



Machine Learning

Unidad # 3 - Aprendizaje Supervisado Avanzado y Aprendizaje No Supervisado

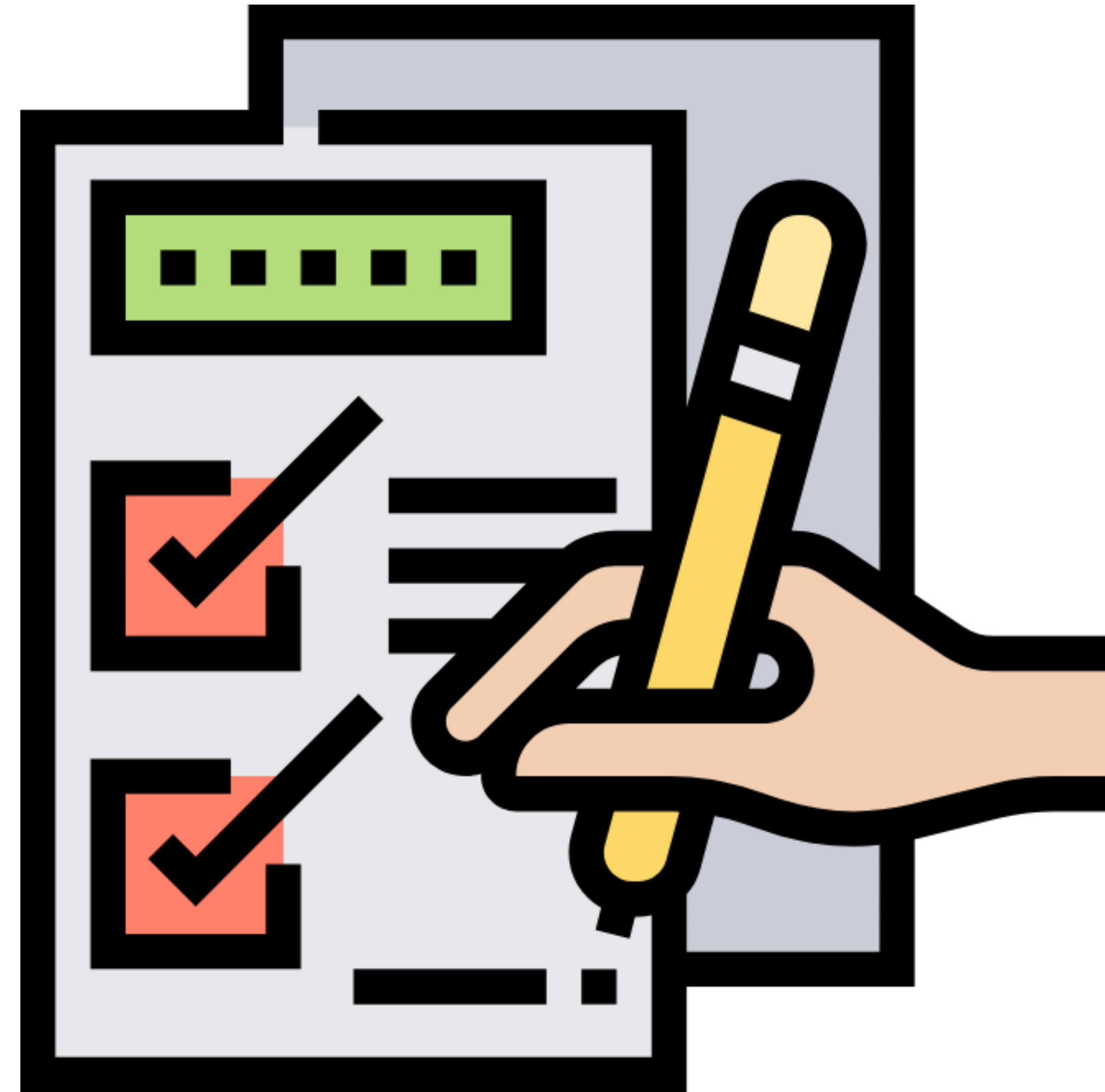
CC57 – 2019-1

Profesor
Andrés Melgar



Competencias a adquirir en la sesión

- Al finalizar la sesión el alumno comprenderá el funcionamiento del **aprendizaje inductivo**.
- Al finalizar la sesión el alumno implementará **modelos algoritmos de regresión** usando conjuntos de datos.
- Al finalizar la sesión el alumno **entenderá** el algoritmo de **regresión lineal**.
- Al finalizar la sesión el alumno **aplicará** el algoritmo de **regresión lineal** para obtener modelos algorítmicos.





