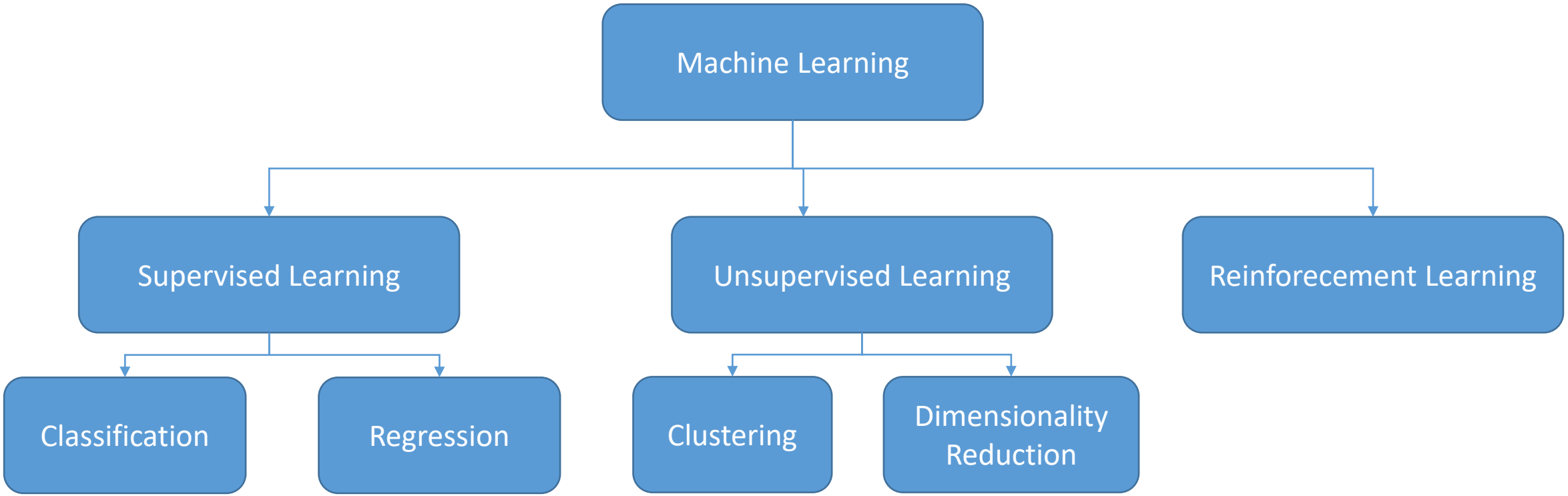
Análisis de regresión

Dr. Raimundo Sánchez

raimundo.sanchez@uai.cl

@raimun2

Tipos de aprendizaje

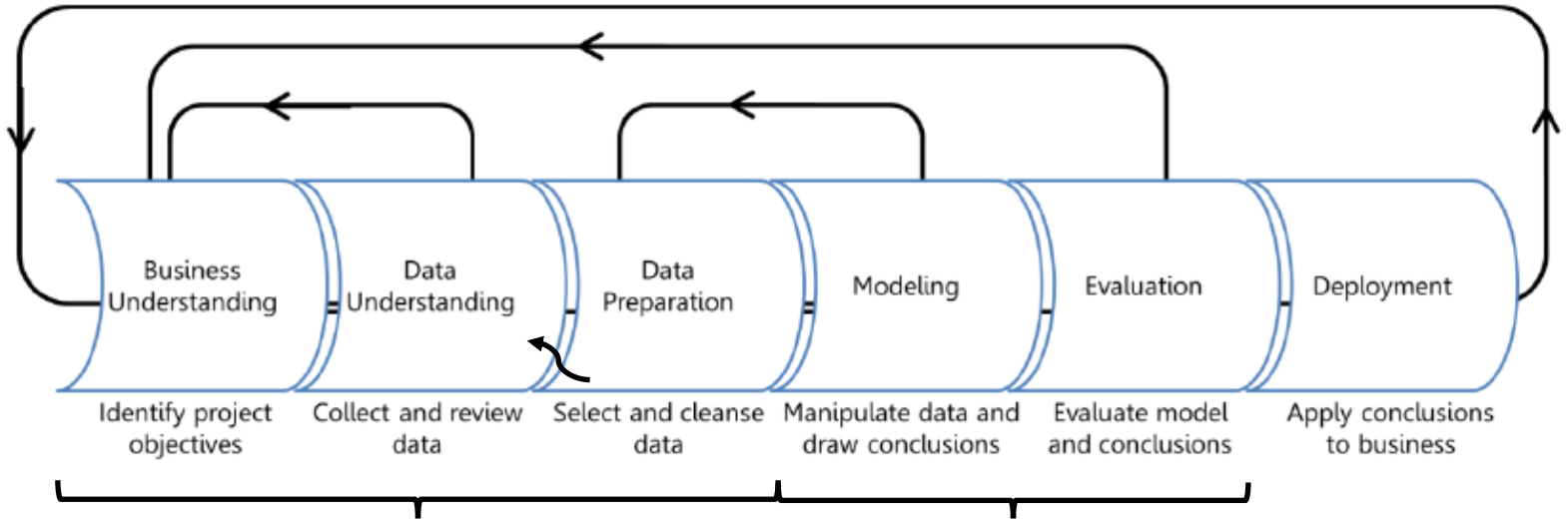


$$\hat{Y} = f(X, Y)$$

$$\hat{Z} = f(X)$$

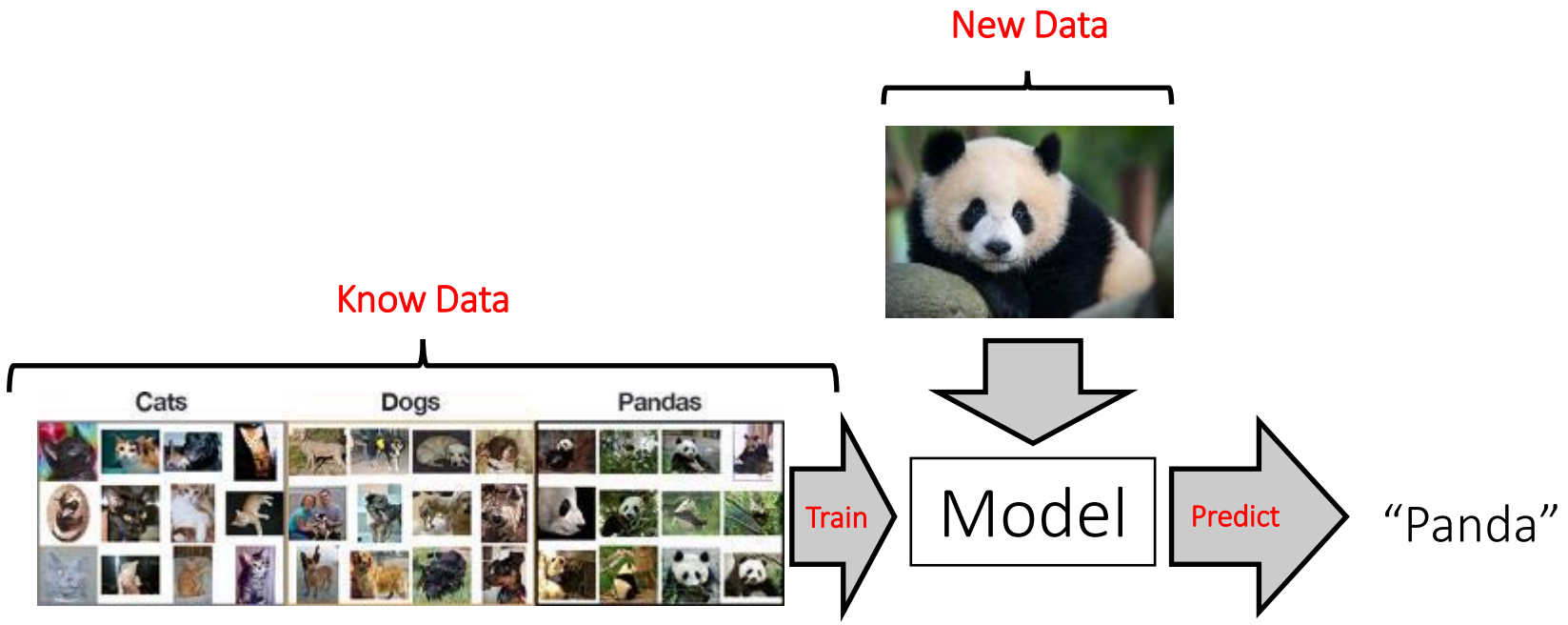
$$\hat{X}_t = f(\hat{X}_{t-n})$$

¿Arte o ciencia?



