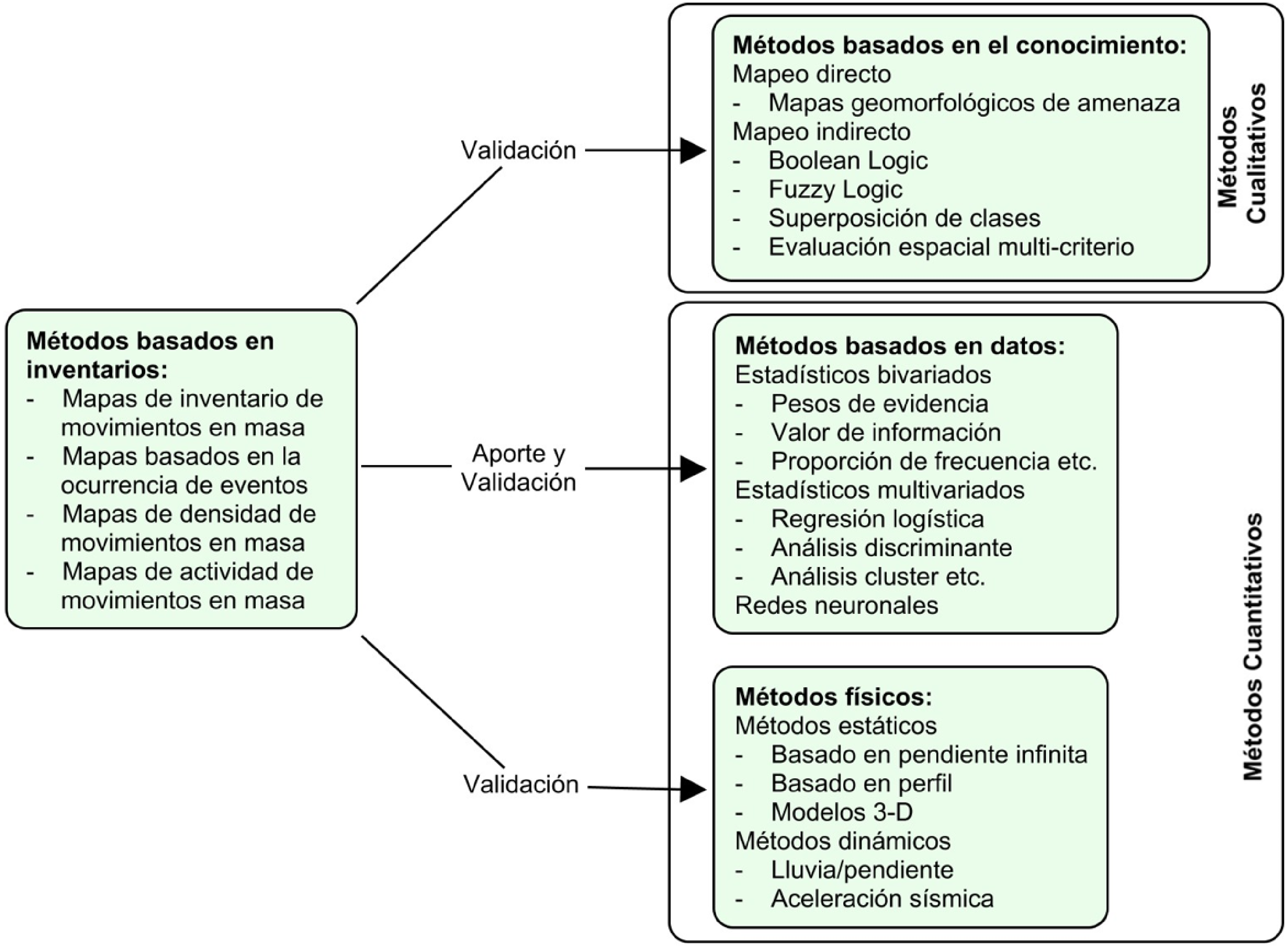
CARTOGRAFÍA GEOTÉCNICA

Machine Learning

Prof. Edier Aristizábal

