

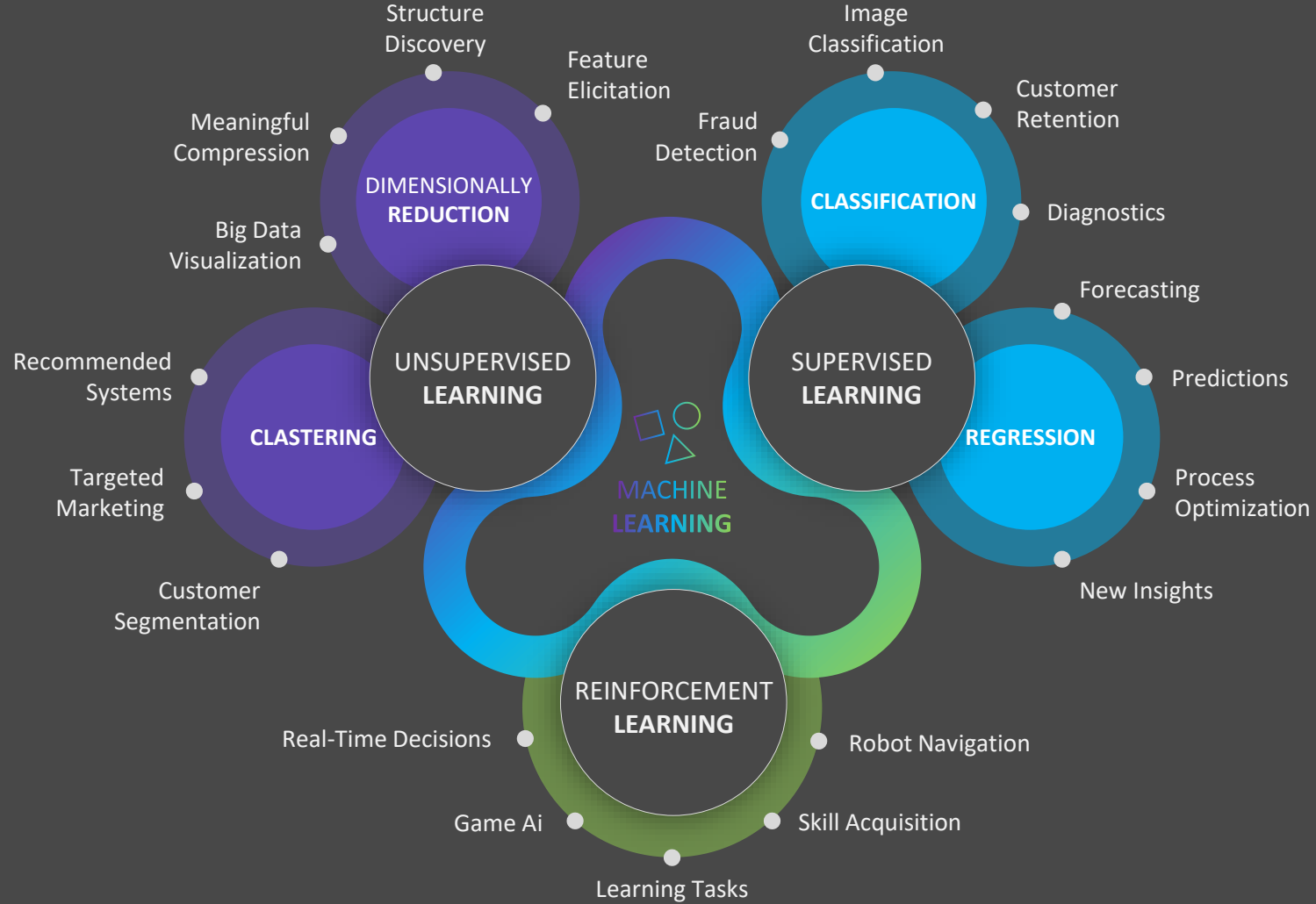


FORECASTING: UNA
APROXIMACIÓN BASICA.

CASO DE USO: PREDICCIÓN DE LA
DEMANDA ELECTRICA-.

[Napoleón Alcides Pérez Arteaga | LinkedIn](#)

Type of Machine Learning



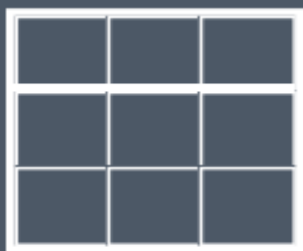
Serie temporal



Una serie temporal (*time series*) es una sucesión de datos ordenados cronológicamente, espaciados a intervalos iguales o desiguales.



El proceso de forecasting consiste en predecir el valor futuro de una serie temporal, bien modelando la serie únicamente en función de su comportamiento pasado (autorregresivo) o empleando otras variables externa.



DATA



PREDICTION



DECISION

DATA
COMPLEXITY

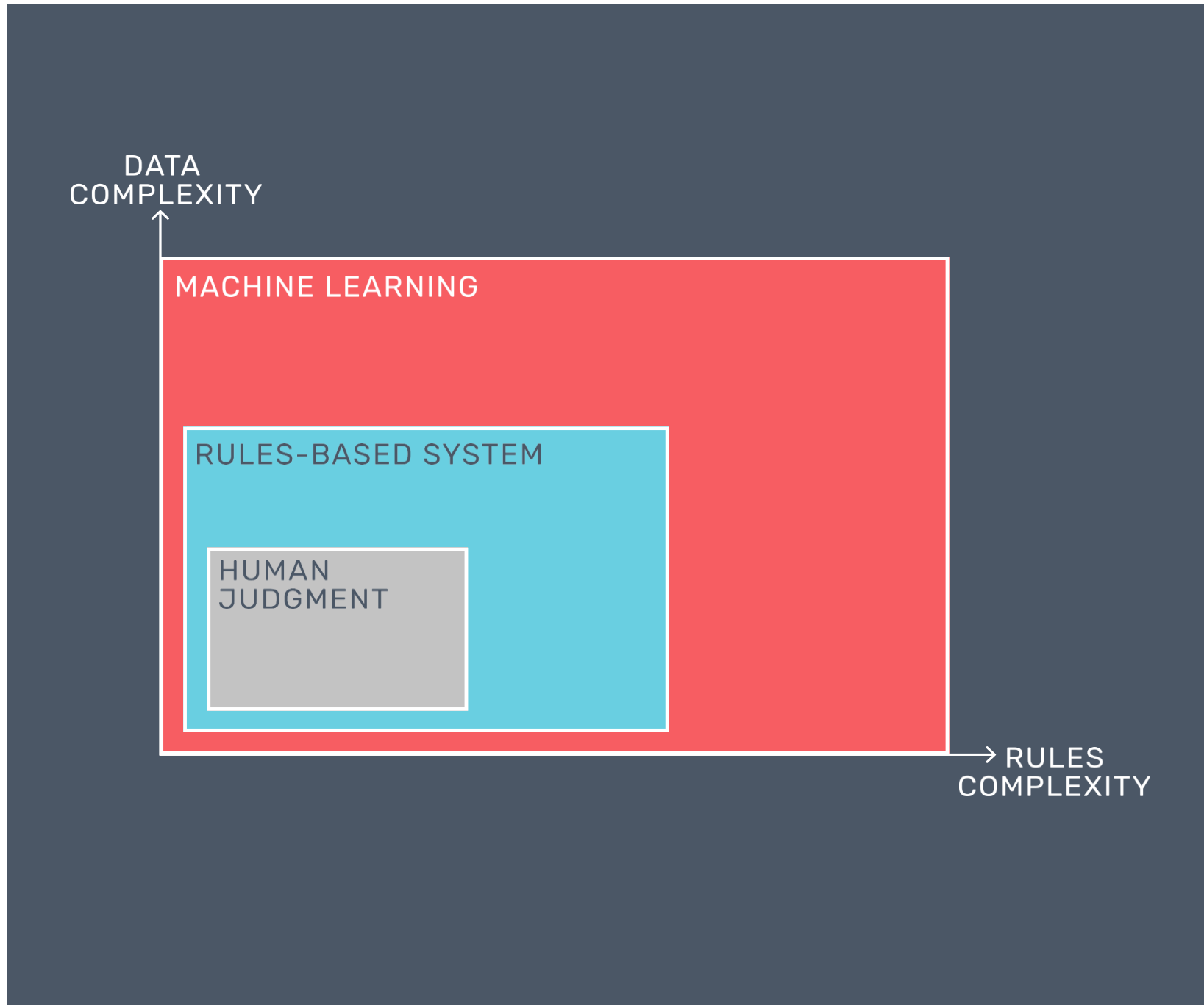


MACHINE LEARNING

RULES-BASED SYSTEM

HUMAN
JUDGMENT

→ RULES
COMPLEXITY



DATA

FEATURES
TARGET
TRAINING
TESTING



PREDICT

WEIGHTED SUM
ACTIVATION

TASK

LINEAR
NON-LINEAR
REGRESSION
CLASSIFICATION

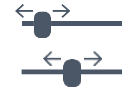


MEASURE

COST
METRICS

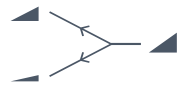
NEURAL NETWORK

NEURONS
LAYERS
ARCHITECTURE



ADJUST

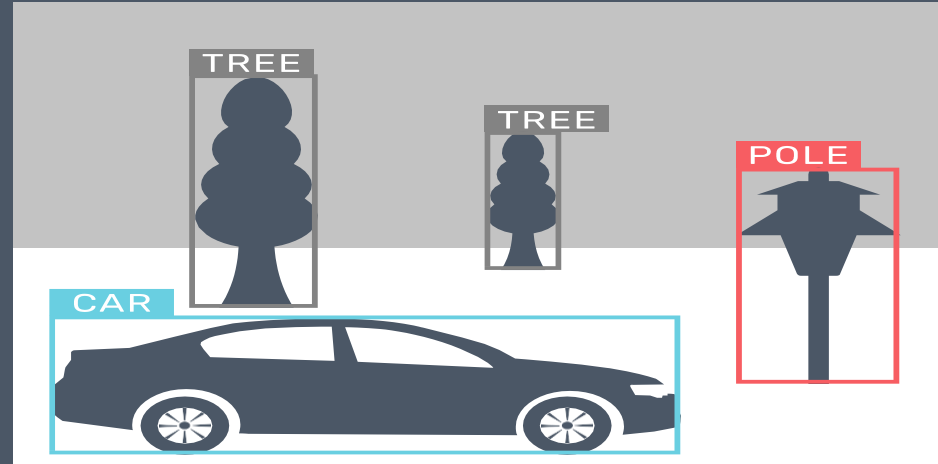
WEIGHTS
BIASES



FEEDBACK

GRADIENTS
BACKPROPAGATION

COMPUTER VISION



NATURAL LANGUAGE PROCESSING

SENTENCE	SENTIMENT
IT'S A GREAT DAY	POSITIVE
I DON'T LIKE MONDAYS	NEGATIVE



MODELOS MAS COMUNES DE FORECASTING

Naïve, SNaïve

Seasonal
decomposition
(+ any model)