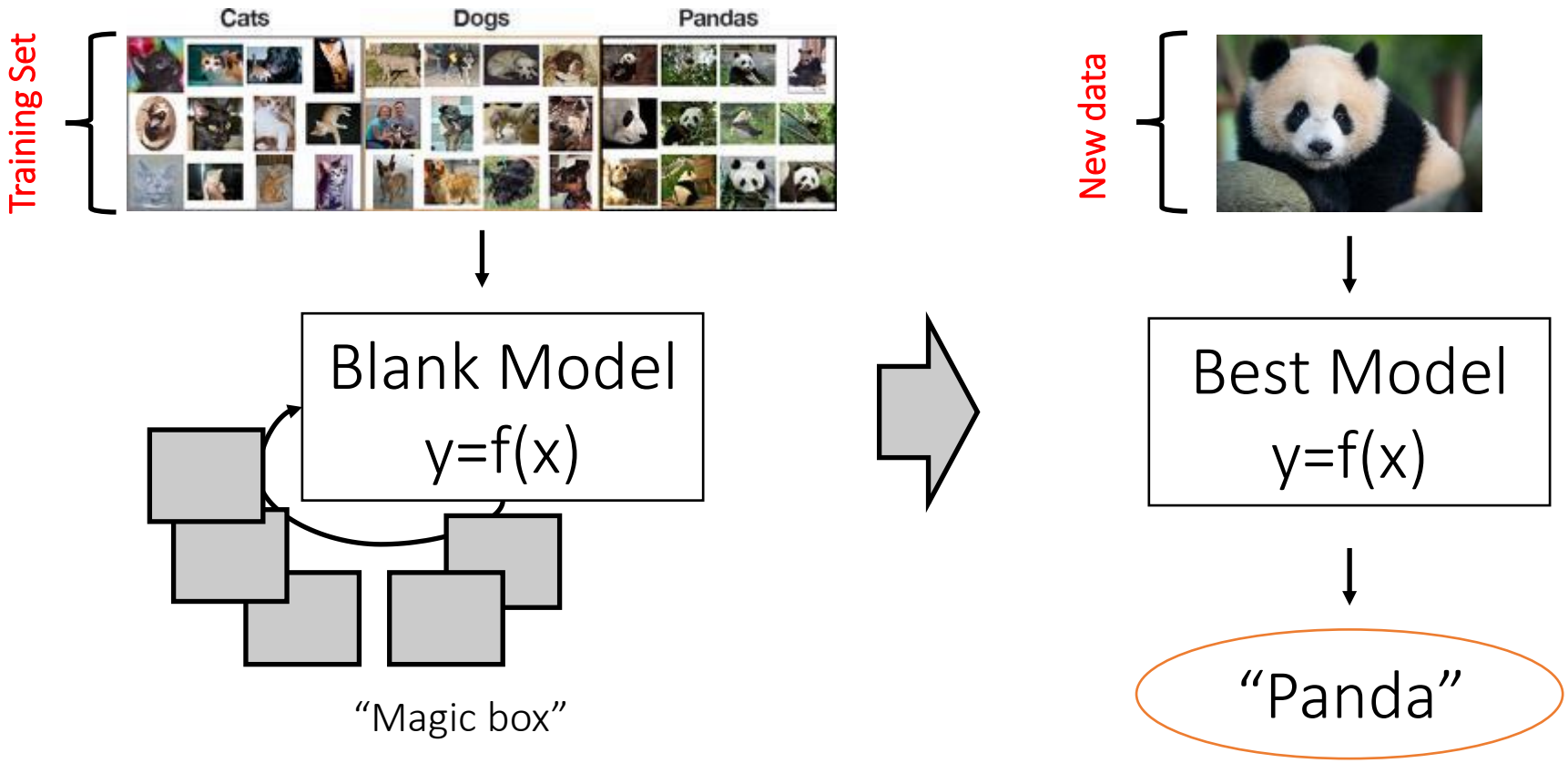
Modelamiento predictivo (aprendizaje supervisado)

Idea general.



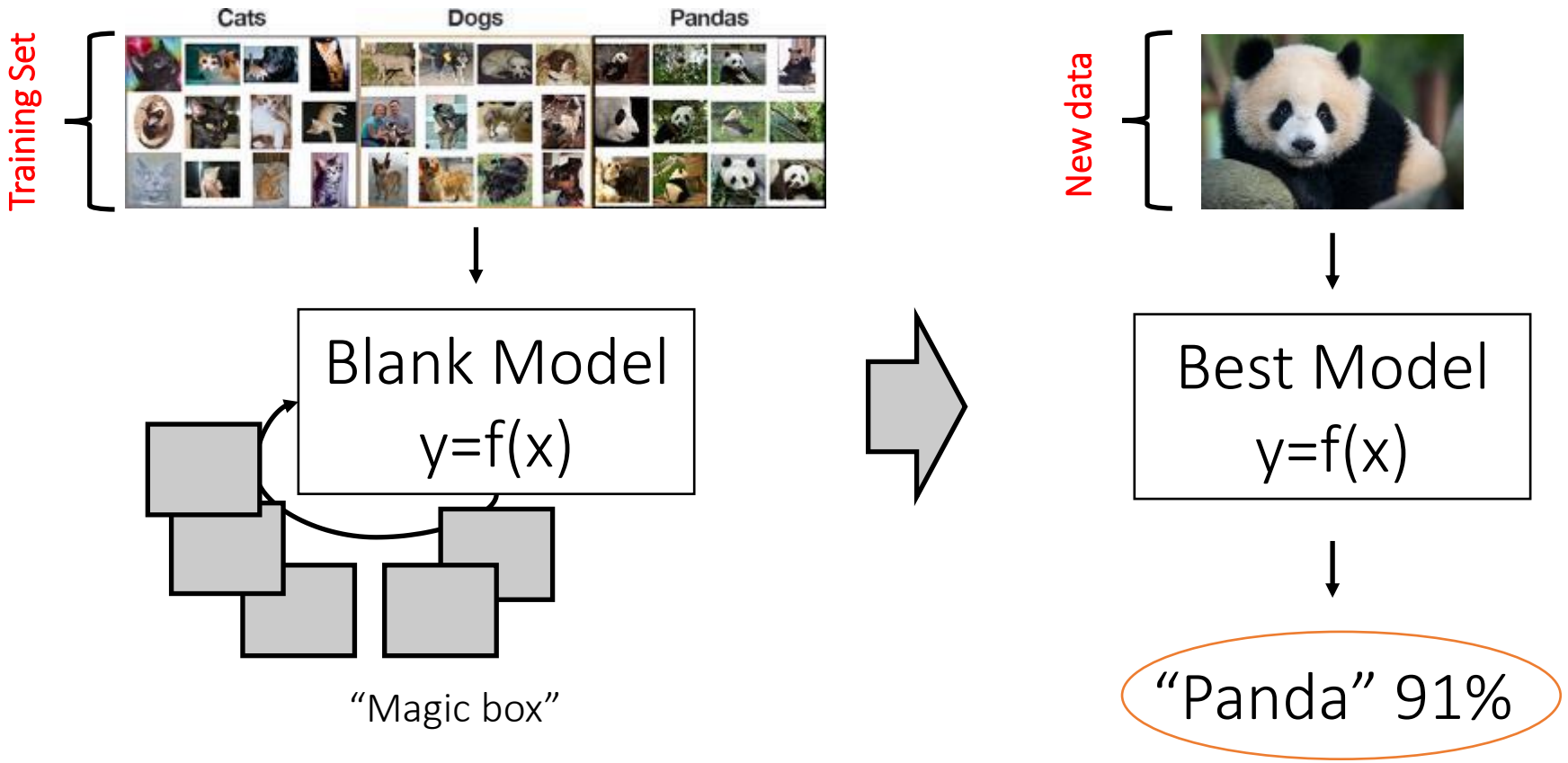
Tenemos que construir, a través de un proceso delicado, el mejor y = f(x) posible.