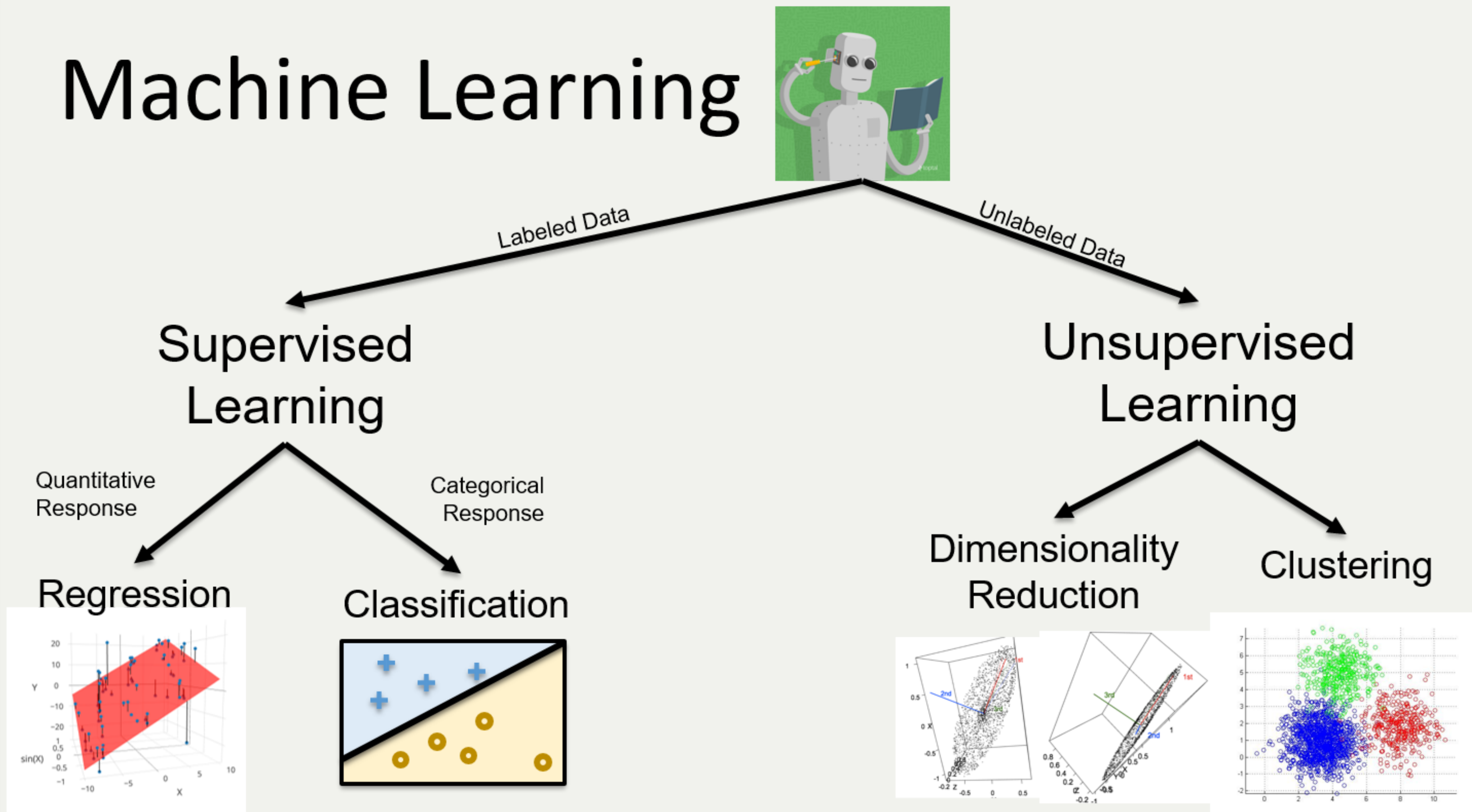
UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MEDELLÍN