Métricas de Evaluación

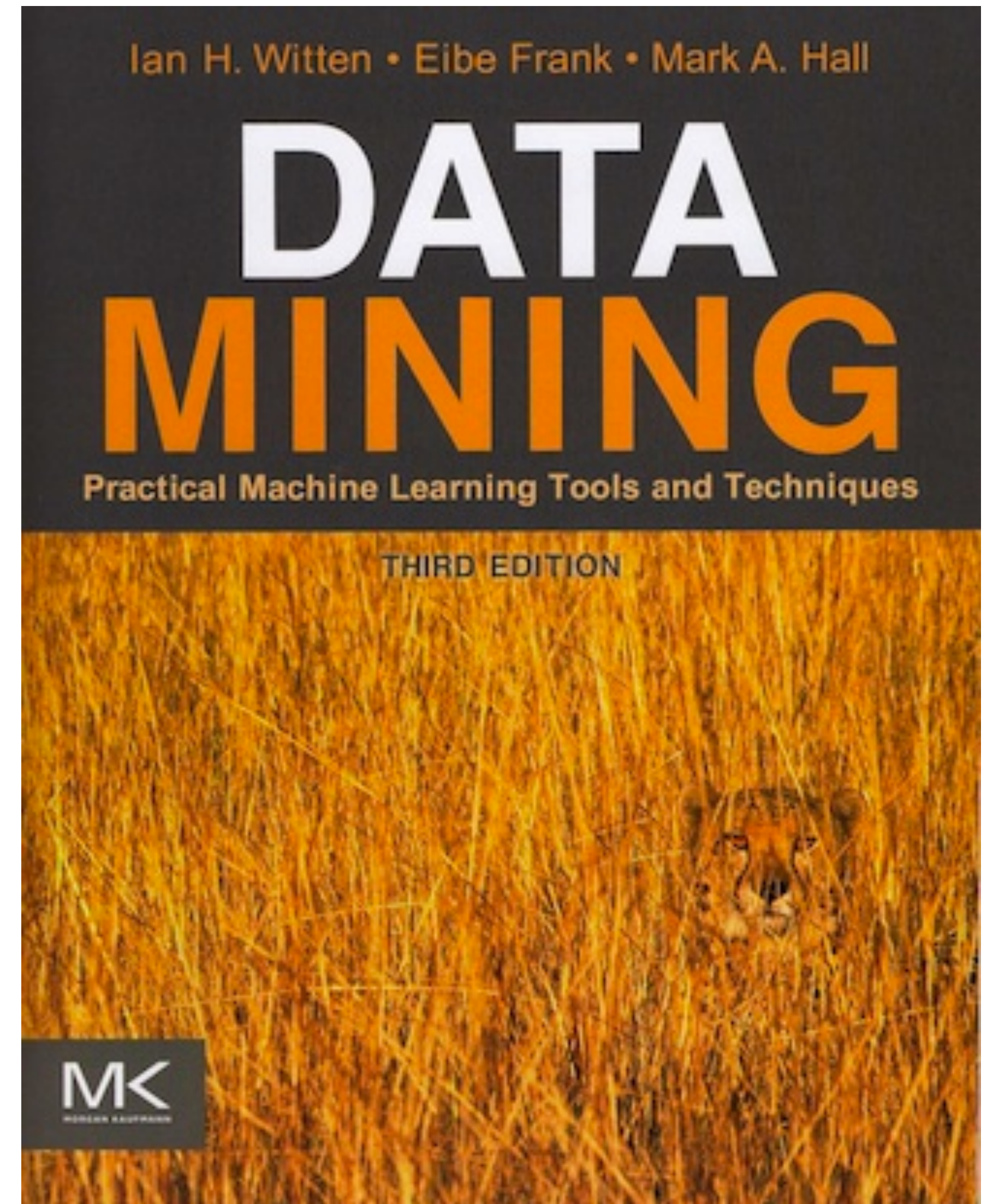
Texto guía

Witten, Ian H., Frank, Eibe, and Hall, Mark A.. 2011. *Data Mining : Practical Machine Learning Tools and Techniques with Java Implementations*. San Francisco: Elsevier Science & Technology.