Enfoque determinístico



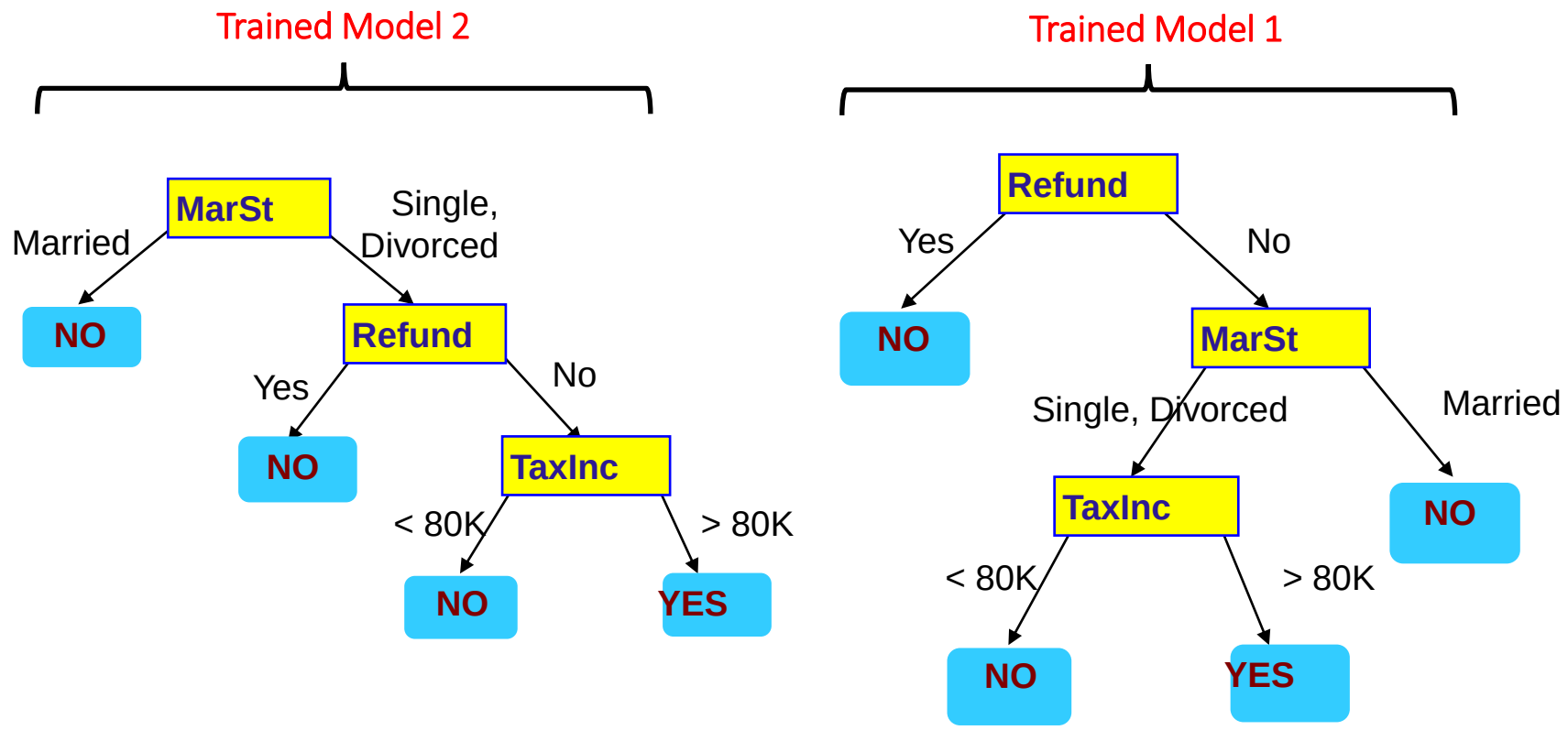
Discriminative

Enfoque probabilístico



Espacio de soluciones

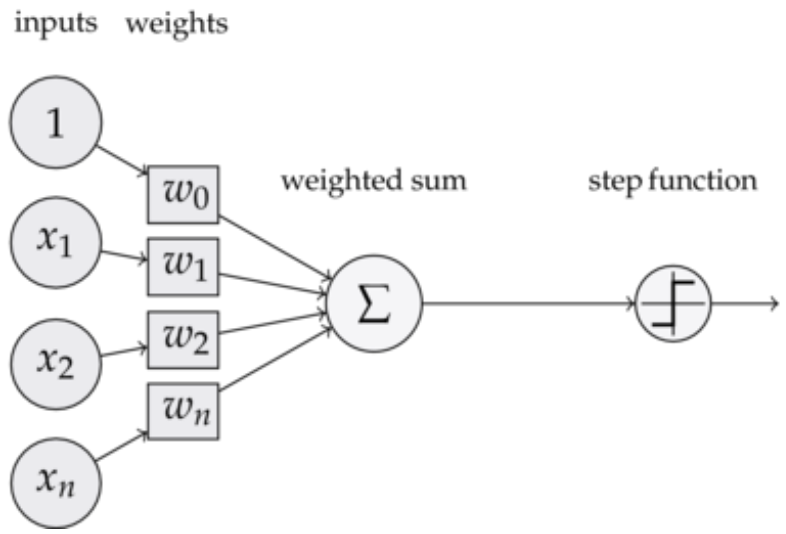
Todas las soluciones posibles



Paramétrico vs no paramétrico

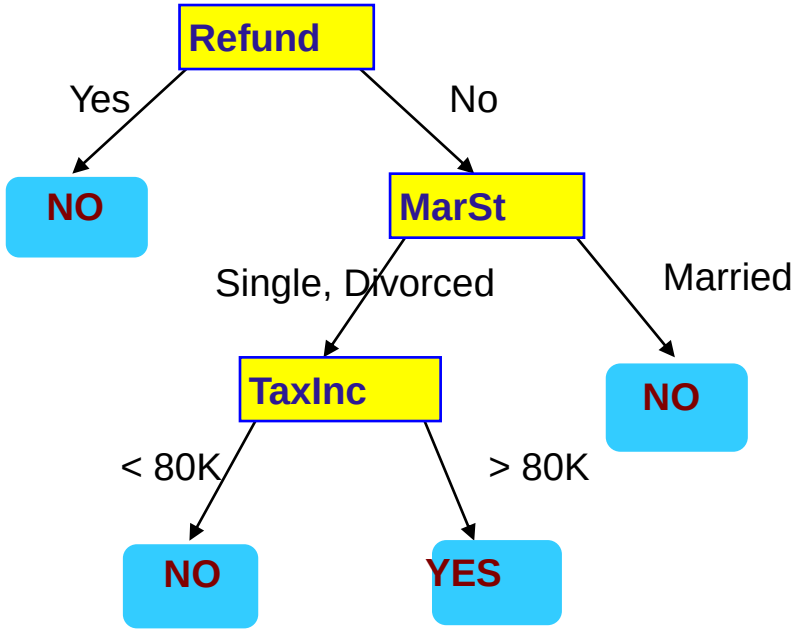
Parametric

Se asume una forma funcional particular
El número de parámetros se fija de antemano
Ejemplos: Naive Bayes, perceptron