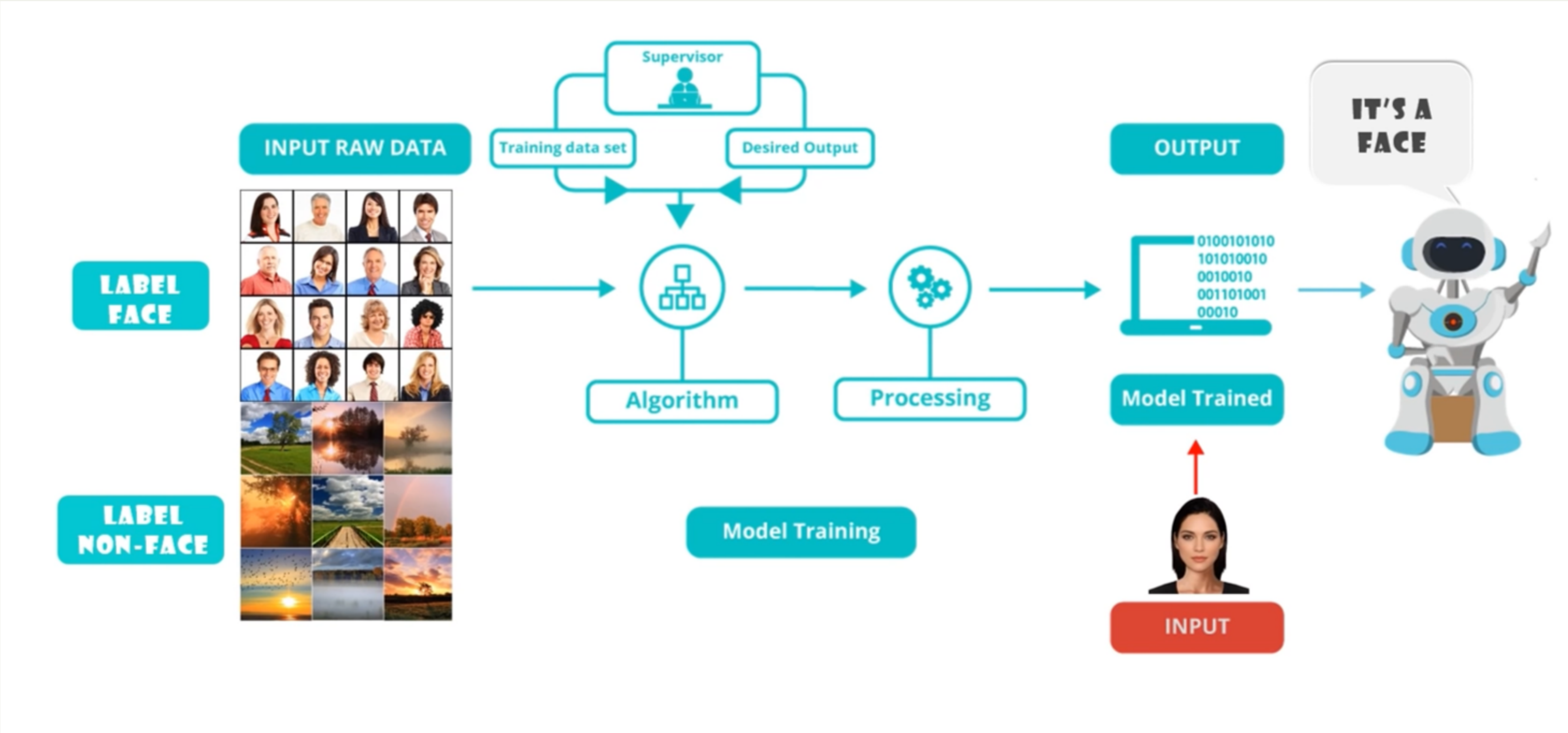
Oct, 16 / 2020



Métodos basados en inventarios:
- Mapas de inventario de movimientos en masa
- Mapas basados en la ocurrencia de eventos
- Mapas de densidad de movimientos en masa
- Mapas de actividad de movimientos en masa
Validación
Aporte y Validación
Validación
Métodos basados en el conocimiento:
Mapeo directo
- Mapas geomorfológicos de amenaza
Mapeo indirecto
- Boolean Logic
- Fuzzy Logic
- Superposición de clases
- Evaluación espacial multi-criterio
Métodos Cualitativos
Métodos basados en datos:
Estadísticos bivariados
- Pesos de evidencia
- Valor de información
- Proporción de frecuencia etc.
Estadísticos multivariados
- Regresión logística
- Análisis discriminante
- Análisis cluster etc.
Redes neuronales
Métodos físicos:
Métodos estáticos
- Basado en pendiente infinita
- Basado en perfil
- Modelos 3-D
Métodos dinámicos
- Lluvia/pendiente
- Aceleración sísmica
Métodos Cuantitativos

Machine Learning




Supervisor
INPUT RAW DATA
Training data set
Desired Output
OUTPUT
IT'S A FACE
LABEL FACE
Algorithm
Processing
0100101010
101010010
0010010
001101001
00010
Model Trained
LABEL NON-FACE
Model Training
INPUT

Step 1: Modelo conceptual.

Step 2: Toma y procesamiento de datos.

Step 3: Análisis exploratorio de datos.

Step 4: Selección de variables (*Feature engineering*).

Step 5: Selección del algoritmo (*Training & Evaluation*).