| | CHAPTER |
|-------------------------------|---------|
| Algorithms: The Basic Methods | 4 |

4.6 LINEAR MODELS

4.6 LINEAR MODELS





Modelos lineales

- Es un modelo de **representación de conocimiento** en aprendizaje de máquina
 - Este modelo se caracteriza por que la **salida** está representada por la **suma de sus atributos**.
 - Generalmente **pesos** son aplicados a los atributos antes de sumarlos
- El **objetivo** del aprendizaje de máquina se centra en obtener buenos valores para los **pesos**
 - Unos valores que permitan que la **salida real** del modelo coincida con la **salida deseada** (corpus supervisado)
 - En este caso, los atributos de entrada y las salidas son todos valores **numéricos**.



Modelos lineales

- Los modelos lineales son muy fáciles de **visualizar** en dos dimensiones
- Es equivalente a trazar una **línea recta** a través de un conjunto de puntos de datos
- ¿Cómo se visualizaría un modelo en **3 dimensiones**?
- ¿Cómo se visualizaría un modelo en **n-dimensiones**?

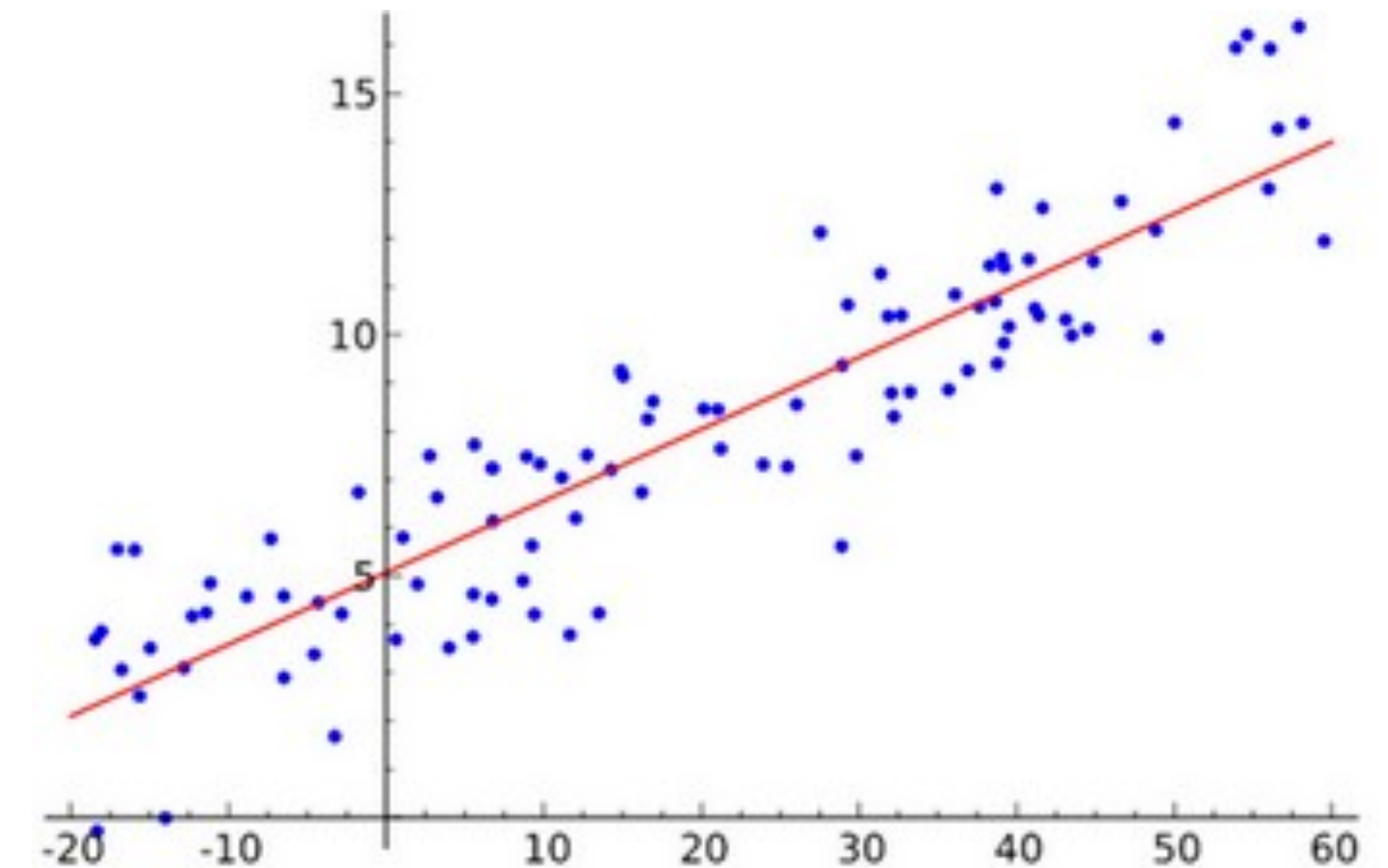
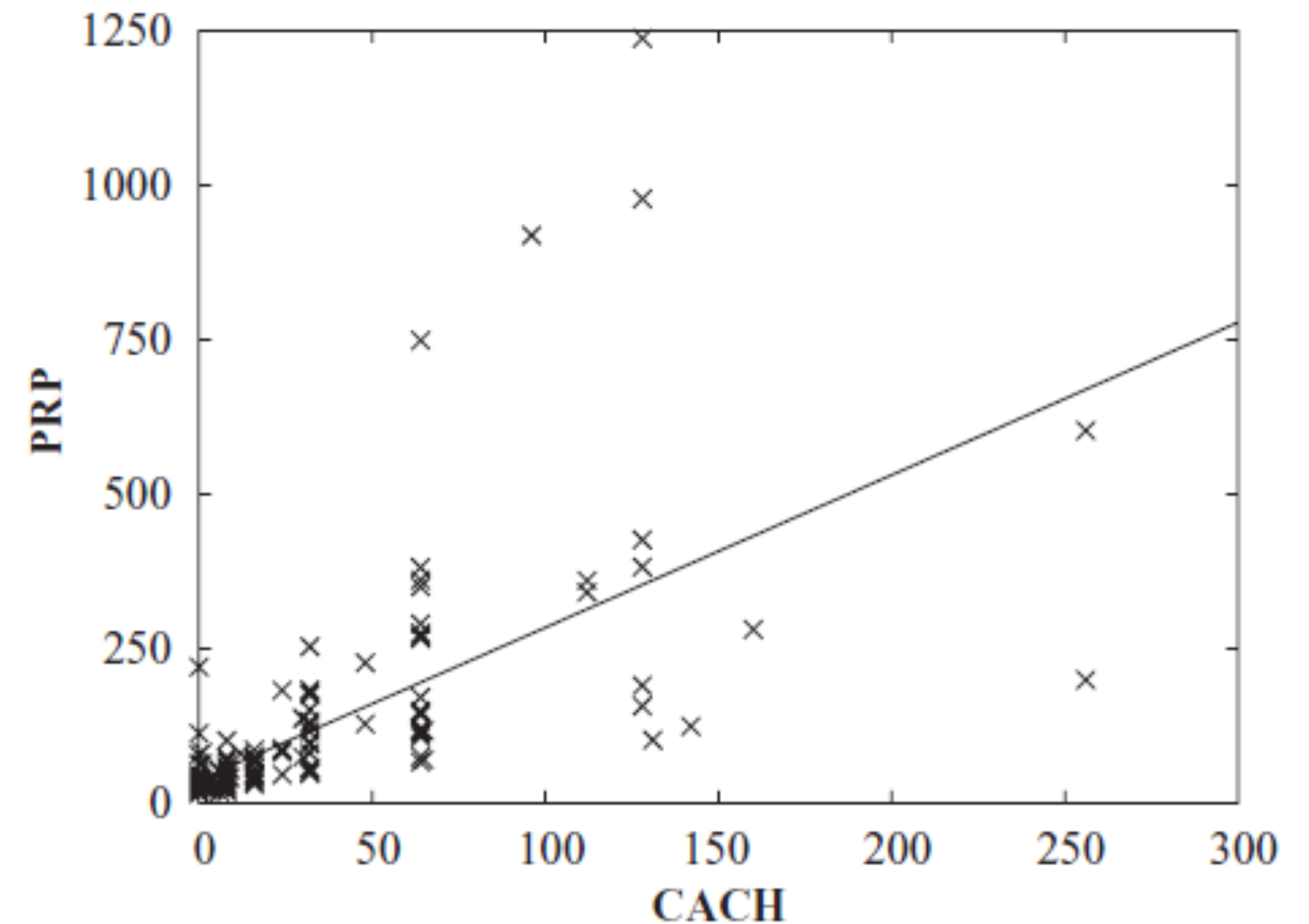