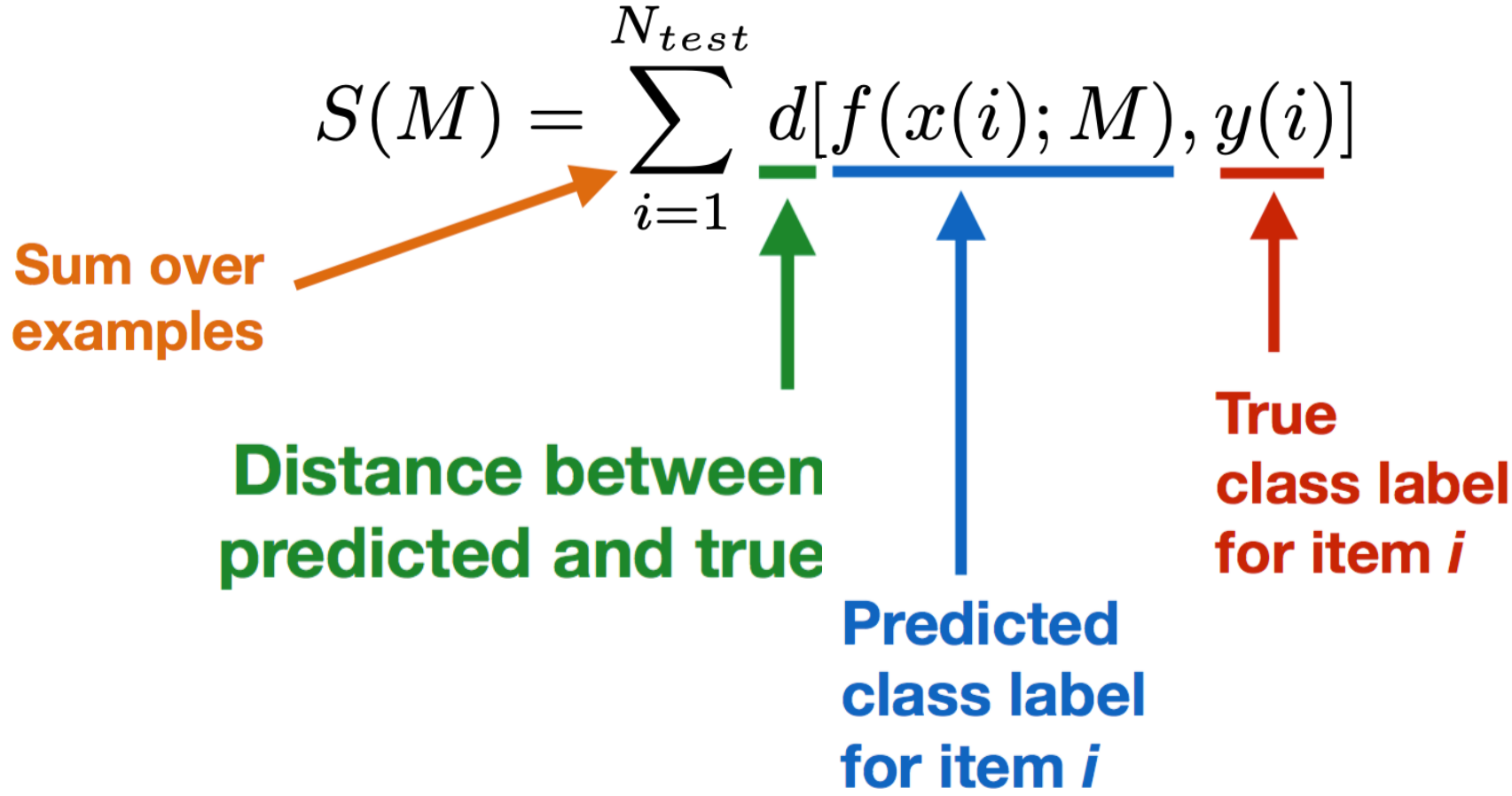
Non-Parametric

Pocas suposiciones se hacen sobre la forma funcional
La estructura del modelo se determina a partir de datos
Ejemplos: classification tree, nearest neighbor



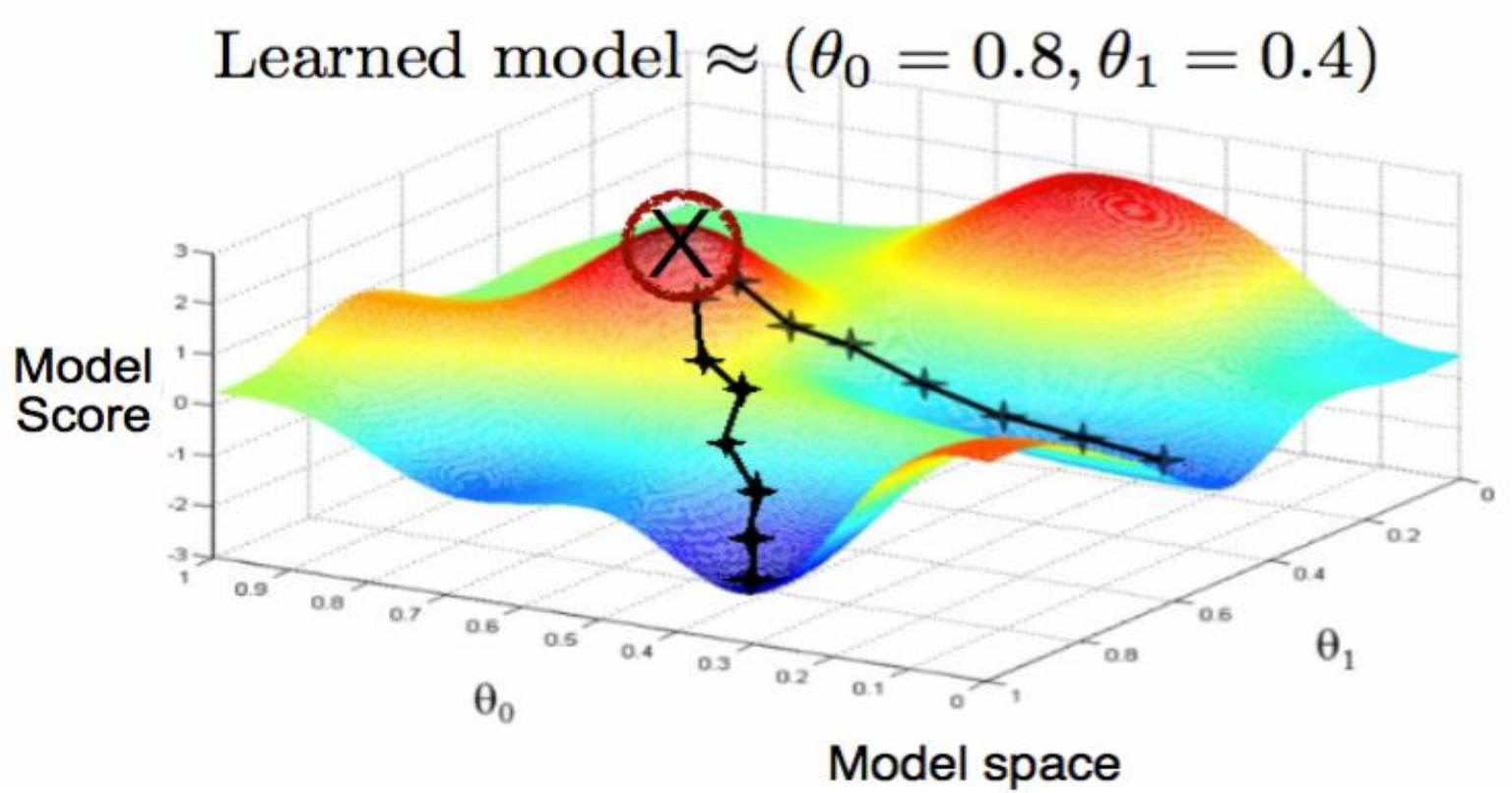
Aprendizaje

Minimizar una función de evaluación



Algoritmo de búsqueda

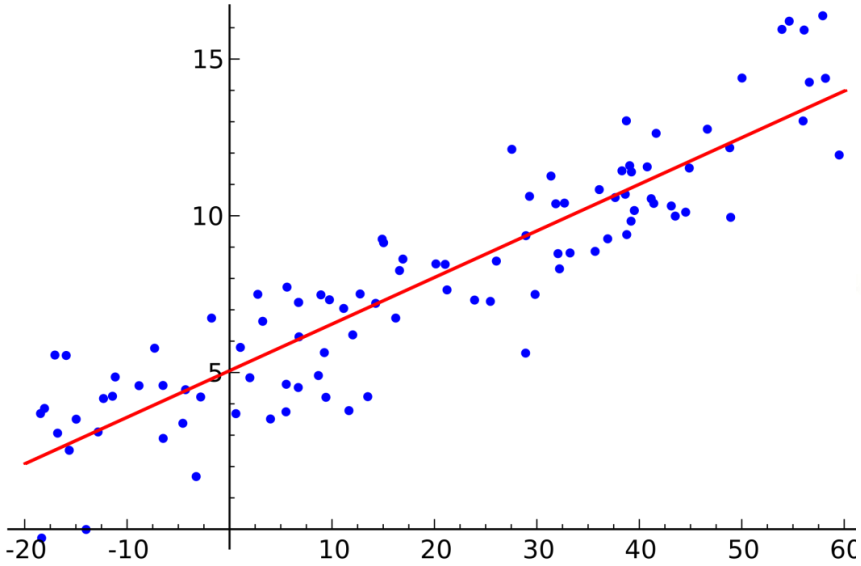
Busca el conjunto de parámetros que maximicen la puntuación del modelo.