Exponential
smoothing

ARIMA,
SARIMA

GARCH

Dynamic linear
models

TBATS

Prophet

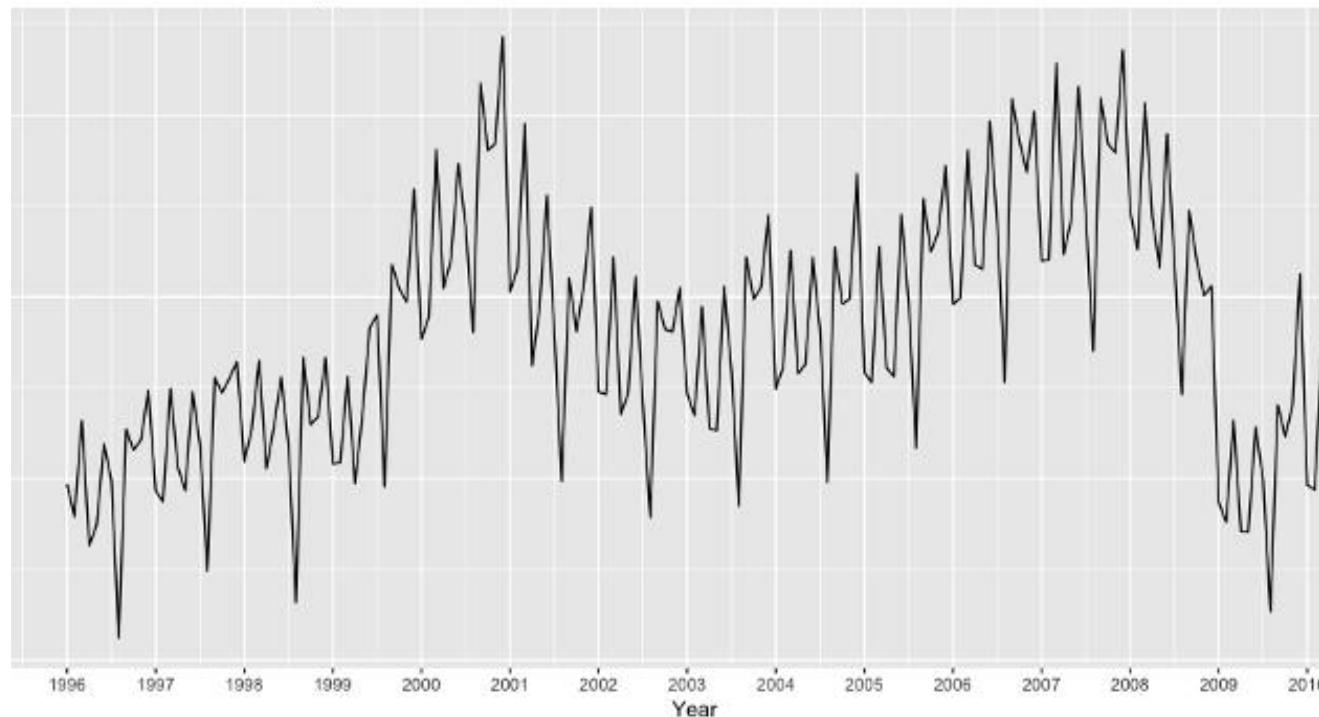
NNETAR

LSTM

SEASONAL DECOMPOSITION

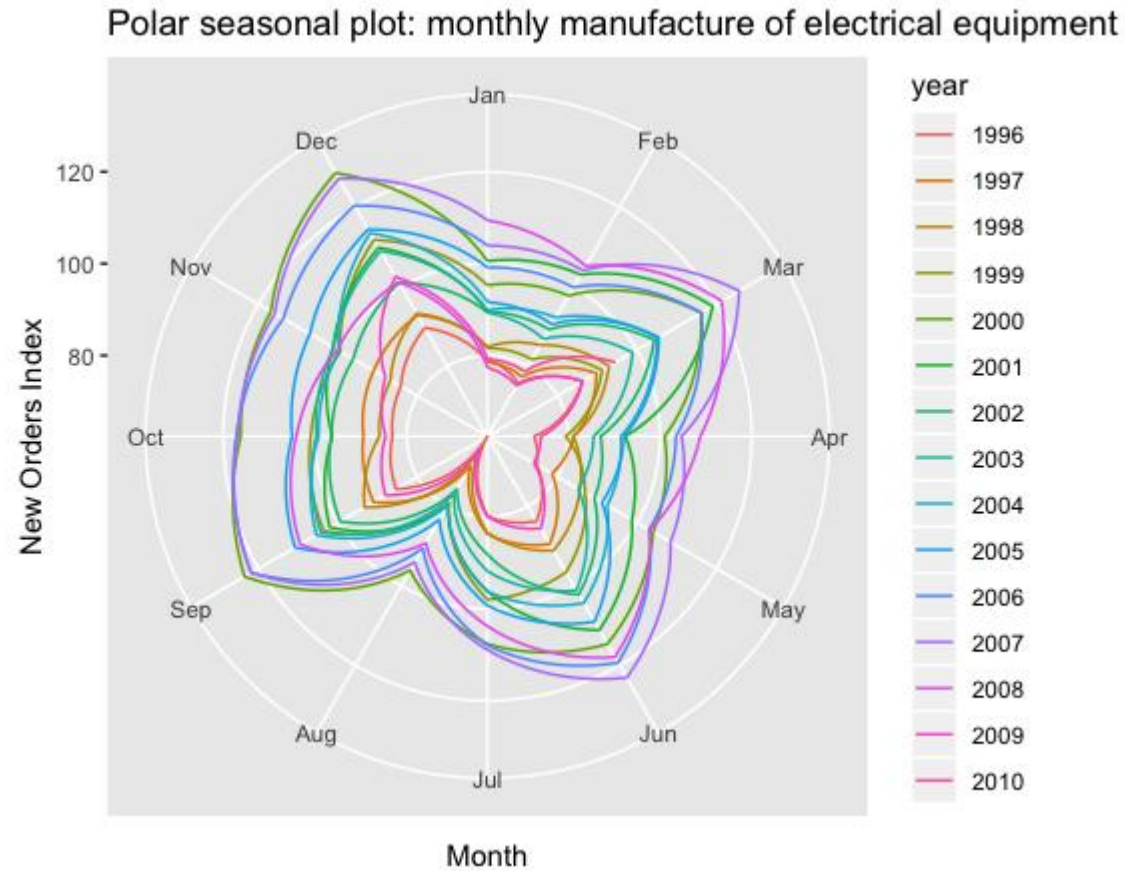
DEMANDA DE FABRICACION

Manufacture of electrical equipment



- El conjunto de datos corresponde a la fabricación mensual de equipos eléctricos (productos informáticos, electrónicos y ópticos) en la zona del euro (17 países) en el período enero 1996-marzo 2012.