Imagen tomada de https://gerardnico.com/data_mining/linear_regression



Modelos lineales

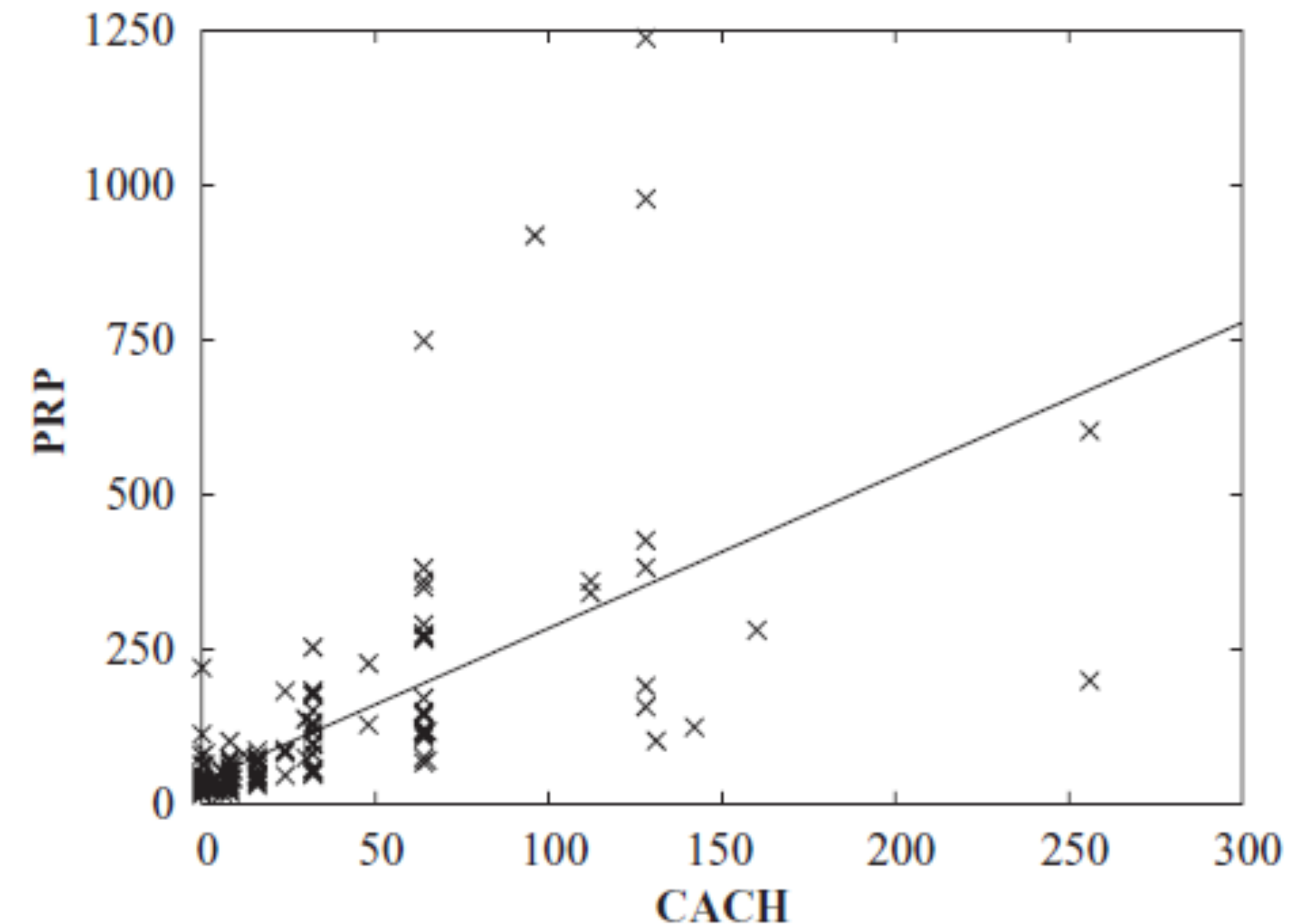
| No. | 1: MYCT | 2: MMIN | 3: MMAX | 4: CACH | 5: CHMIN | 6: CHMAX | 7: class |
|-----|---------|---------|---------|---------|----------|----------|----------|
| | Numeric | Numeric | Numeric | Numeric | Numeric | Numeric | Numeric |
| 1 | 125.0 | 256.0 | 6000.0 | 256.0 | 16.0 | 128.0 | 198.0 |
| 2 | 29.0 | 800... | 3200... | 32.0 | 8.0 | 32.0 | 269.0 |
| 3 | 29.0 | 800... | 3200... | 32.0 | 8.0 | 32.0 | 220.0 |
| 4 | 29.0 | 800... | 3200... | 32.0 | 8.0 | 32.0 | 172.0 |
| 5 | 29.0 | 800... | 1600... | 32.0 | 8.0 | 16.0 | 132.0 |
| 6 | 26.0 | 800... | 3200... | 64.0 | 8.0 | 32.0 | 318.0 |
| 7 | 23.0 | 160... | 3200... | 64.0 | 16.0 | 32.0 | 367.0 |
| 8 | 23.0 | 160... | 3200... | 64.0 | 16.0 | 32.0 | 489.0 |
| 9 | 23.0 | 160... | 6400... | 64.0 | 16.0 | 32.0 | 636.0 |