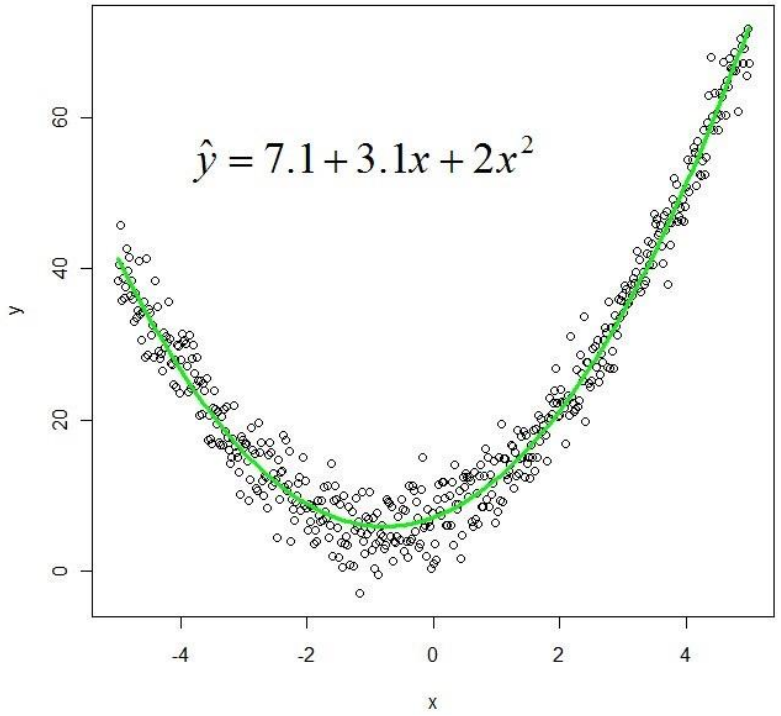


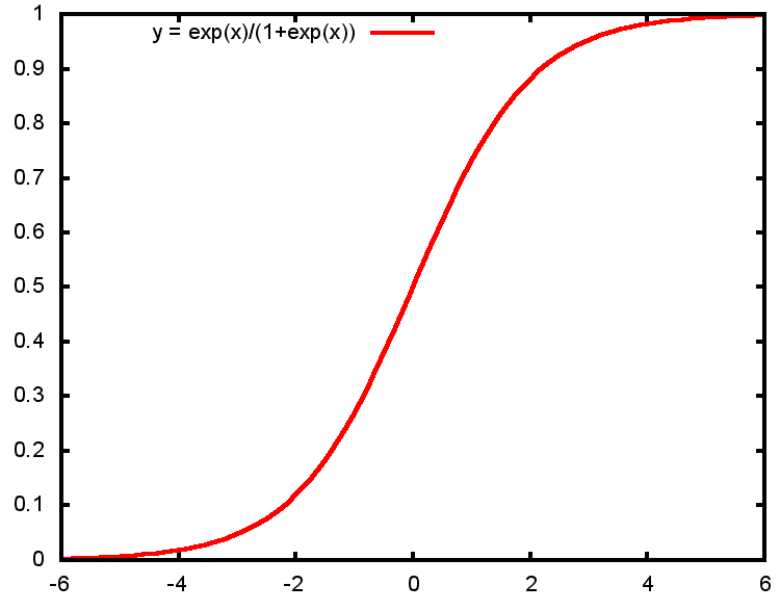
Análisis de regresión

Los modelos de regresión buscan una relación entre una variable dependiente y una o varias variables explicativas.

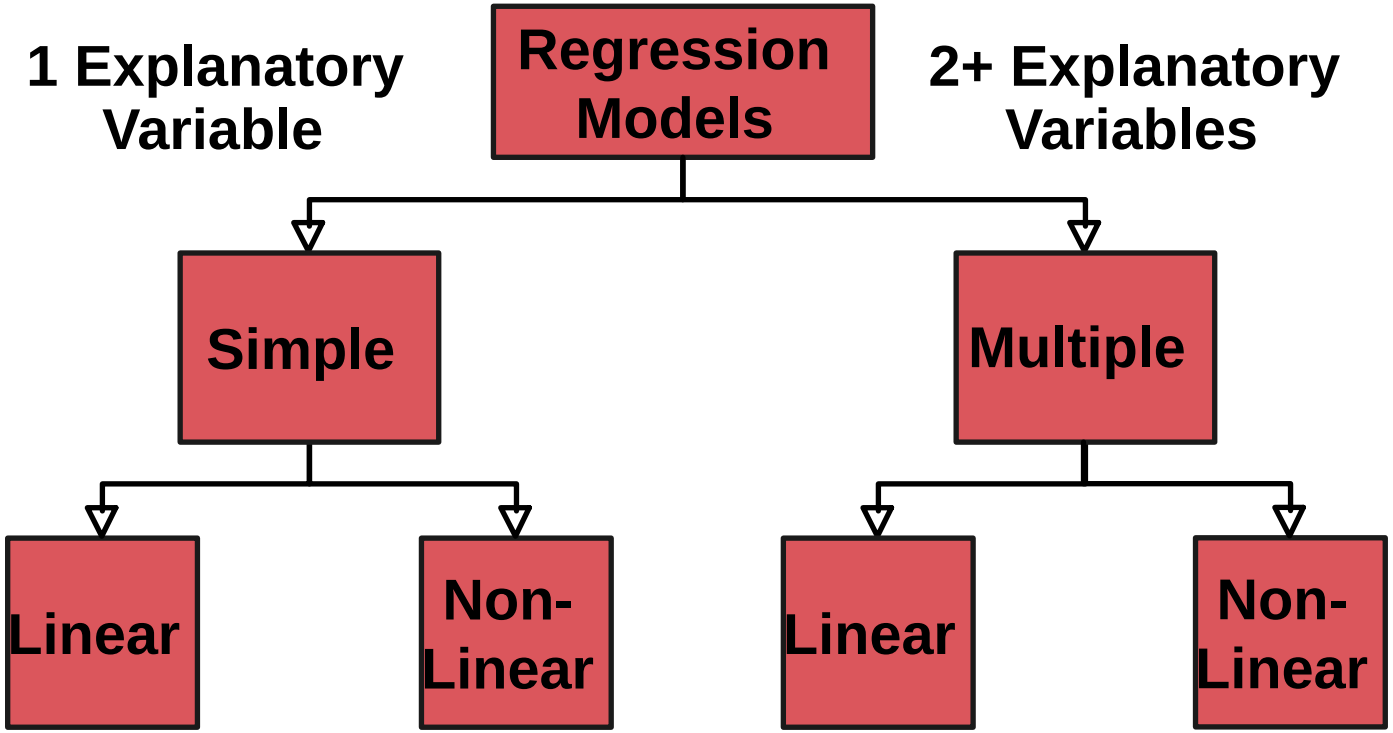
Los modelos de regresión se utilizan principalmente para predicción y estimación.







Tipos de regresión



Modelo de regresión lineal simple

En un modelo de regresión lineal, la relación entre variables es una función lineal en los parámetros

$$Y = \beta_0 + \beta_1 X + \epsilon$$

$$Y = \beta_0 + \beta_1 X^2 + \epsilon$$

$$Y = \beta_0 + \beta_1 X + \beta_1 X^2 + \cdots + \beta_7 X^7 + \epsilon$$

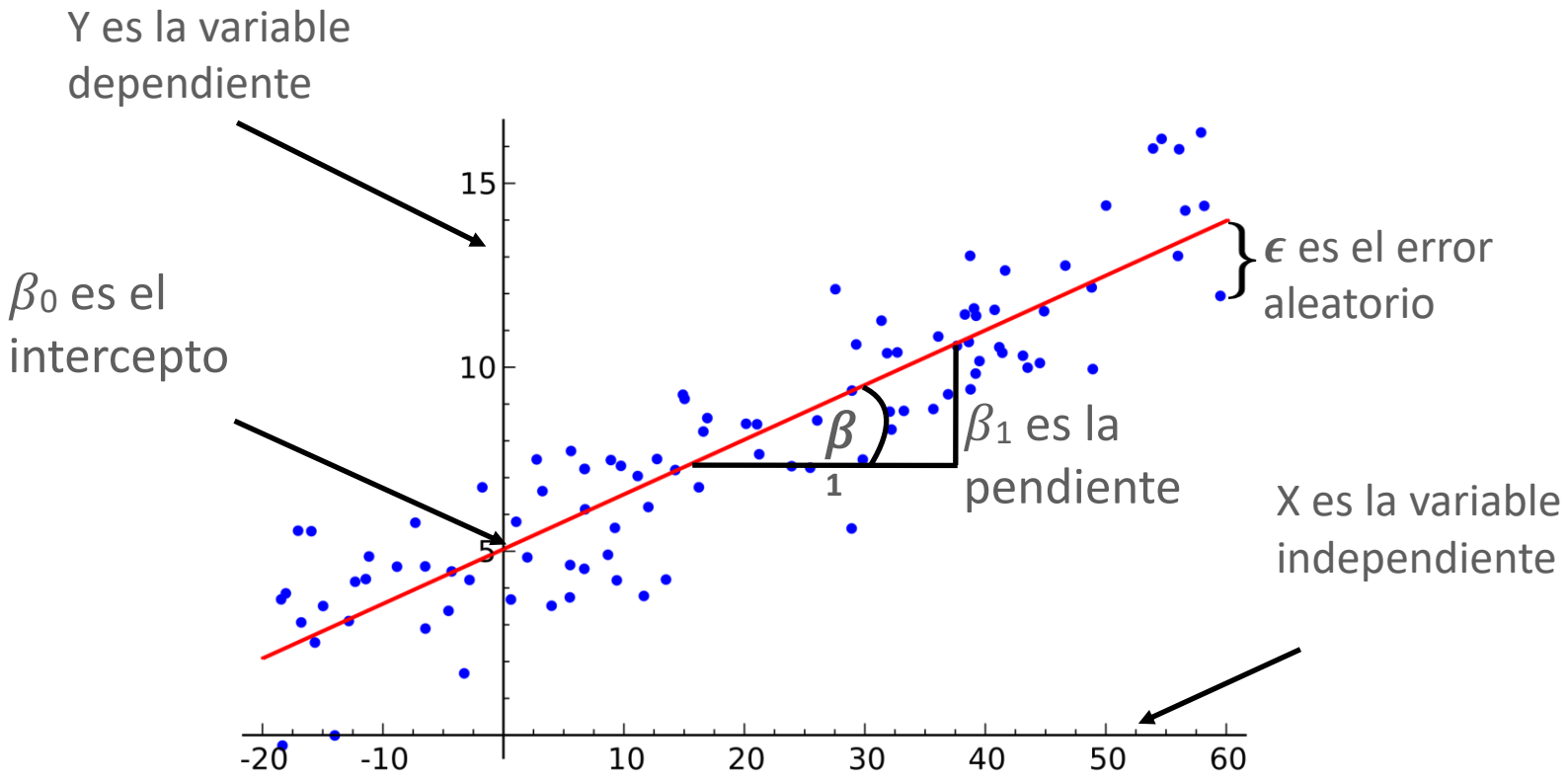
En un modelo de regresión no lineal, la función de respuesta no es lineal