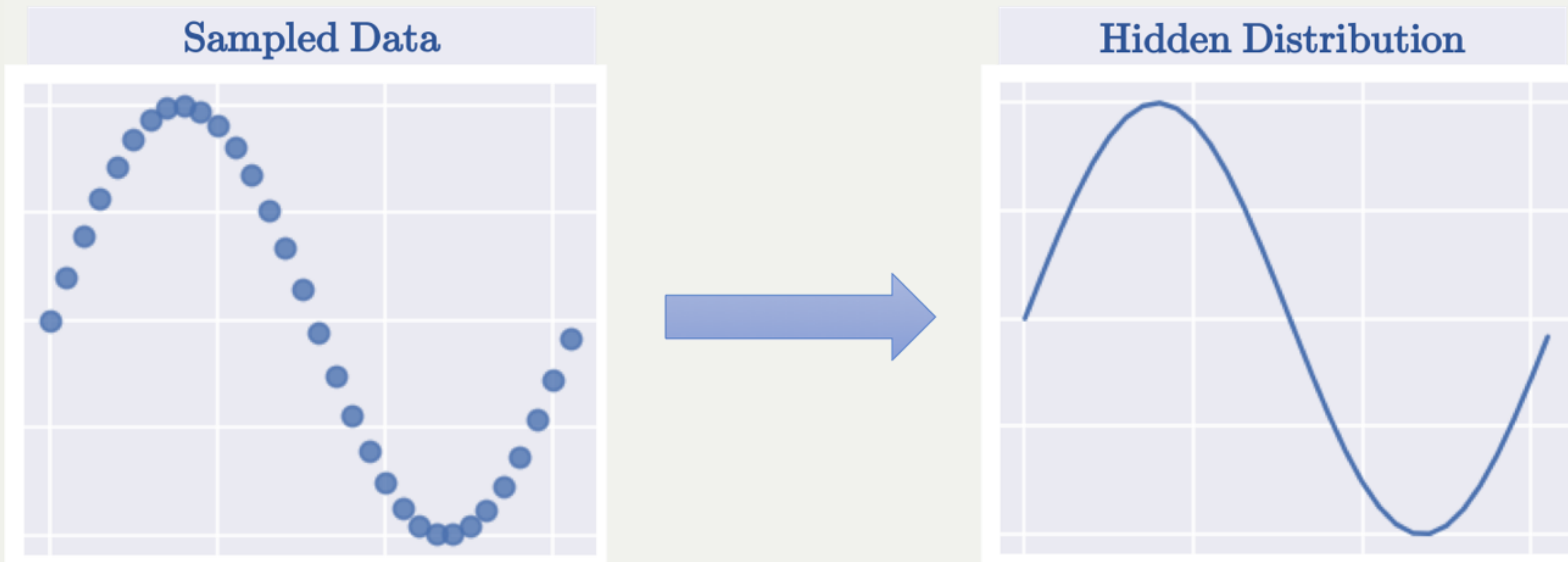
Step 6: Optimización del modelo (*Hyperparameters selection*).

