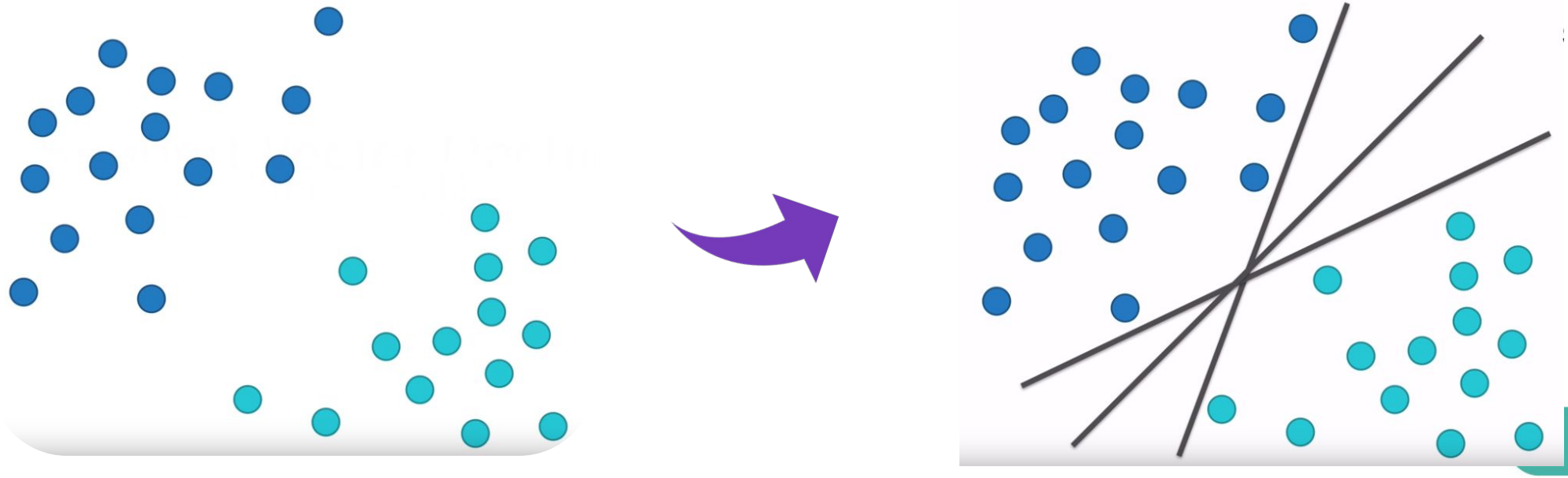
FUNDACIÓN YPF

Support Vector Machines

FUNDACIÓN YPF

Support Vector Machines

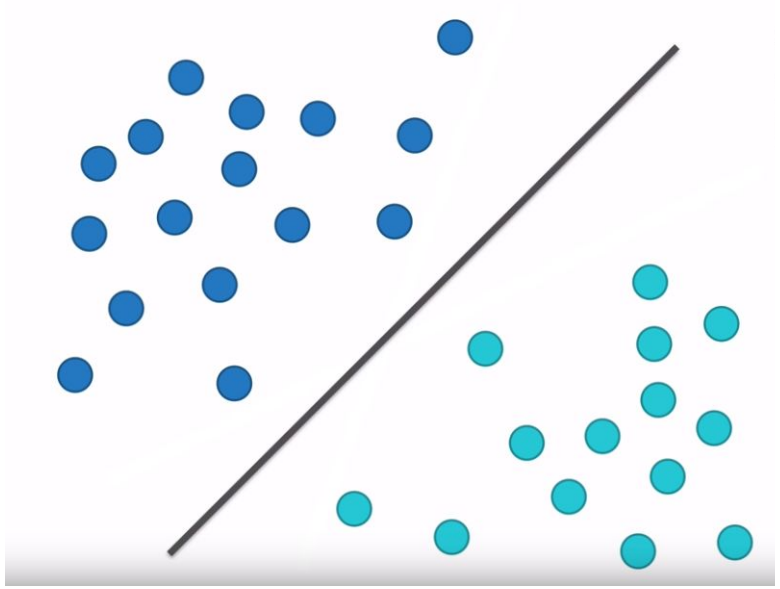
Tarea: Separar los puntos azules de los celestes



FUNDACIÓN YPF

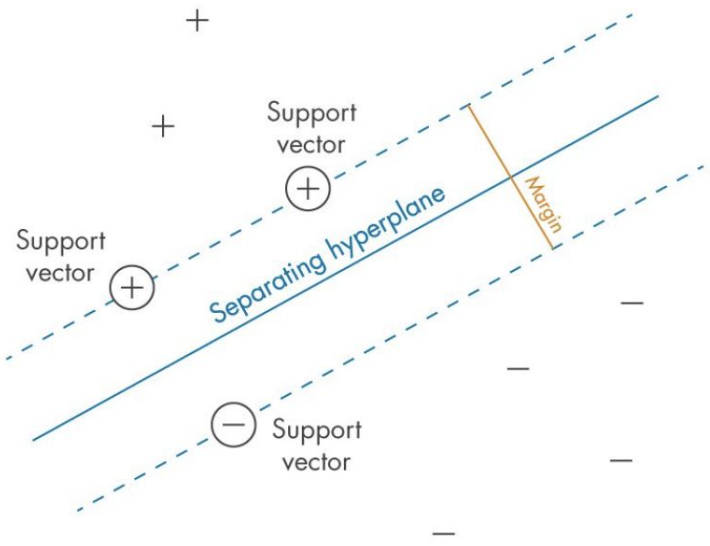
Support Vector Machines

- SVM elige la recta que separa los puntos de la “**mejor manera posible**”