$$Y = \beta_0 \exp(\beta_1 X) + \epsilon$$

Componentes

Modelo de regresión lineal simple con una variable predictora

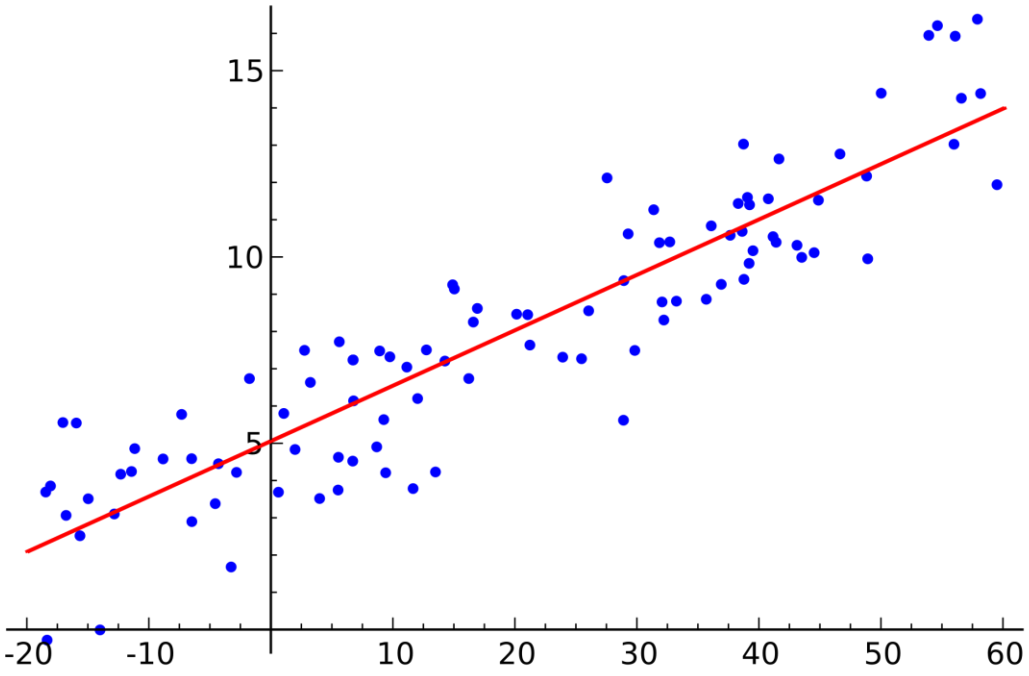
$$Y = \beta_0 + \beta_1 X + \epsilon$$



Objetivos

- Buscamos estimar los parámetros β0 y β1 que minimizan/maximizan una función de puntuación
- Con estos parámetros y xi podemos calcular ŷi, la predicción de yi

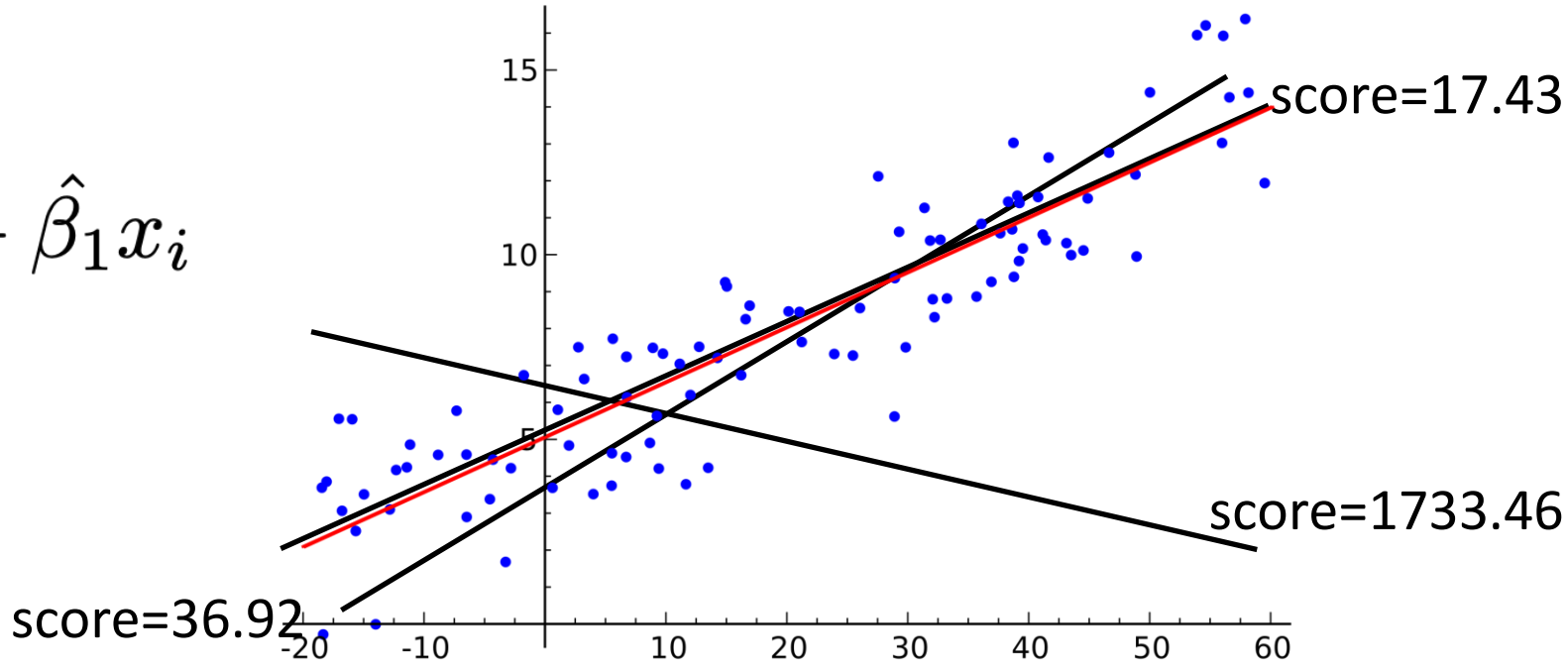
$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$



Espacio de soluciones

El espacio modelo viene determinado por todos los valores posibles de los parámetros β0 y β1