Modelos lineales

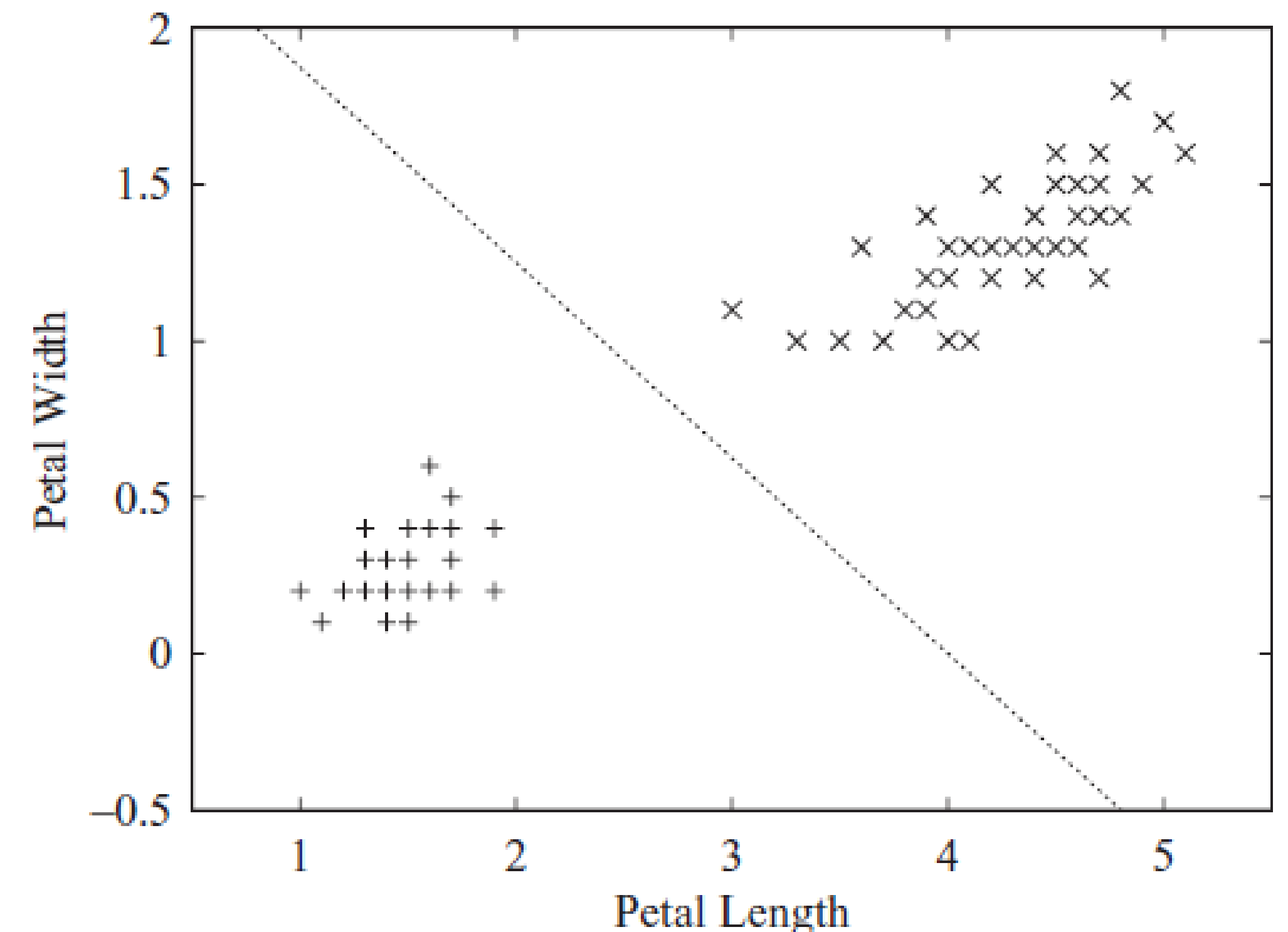
- La figura anterior muestra una **línea ajustada** a los datos de rendimiento del CPU.
 - Se utiliza sólo el atributo de caché como entrada.
 - El rendimiento del CPU se muestra en el eje vertical.
 - La memoria caché en el eje horizontal.
 - Ambos atributos son **numéricos**.
 - La línea recta representa la **ecuación** de predicción es $PRP = 37.06 + 2.47 \text{ CACH}$
- Usando la fórmula es posible realizar **predicción** con datos de nuevas instancias.