DETECTAR PATRONES ESTACIONALES



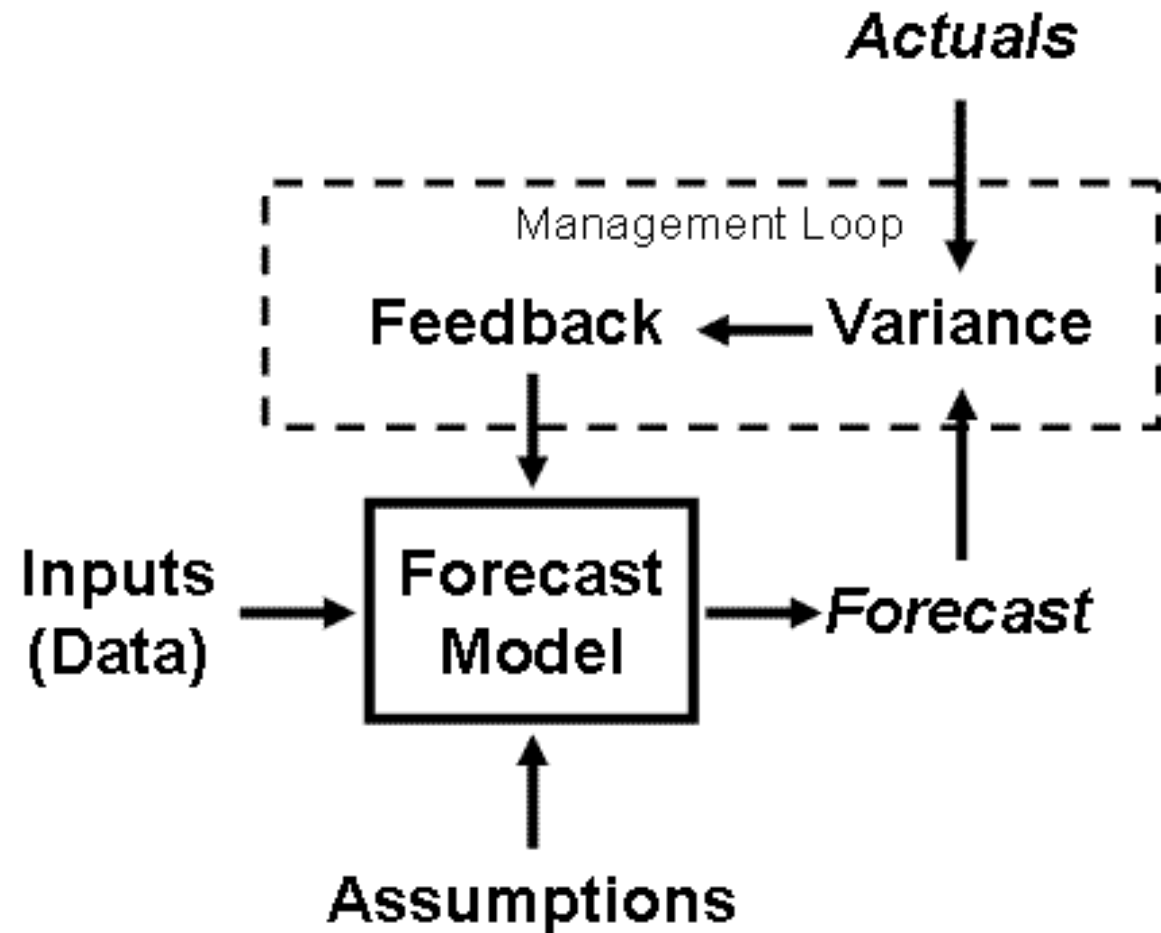
Pronóstico de series temporales

- La previsión de series temporales se produce cuando se realizan predicciones científicas basadas en datos históricos con marca de tiempo.

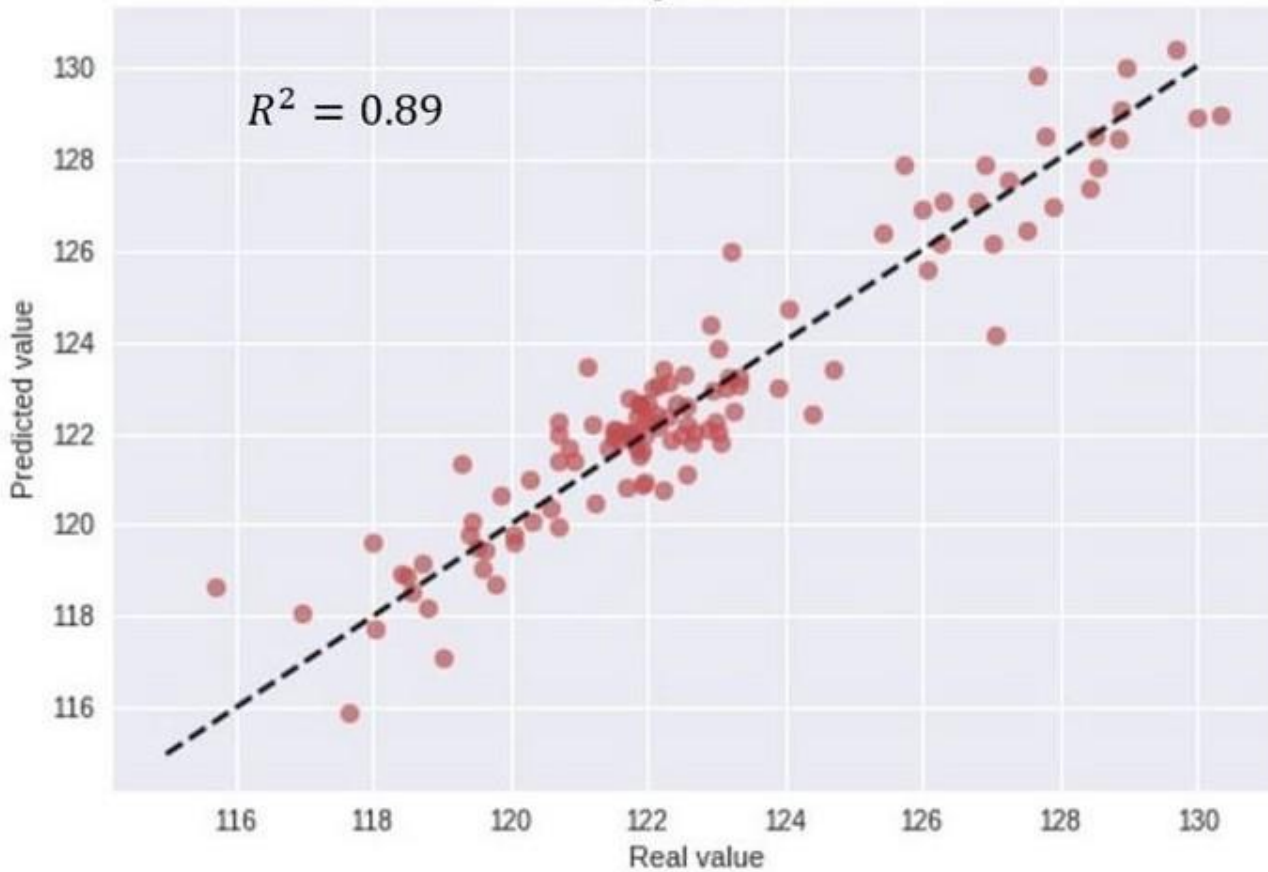


PROCESO GLOBAL

- El proceso de Forecasting consiste en predecir el valor futuro de una serie temporal, bien modelando la serie temporal únicamente en función de su comportamiento pasado (autorregresivo) o empleando otras variables externas a la serie temporal.



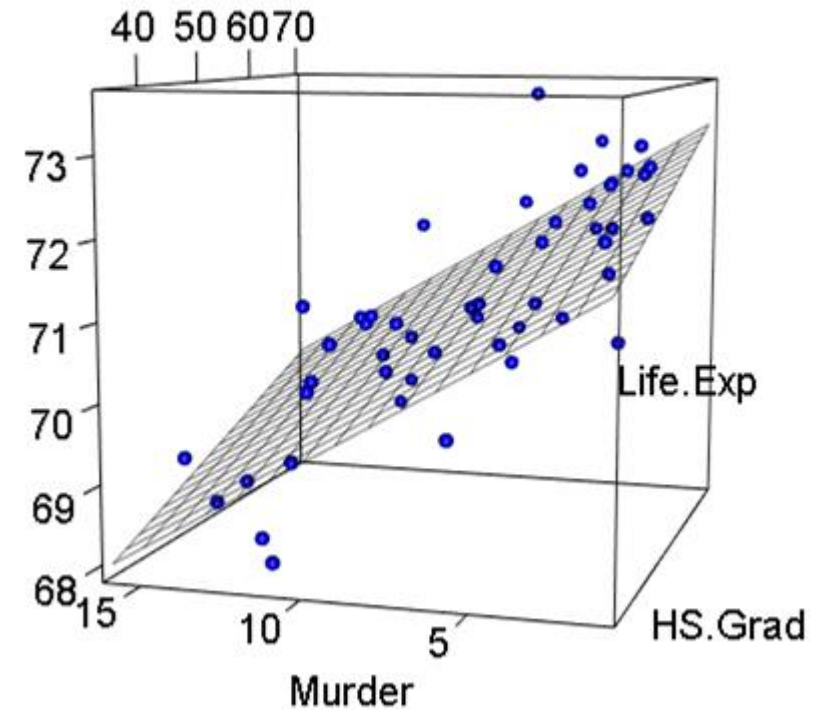
Original data



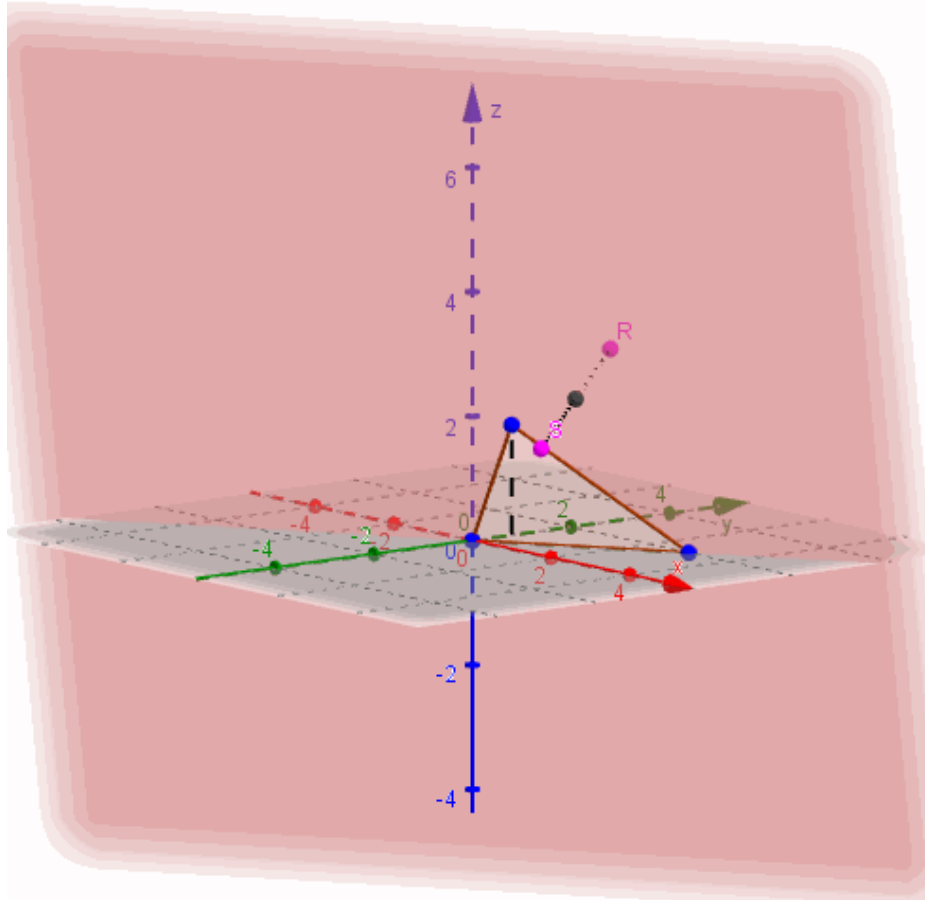
En conclusión, R^2 es la proporción entre lo bueno que es nuestro modelo y lo bueno que es el modelo medio ingenuo.