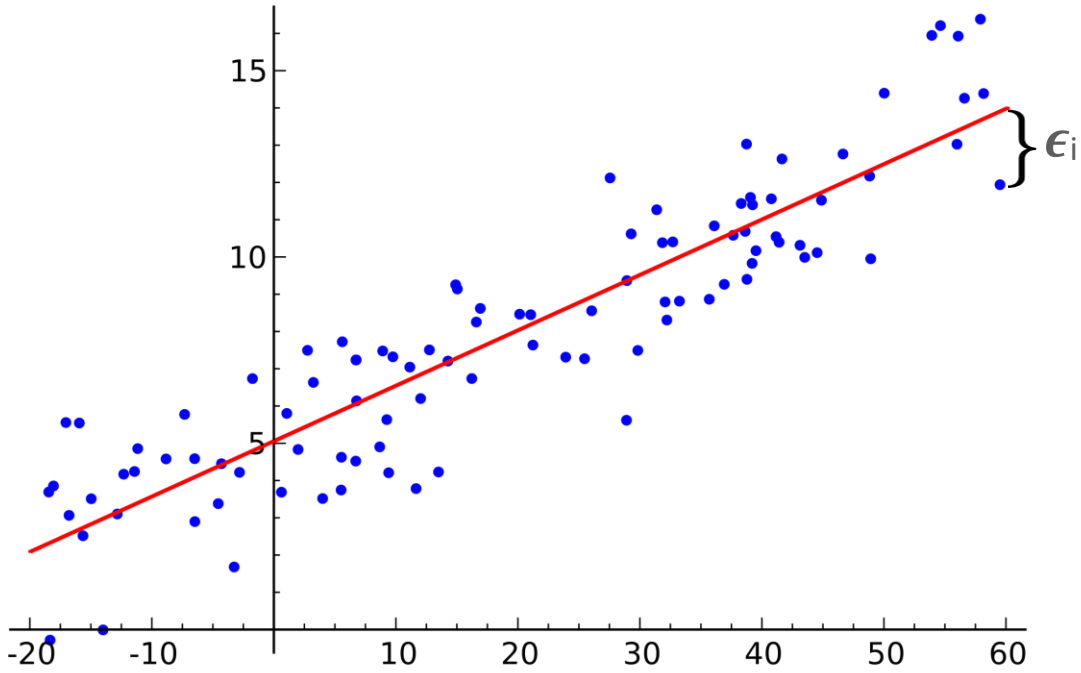
$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$



Aprendizaje

Una de las funciones de puntuación más típicas es el error cuadrático medio (MSE), pero hay diferentes funciones de puntuación disponibles

$$MSE(M) = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{N}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = \sum_{i=1}^{N}(\hat{\epsilon}_i)^2$$



Aprendizaje 2

Usando MSE, podemos determinar analíticamente el mejor valor para la obtención de β0 y β1:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(x_i - \overline{X})(y_i - \overline{Y})}{\sum_{i=1}^{N}(x_i - \overline{X})^2}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$$

media de los datos observados

- **Interpretación de los parámetros:**
  β_1 => Y estimado cambia en β1 para cada unidad que X aumente
  β_0 => Valor medio de Y Cuando X=0

Máxima verosimilitud

Otra función de puntuación es la función de verosimilitud.

La verosimilitud de un conjunto de valores de parámetro, β, dados los resultados x, es igual a la probabilidad de esos resultados observados dados esos valores de parámetro.

Teniendo en cuenta que el error se modela a través de una distribución Normal con varianza conocida, la verosimilitud es:

$$L(M) = \prod_{i=1}^{n} P(\hat{\epsilon}_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2\right)$$

$$ln(L(\beta_0, \beta_1)) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

El procedimiento de búsqueda debe encontrar el mejor valor para β0 y β1.

Afortunadamente, hay una solución de forma cerrada para β0 y β1, y obtiene los mismos parámetros que el uso de la función de puntuación MSE.

Coeficiente de determinación

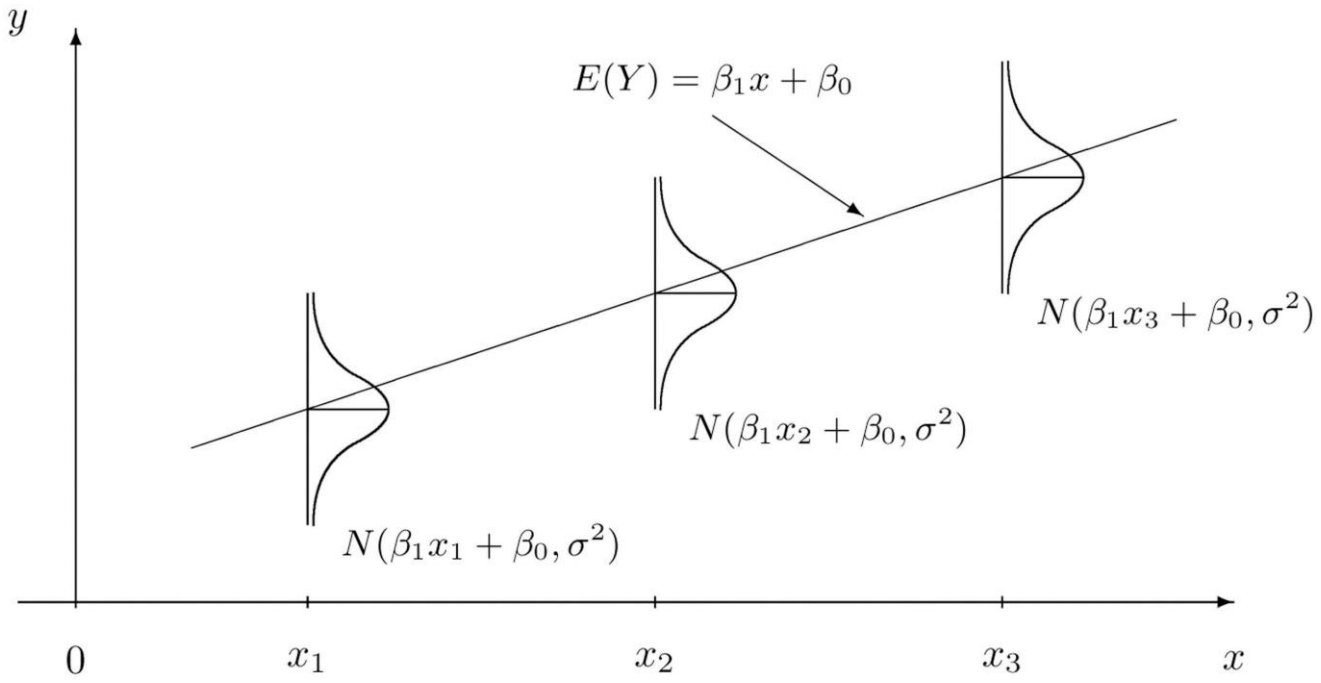
Una estadística que se utiliza para cuantificar qué tan bien la línea ajustada describe los datos es el coeficiente de determinación.

Se define como la relación entre la suma de cuadrados de la regresión y la suma total de cuadrados.

$$r^2 = \frac{\text{Regression sum of squares}}{\text{Total sum of squares}} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}$$

Este coeficiente mide la proporción de la variación total en y1,...,yn que se explica por el modelo

Distribución de los parámetros



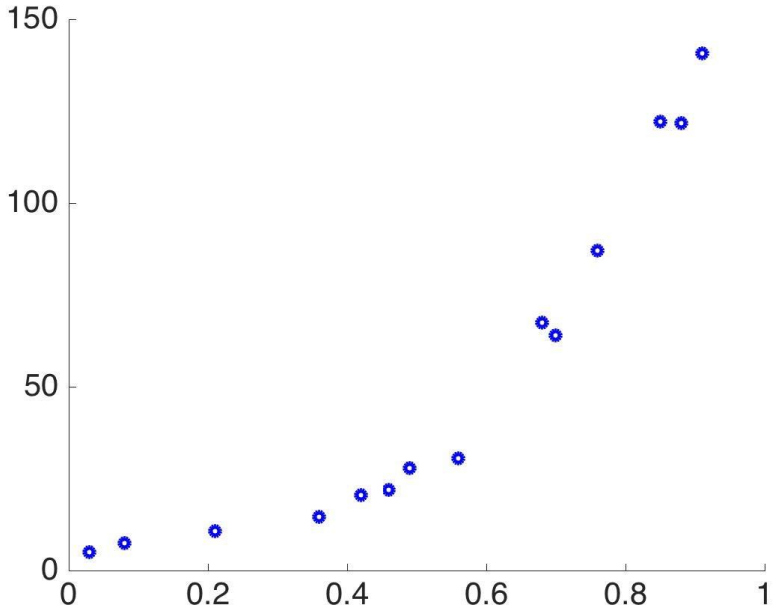

```
Call:
lm(formula = height ~ age + no_siblings, data = ageandheight)

Residuals:
     Min       1Q   Median       3Q      Max
-0.28029 -0.22490 -0.02219  0.14418  0.48350

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 64.95872    0.55752 116.515 1.28e-15 ***
age          0.63516    0.02254  28.180 4.34e-10 ***
no_siblings -0.01137    0.05893  -0.193    0.851
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2693 on 9 degrees of freedom
Multiple R-squared:  0.9888,    Adjusted R-squared:  0.9863
F-statistic: 397.7 on 2 and 9 DF,  p-value: 1.658e-09
```