Modelos lineales

- Los modelos lineales pueden ser usados también para problemas de **clasificación**
- En este caso, la línea producida por este modelo se utiliza para **separar** a las clases.
- Permite decidir si es que una instancia **pertenece** o no a una clase (¿Está encima o debajo de la línea?).
- A estas líneas se les suele denominar la **frontera de decisión**.





Regresión Lineal

- Cuando la **salida** es **numérica** y **todos los atributos son numéricos** la regresión lineal es la técnica natural a considerar.
- Este es un método basado en la estadística. La idea es expresar la clase como una **combinación lineal** de los atributos, con pesos predeterminados:

$$x = w_0 + w_1a_1 + w_2a_2 + \cdots + w_ka_k$$

- Donde:
 - X es el valor de la clase a predecir.
 - a_1, a_2, \dots, a_k son los atributos.
 - w_0, w_1, \dots, w_k son los pesos.



Regresión Lineal

- La regresión lineal es un método excelente y simple para la **predicción numérica**.
- Ha sido ampliamente utilizado en aplicaciones **estadísticas** durante décadas.
- Pero tienen la desventaja de la **linealidad**.
- Si los datos exhibe una dependencia no lineal, la predicción se realizará tomando como base la **diferencia de los cuadrados**, lo cual podría no ser la mejor solución.



Regresión Lineal

- Los **pesos** son calculados a partir de los datos de entrenamiento.
- Para predecir el **valor de la clase**, se parte de la siguiente fórmula:

$$x = w_0 + w_1a_1 + w_2a_2 + \cdots + w_k a_k$$

- El algoritmo centra su interés en la **diferencia** del valor real (corpus) y del valor predicho (modelo).
- Se escogen los pesos de forma tal que se **minimicen los cuadrados de las diferencias** entre el valor real y el valor predicho.



Regresión Lineal

- Dada n instancias, el cuadrado de la diferencia sería:

$$\sum_{i=1}^n (x_i - \sum_{j=0}^k w_j \times a_j)^2$$

- Para obtener los pesos, se tiene que solucionar un problema matemático de minimización.



Competencias a adquirir en la sesión

- Al finalizar la sesión el alumno comprenderá el funcionamiento del **aprendizaje inductivo**.
- Al finalizar la sesión el alumno implementará **modelos algoritmos de regresión** usando conjuntos de datos.
- Al finalizar la sesión el alumno **entenderá** el algoritmo de **regresión lineal**.
- Al finalizar la sesión el alumno **aplicará** el algoritmo de **regresión lineal** para obtener modelos algorítmicos.