El R cuadrado se utiliza en la regresión múltiple para ver el grado de intensidad o efectividad que tienen las variables independientes en explicar la variable dependiente.

Multiple Regression \rightarrow $y = m_1x_1 + m_2x_2 + b$



INTUICION MATEMATICA



$$AX + BY + CZ + D = 0$$

Plano con vector (A, B, C)

De las ecuaciones paramétricas tenemos:

$$x = x_0 + ta \rightarrow \frac{x - x_0}{a} = t$$

$$y = y_0 + tb \rightarrow \frac{y - y_0}{b} = t$$

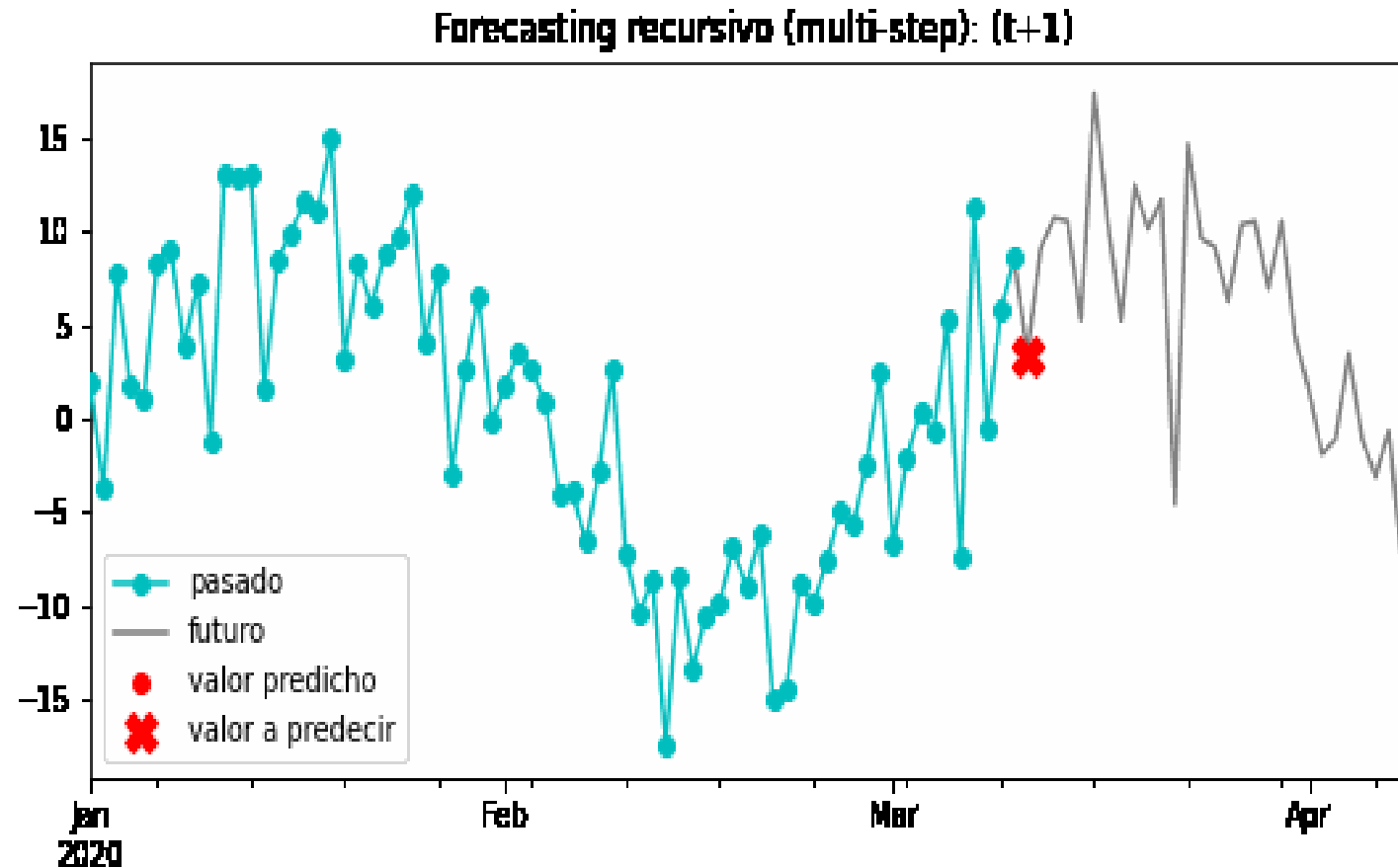
$$z = z_0 + tc \rightarrow \frac{z - z_0}{c} = t$$

Finalmente,

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c} = t$$

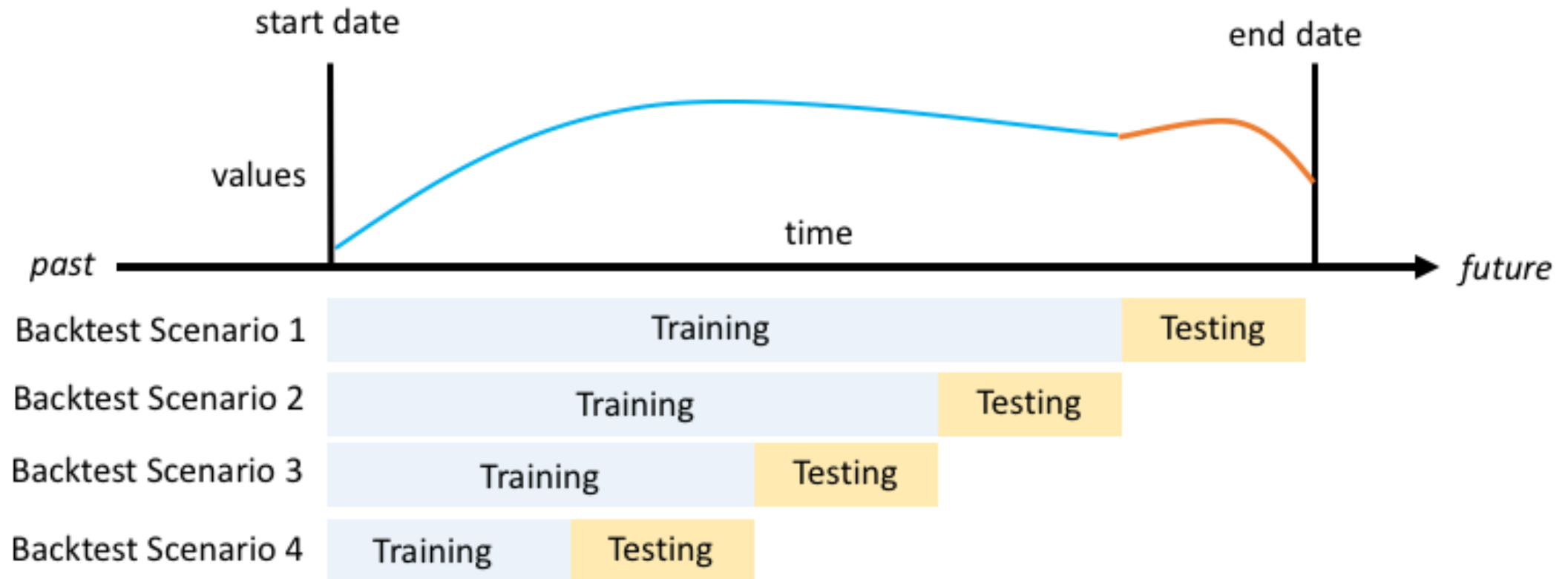
*Ecuación de la recta
en forma simétrica*

FORECASTING RECURSIVO



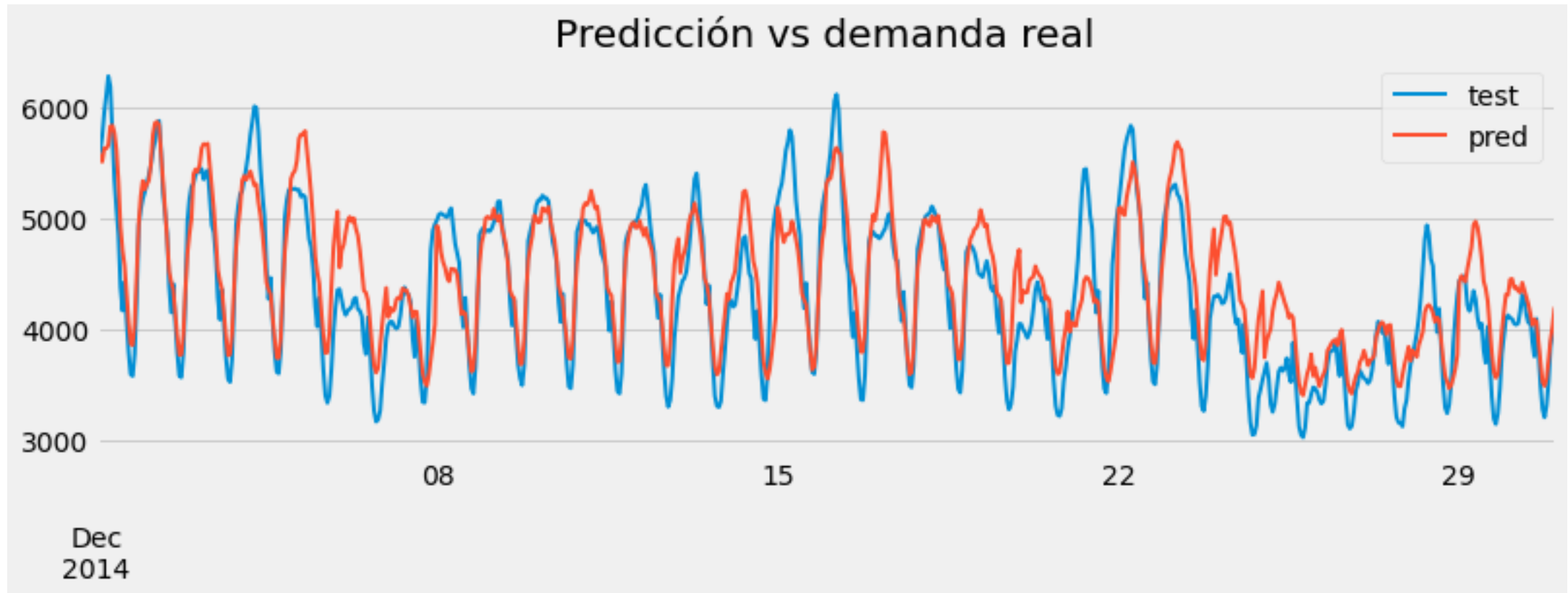
Dado que, para predecir el momento t_n se necesita el valor de $t(n-1)$, y $t(n-1)$ se desconoce, es necesario hacer predicciones recursivas en las que, cada nueva predicción, se basa en la predicción anterior