Linealización de modelos

Este método transforma todo el conjunto de datos en una forma lineal, aplica una sola regresión lineal y realiza la transformación inversa

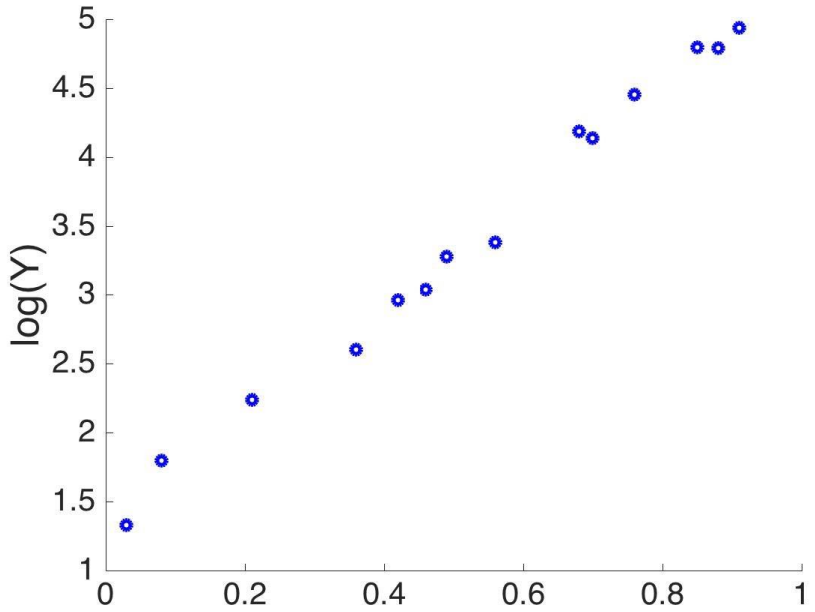
Observe data

$$Y = \beta_0 \exp(\beta_1 X)\epsilon$$



Transform variable

$$\log(Y) = \log(\beta_0) + \beta_1 X + \log(\epsilon)$$

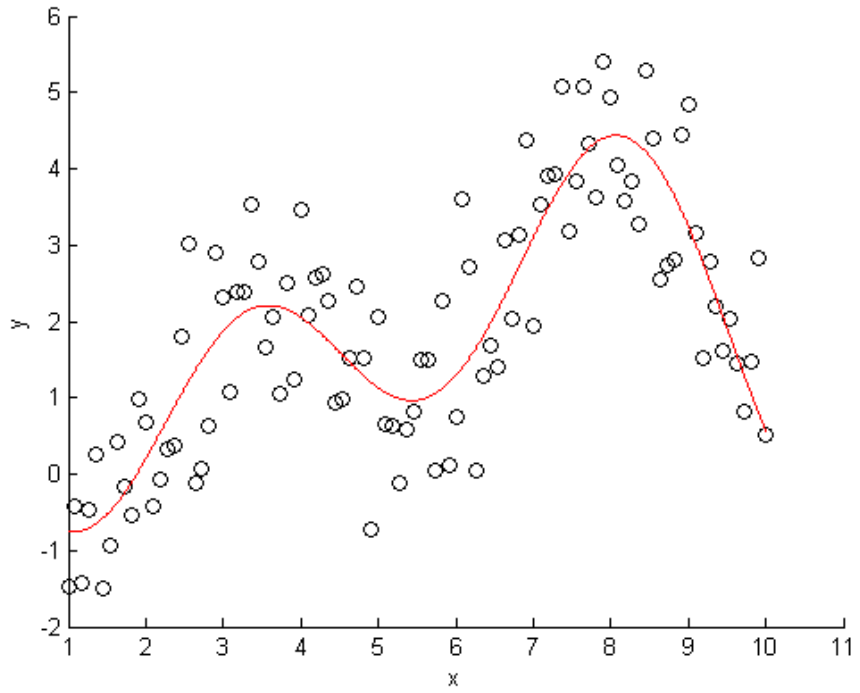


Regresión polinomial

Las regresiones polinómicas se utilizan cuando los datos son polinomios, o se desconoce la función curvilínea, pero una regresión polinómica se ajusta a los datos reales

Un modelo de regresión polinómica de orden p, para un único predictor X, es

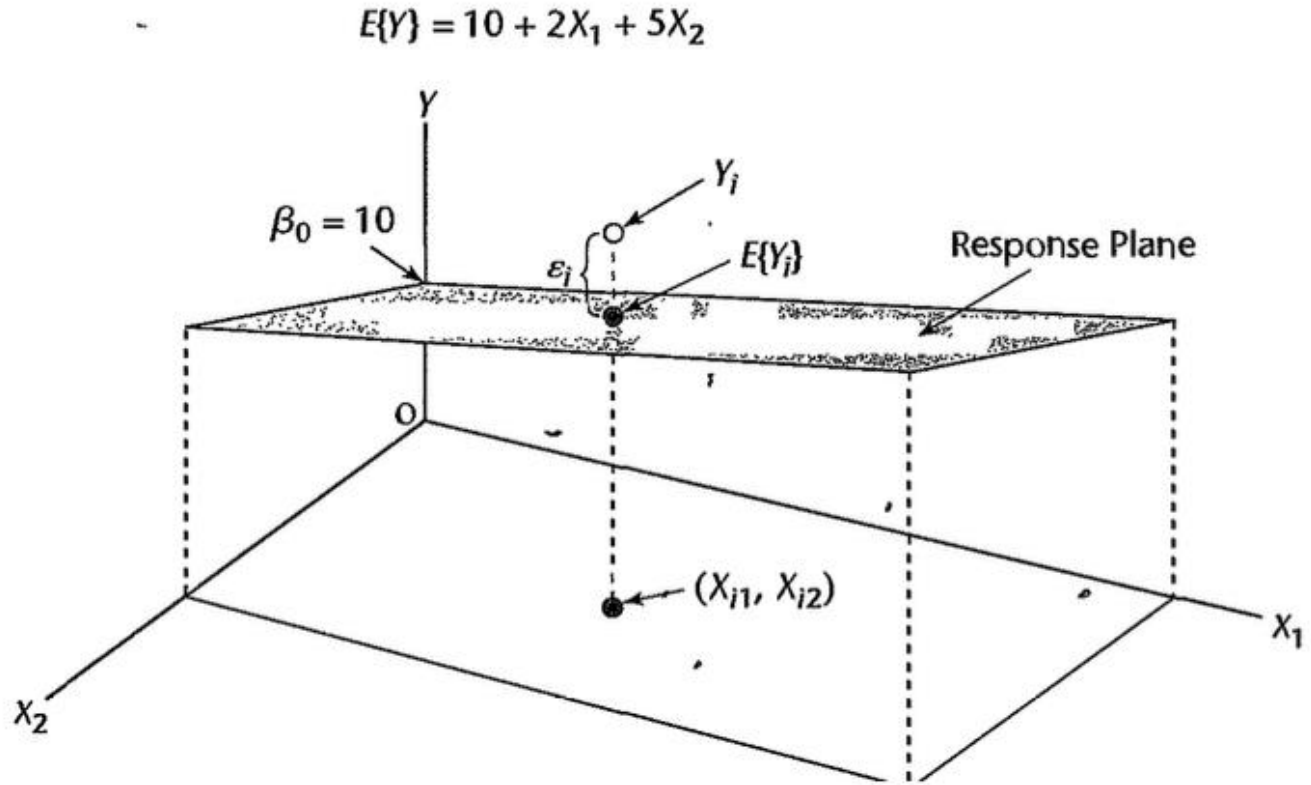
$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \cdots + \beta_p X^p + \epsilon$$



Regresión lineal múltiple

La variable dependiente está relacionada linealmente con varias variables predictoras, generando un hiperplano.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \epsilon$$



Determinando un modelo

Dadas varias variables, existen diversos métodos para determinar aquellos atributos que generan el mejor modelo:

- Fuerza bruta: probar todos los modelos posibles (las combinaciones en todas las variables, obteniendo un total de modelos 2p), y evaluar su error, utilizando r2
- Selección/generación hacia adelante: Dado un modelo vacío (modelo sin variables), la selección hacia delante agrega, secuencialmente, la variable que tiene la mayor mejora en una medida de error (hasta que el error no se mejora significativamente).
- Selección/generación hacia atrás: Dado un modelo completo (modelo con todas las variables), la eliminación hacia atrás elimina, secuencialmente, la variable que tiene el decremento más bajo en una medida de error (esto no es significativo).

Análisis de regresión

Dr. Raimundo Sánchez

raimundo.sanchez@uai.cl

@raimun2