FUNDACIÓN YPF

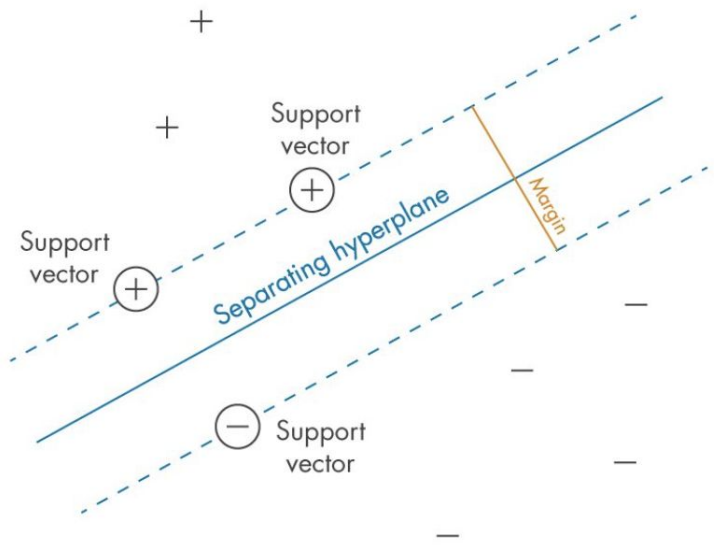
Support Vector Machines



- Establece un hiperplano que separa los puntos **maximizando** el margen.
  - La distancia del margen al vector de soporte es la mayor que se puede encontrar.
  - Es el mayor margen que se puede encontrar.

FUNDACIÓN YPF

Support Vector Machines



Margen: Es la distancia de las instancias más cercanas a la recta de decisión.

Vectores de soporte: Las instancias más cercanas al margen. Es una de las pocas características de los datos que le importa al algoritmo.

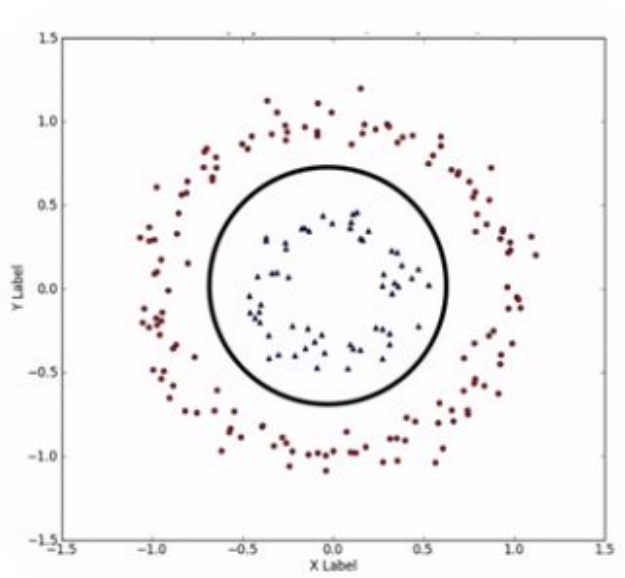
Lo que hace el algoritmo es encontrar la recta que maximice el margen.

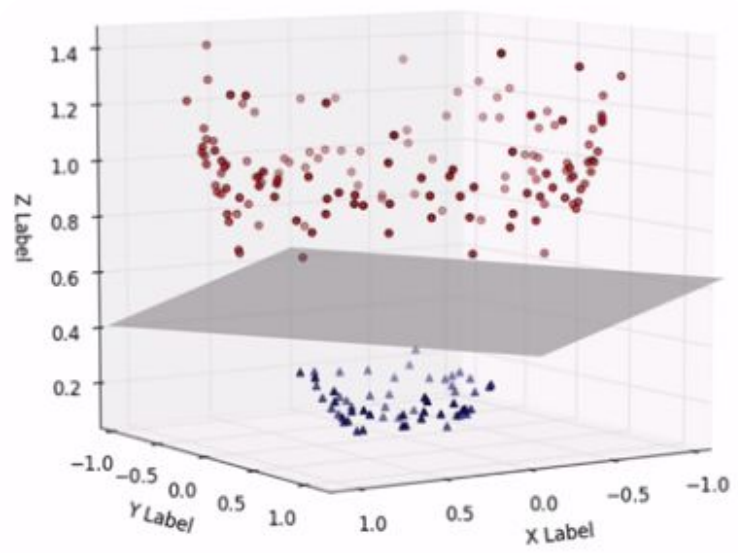
Es un problema de optimización, lo que conlleva a una solución computacionalmente eficiente.

FUNDACIÓN YPF

Support Vector Machines

Kernels: Se utilizan cuando los datos directamente no se pueden separar con una línea





FUNDACIÓN YPF

Support Vector Machines

- Clasificación y Regresión
- **Ventajas**:
  - Eficaz en espacios de alta dimensión
  - Eficiente en memoria
  - El uso de kernels lo convierten en muy versátil
- **Desventajas**:
  - El uso de kernels tiende a llevar al algoritmo a overfitting
  - Su mejor desempeño se da para problemas de clasificación

FUNDACIÓN YPF

Sección práctica:

Trabajamos con la Notebook 15
ajustando los primeros modelos

FUNDACIÓN YPF

En la sala general
Aprendizaje Supervisado - Parte 2

Trabajamos con la Notebook 15

Demostraremos cómo ajustar los modelos de regresión y clasificación vistos

FUNDACIÓN YPF

Desafío 11

- Para la siguiente clase, repasar y ejercitar la notebook 15.

FUNDACIÓN YPF

¿Alguna consulta?

FUNDACIÓN YPF

¡Muchas gracias!