METRICAS DEL MODELO

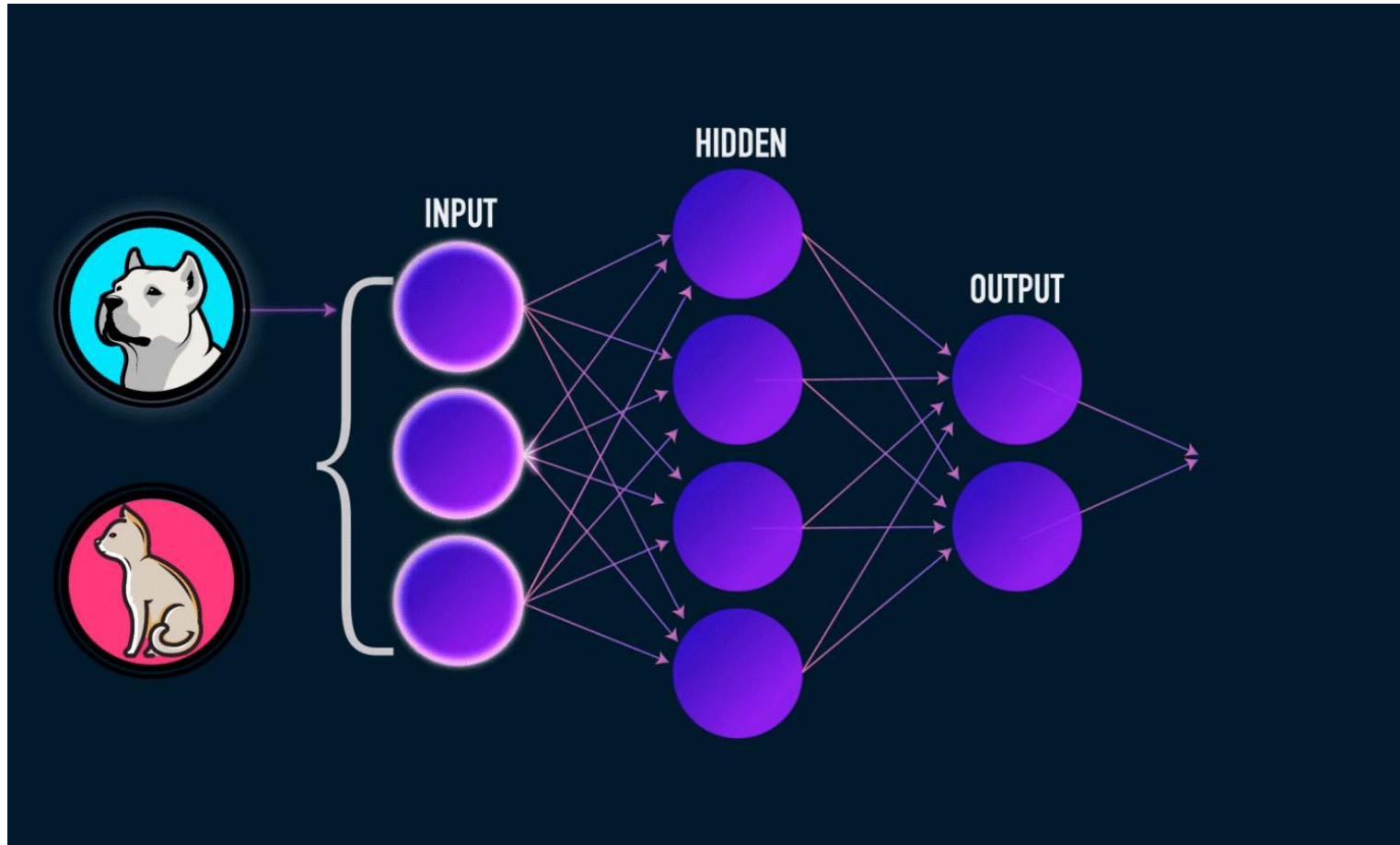


CASO DE USO

- Se dispone de una serie temporal con la demanda eléctrica (MW) del estado de Victoria (Australia) desde 2011-12-31 al 2014-12-31. Se pretende generar un modelo de *forecasting* capaz de predecir la demanda energética del día siguiente a nivel horario.

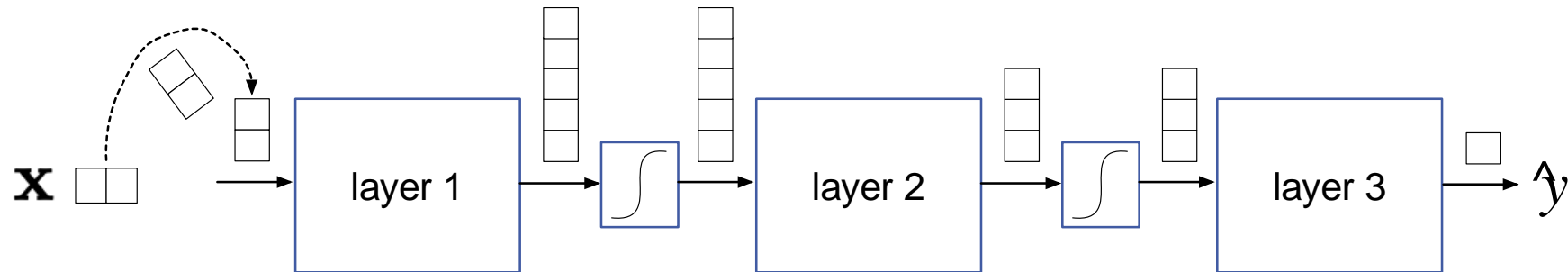


RED NEURONAL



RED NEURONAL

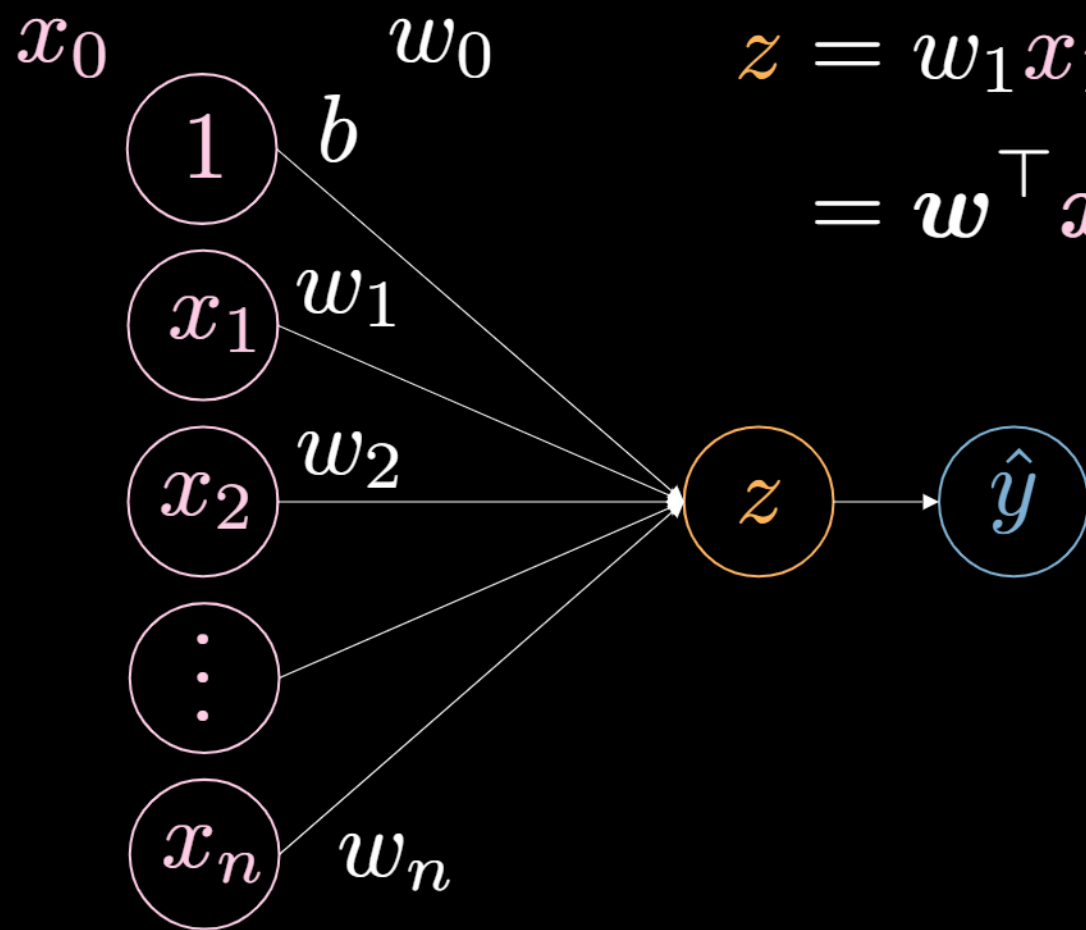
- Combinación lineal y no lineal de transformaciones
- Lineal: $z^{[layer]} = W^{[layer]} \mathbf{x}^T + \mathbf{b}^{[layer]}$
 - Nonlineales: $\mathbf{a}^{[layer]} = \sigma(z^{[layer]})$; FUNCION DE ACTIVACION
- Las neuronas tienen multiples capas.



Transforma el vector \mathbf{x} sin procesar en características cada vez mejores, la capa lineal final puede hacer una predicción excelente

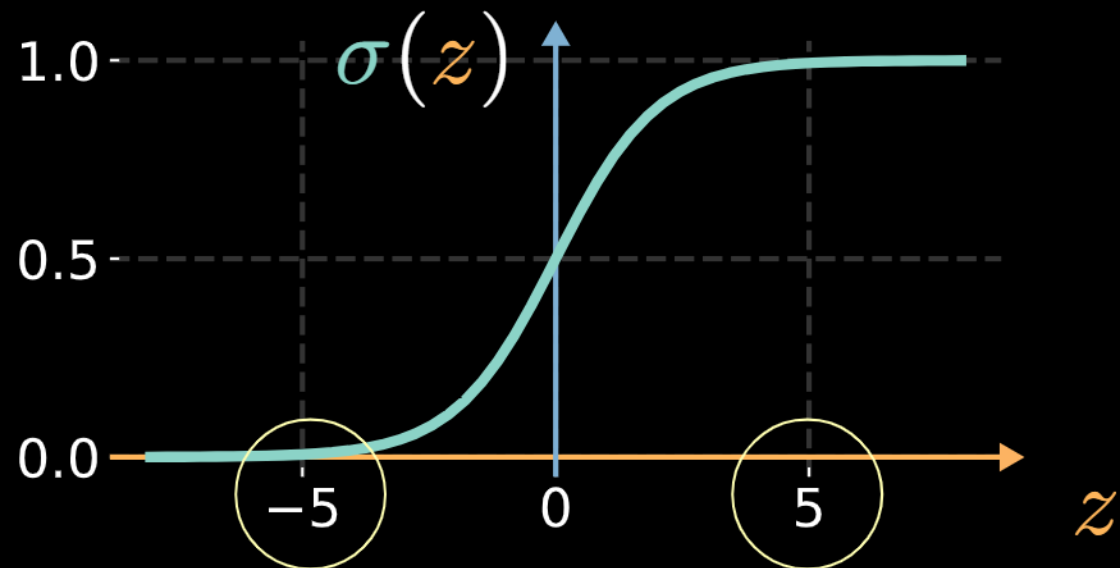
Neuron / perceptron

$$\sigma(z) = \frac{1}{1 + \exp(-z)}$$

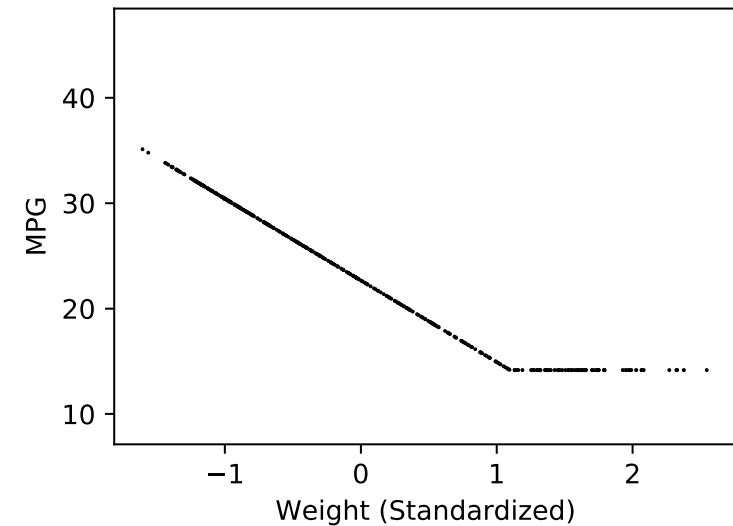
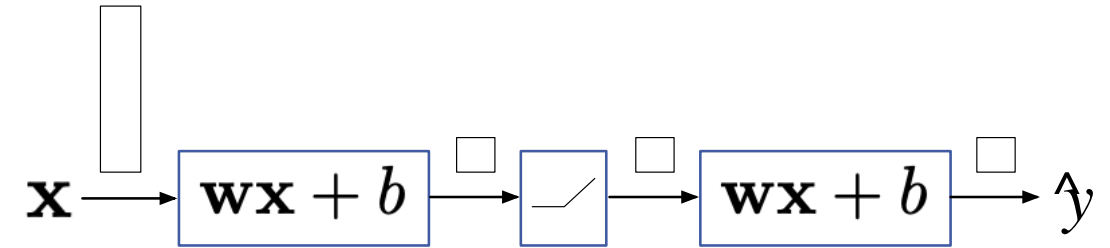
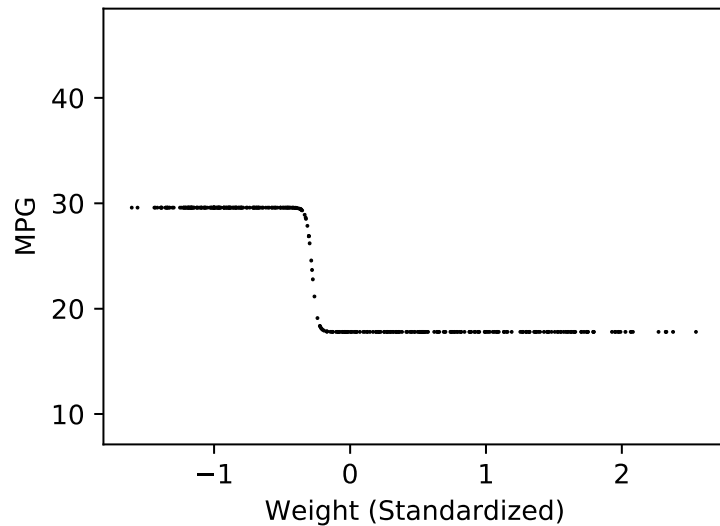
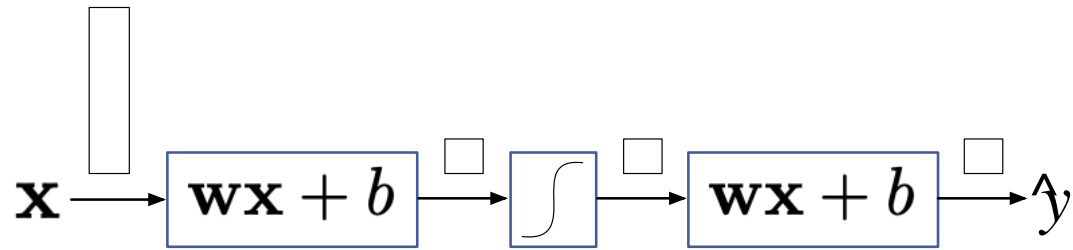


$$z = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n + b$$
$$= \mathbf{w}^\top \mathbf{x} + b = \tilde{\mathbf{w}}^\top \tilde{\mathbf{x}}$$

$$\hat{y} = \sigma(z)$$



No linealidad



ReLU idea here: Dibujar dos líneas y capturar la intersección

Neural net training

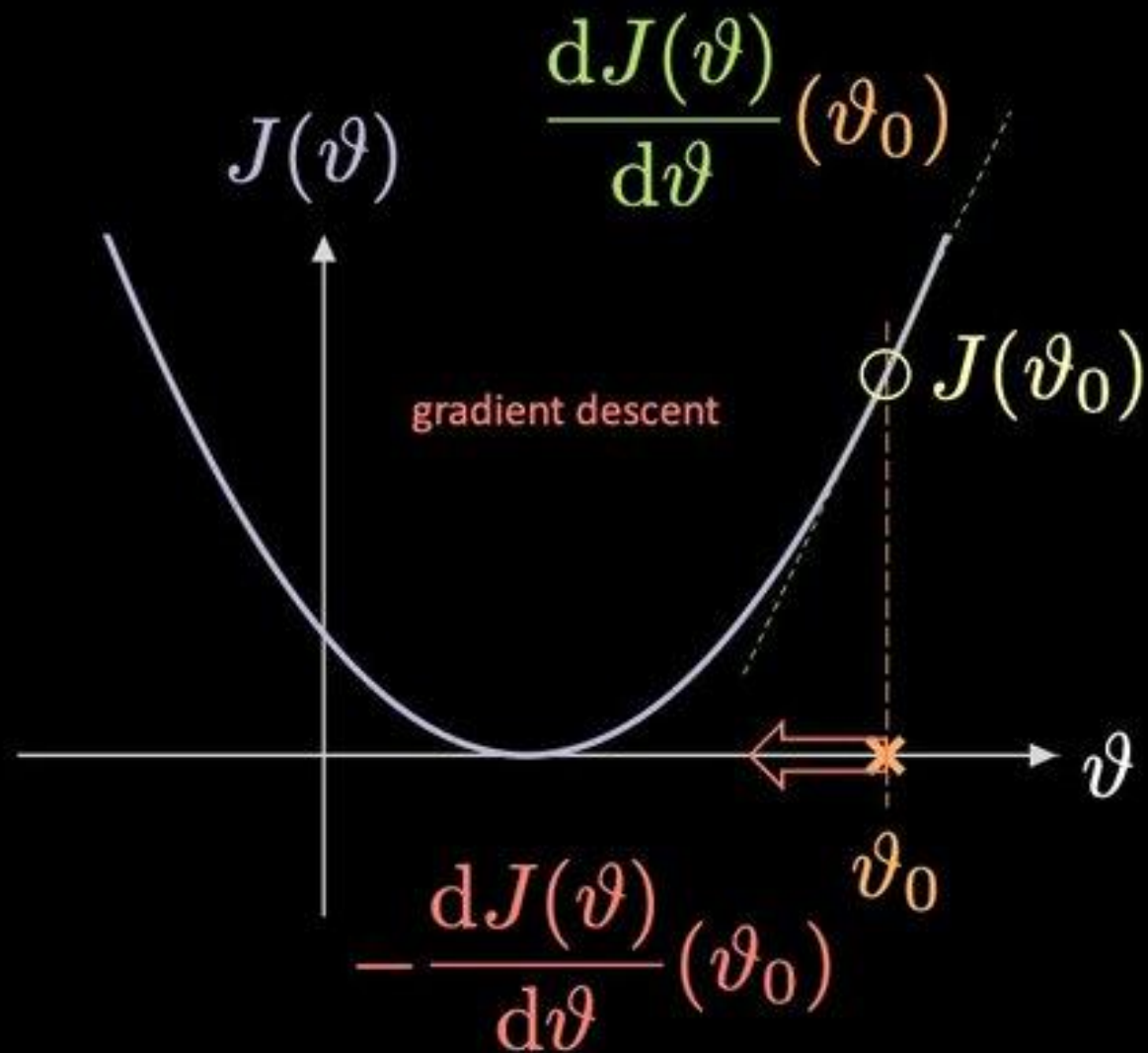
$$\Theta \doteq \{\mathbf{W}_h, \mathbf{b}_h, \mathbf{W}_y, \mathbf{b}_y\}$$

$$J(\Theta) \doteq \mathcal{L}(\tilde{\mathbf{Y}}(\Theta), \mathbf{c}) \in \mathbb{R}^+$$

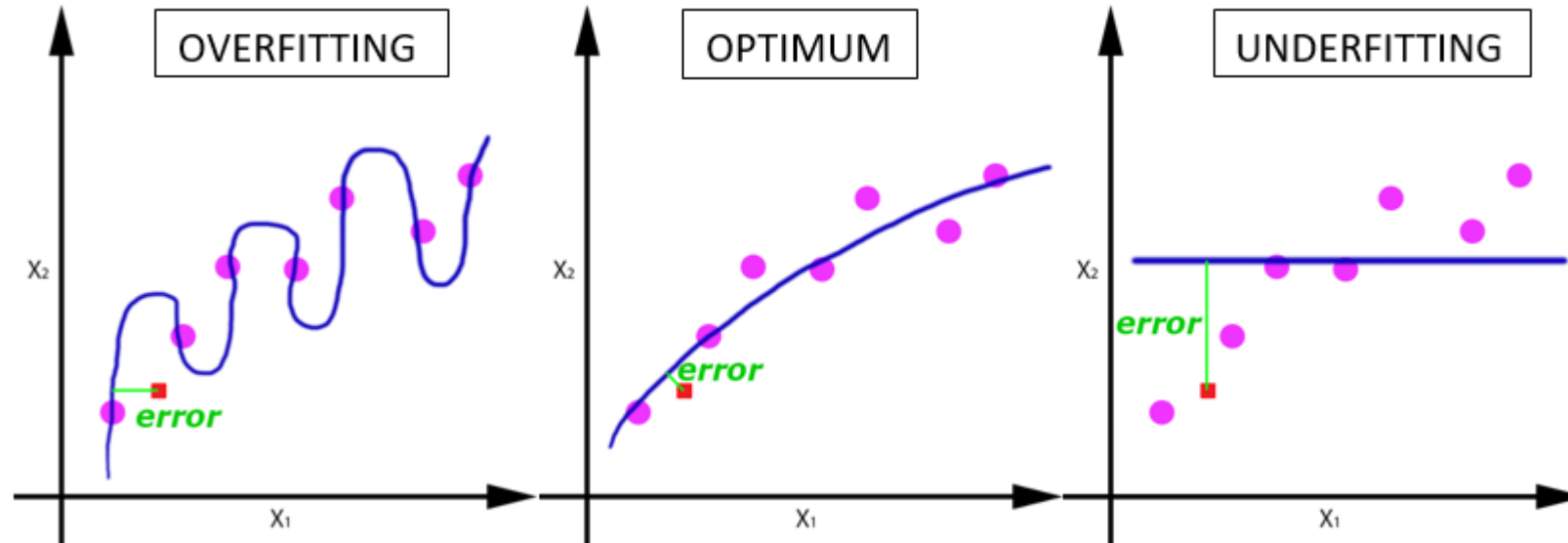
$$\frac{\partial J(\Theta)}{\partial \mathbf{W}_y} = \frac{\partial J(\Theta)}{\partial \tilde{\mathbf{y}}} \frac{\partial \tilde{\mathbf{y}}}{\partial \mathbf{W}_y}$$

$$\frac{\partial J(\Theta)}{\partial \mathbf{W}_h} = \frac{\partial J(\Theta)}{\partial \tilde{\mathbf{y}}} \frac{\partial \tilde{\mathbf{y}}}{\partial \mathbf{h}} \frac{\partial \mathbf{h}}{\partial \mathbf{W}_h}$$

back-propagation



PROBLEMAS CON LOS DATOS

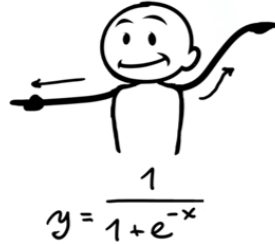


FUNCIONES DE ACTIVACION

- **FUNCIONES DE ACTIVACIONES NO LINEALES** :se utilizan al final de una unidad oculta para introducir complejidades no lineales en el modelo.

- **Stanford(CS229)**

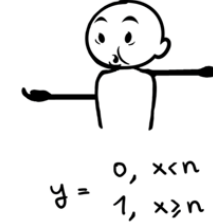
Sigmoid



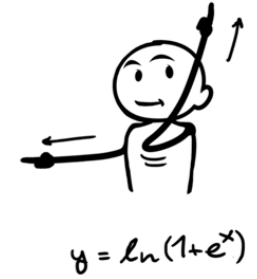
Tanh



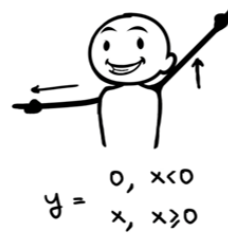
Step Function



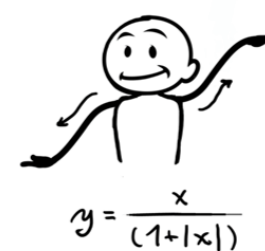
Softplus



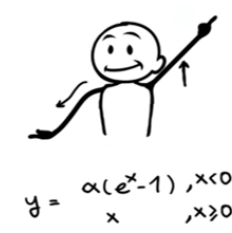
ReLU



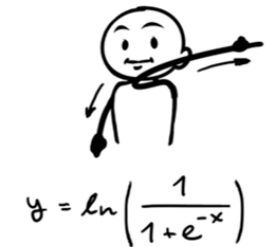
Softsign



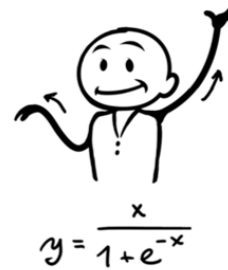
ELU



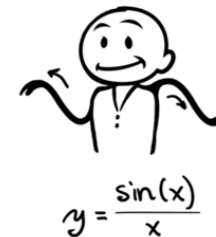
Log of Sigmoid



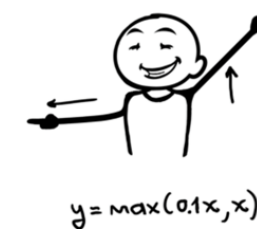
Swish



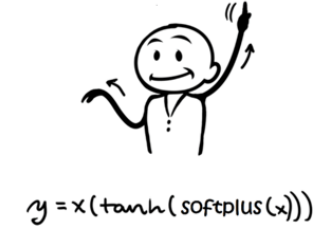
Sinc



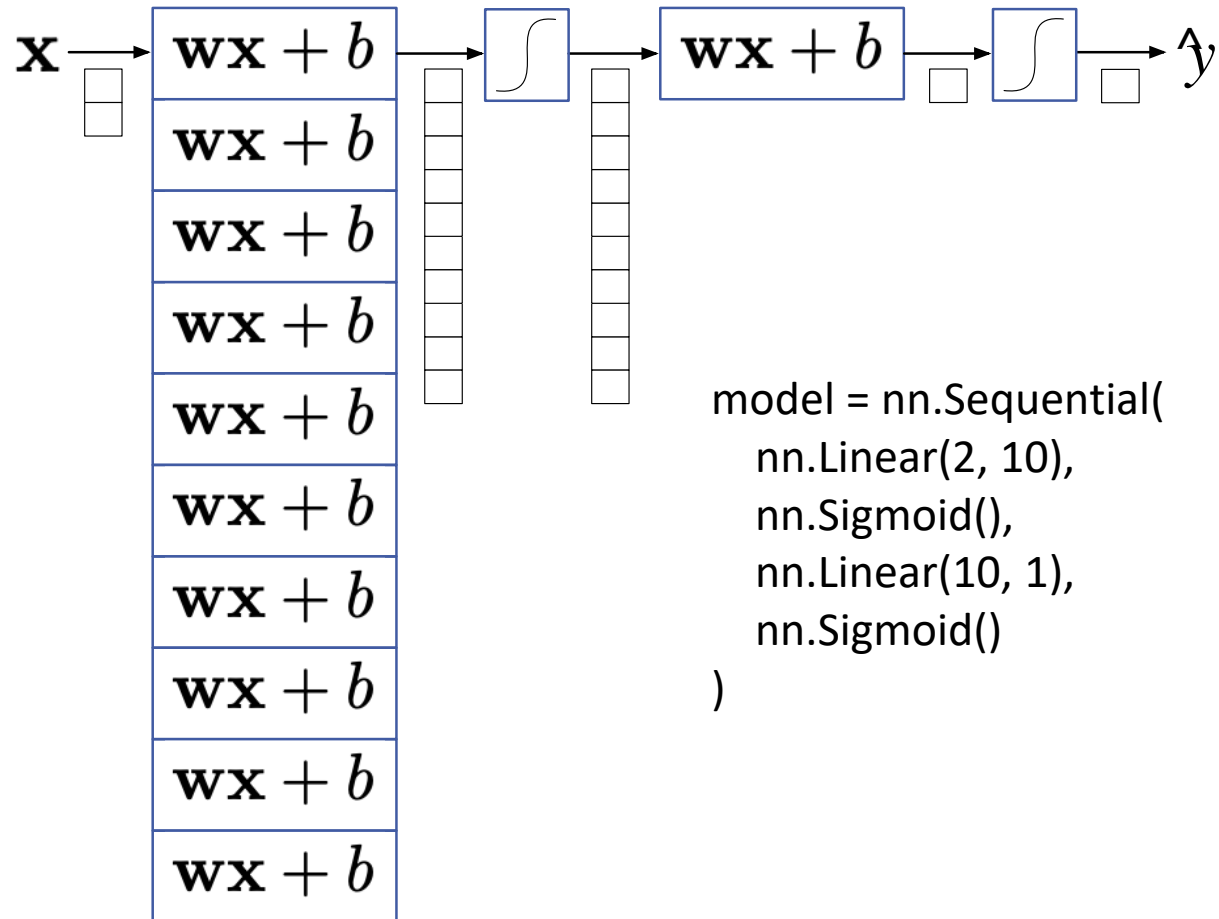
Leaky ReLU



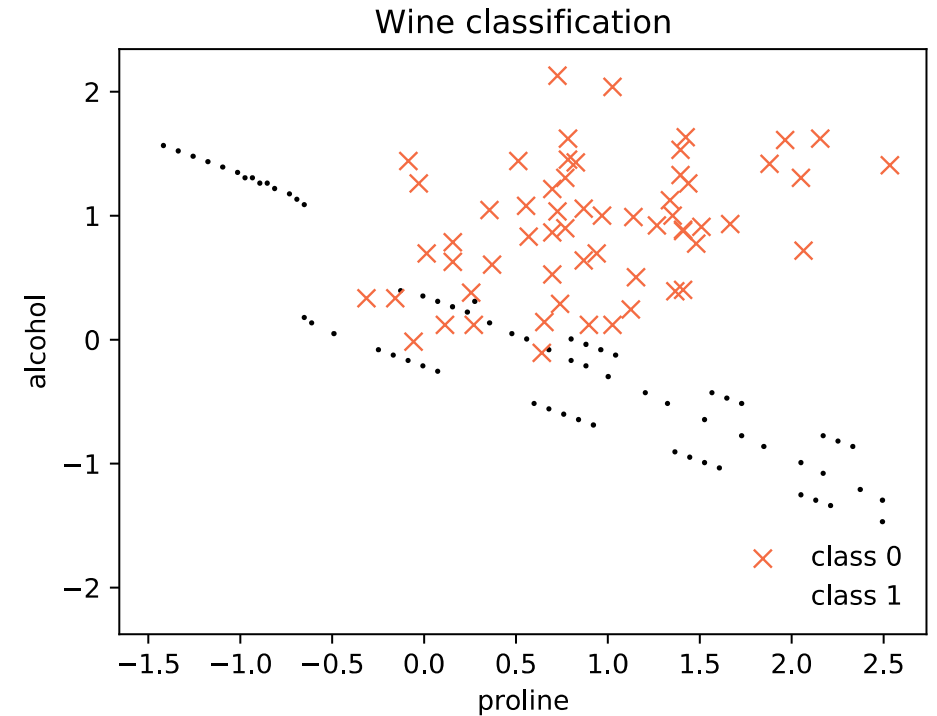
Mish



Mas neuranas: mayor complejidad de region de decisi3n



```
model = nn.Sequential(  
    nn.Linear(2, 10),  
    nn.Sigmoid(),  
    nn.Linear(10, 1),  
    nn.Sigmoid()  
)
```

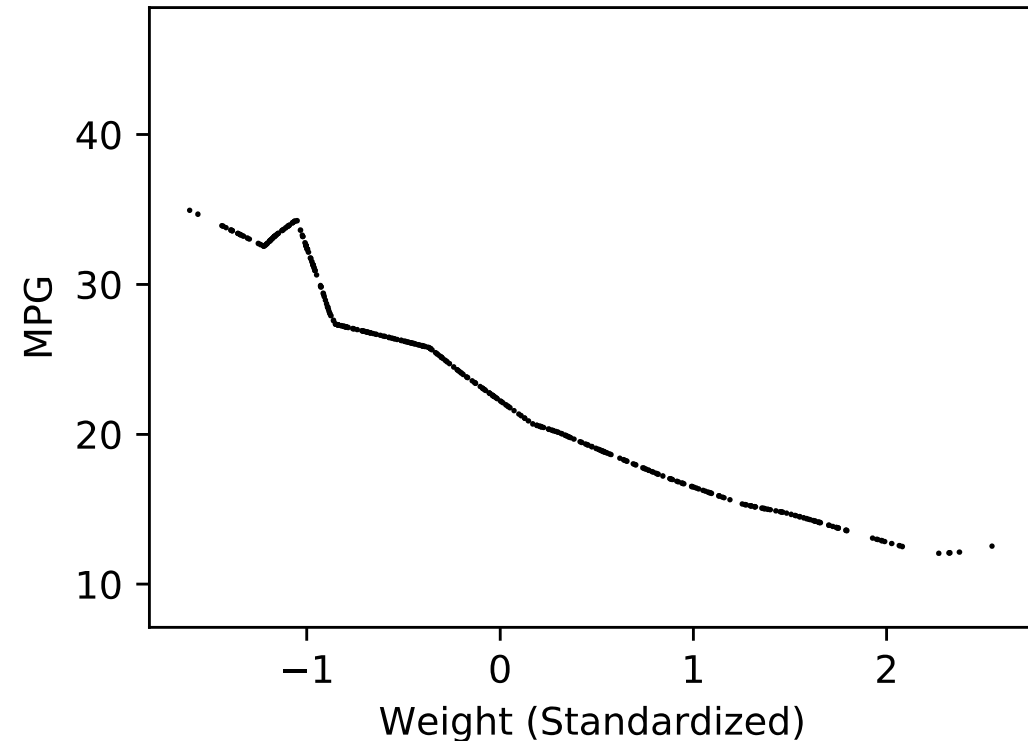


Likely overfit

Demasiada fuerza puede conducir a un sobreajuste

- Los modelos con demasiados parámetros se sobre ajustarán fácilmente si entrenamos durante mucho tiempo

```
model = nn.Sequential(  
    nn.Linear(1, 1000),  
    nn.ReLU(),  
    nn.Linear(1000, 1)  
)
```



Proceso de entrenamiento

Preparar los datos

Normalizar variables
numéricas

onehot para variables
categoricas.

Proceso de entrenamiento

Separar al menos un conjunto de validación del conjunto de entrenamiento

Elegir la arquitectura de la red, y una adecuada función de pérdida

Elegir hyper-parameters, como dropout rate.

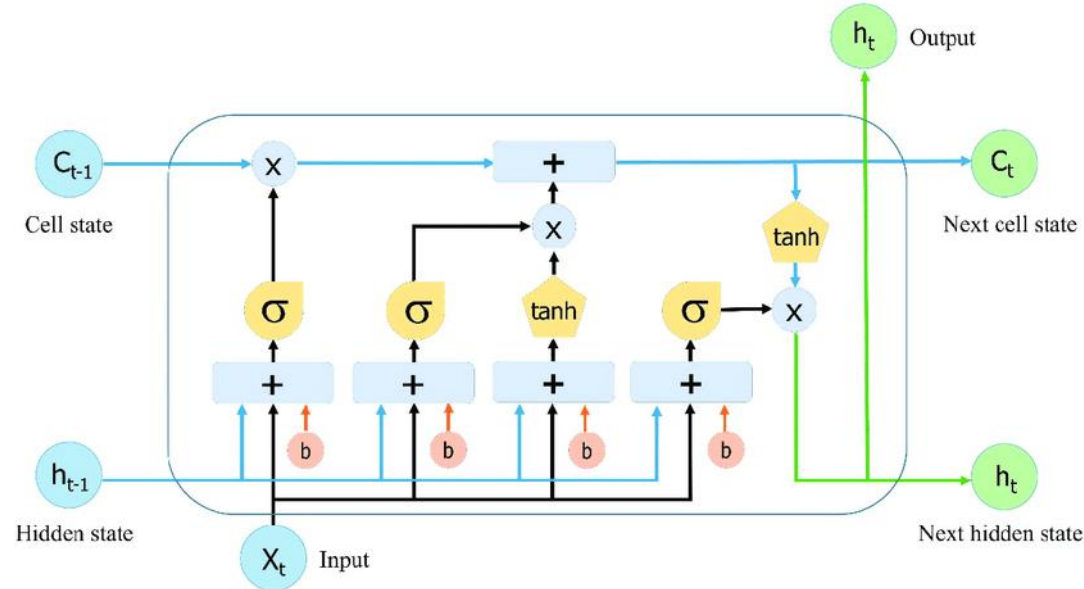
Proceso de entrenamiento

Elige un ritmo de aprendizaje, número de epoca(Pasar los datos por toda la red)

Ejecutar ciclo de entrenamiento(hasta que el error de validación aumente o el número de épocas)

Interar pasos 3 a 5.

LSTM



Inputs:

- X_t Current input
- C_{t-1} Memory from last LSTM unit
- h_{t-1} Output of last LSTM unit

Outputs:

- C_t New updated memory
- h_t Current output

Nonlinearities:

- σ Sigmoid layer
- \tanh Tanh layer
- b Bias

Vector operations:

- \times Scaling of information
- $+$ Adding information

Las redes LSTM son un tipo de red neuronal recurrente (RNN) que se desarrolló para hacer frente a las circunstancias en las que los RNN fallaron. Cuando se trata de RNN, son redes que funcionan con entradas de corriente teniendo en cuenta las salidas anteriores (retroalimentación) y manteniéndolas en la memoria durante un breve período de tiempo (memoria a corto plazo).