Step 7: Predecir

Bias-Variance trade off

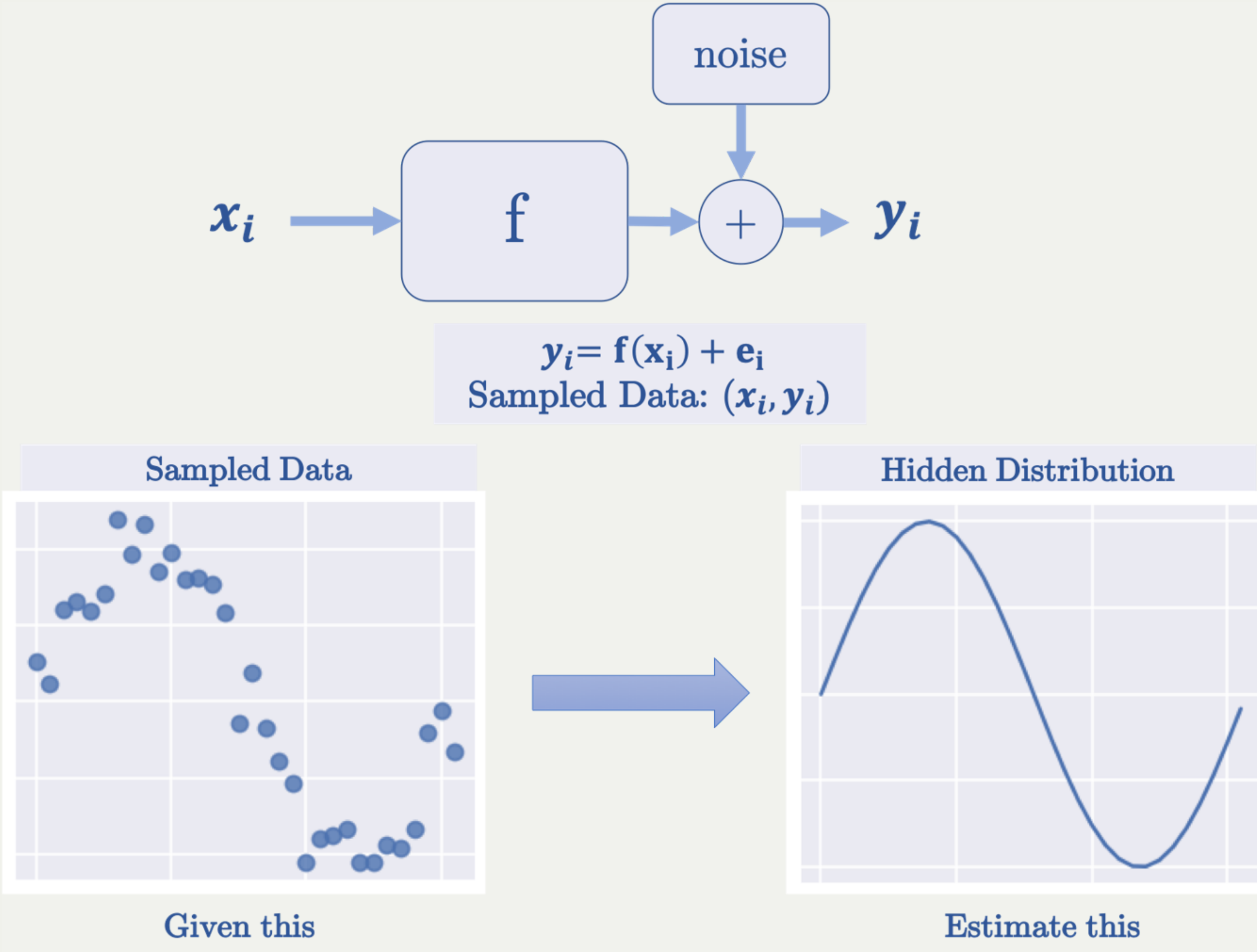
$$y_i = \mathbf{f}(\mathbf{x}_\mathrm{i})$$

Sampled Data: (x_i, y_i)

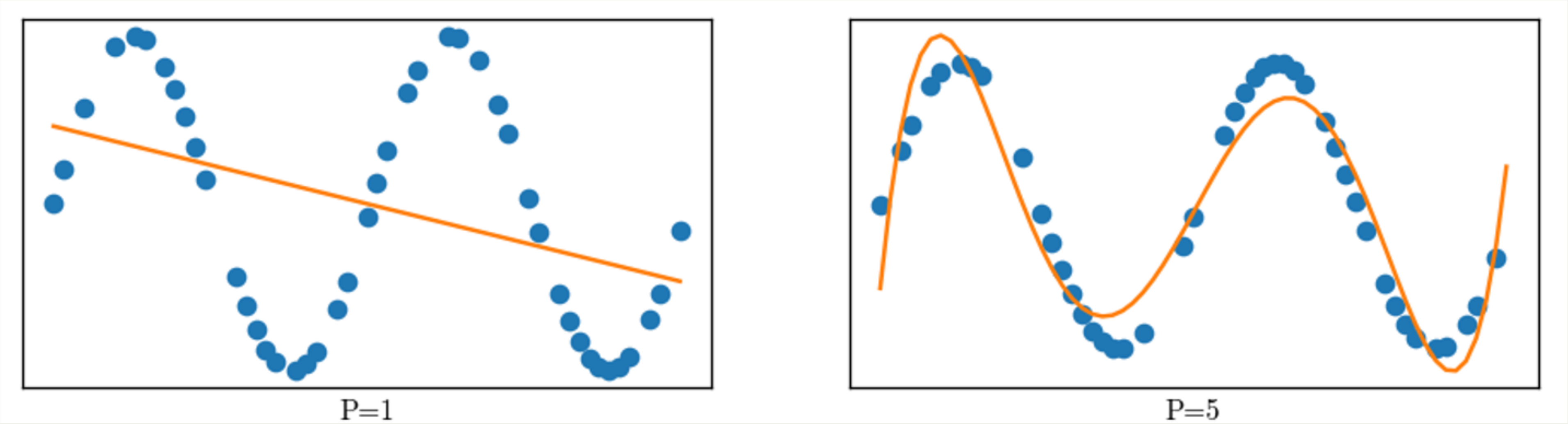


Given this

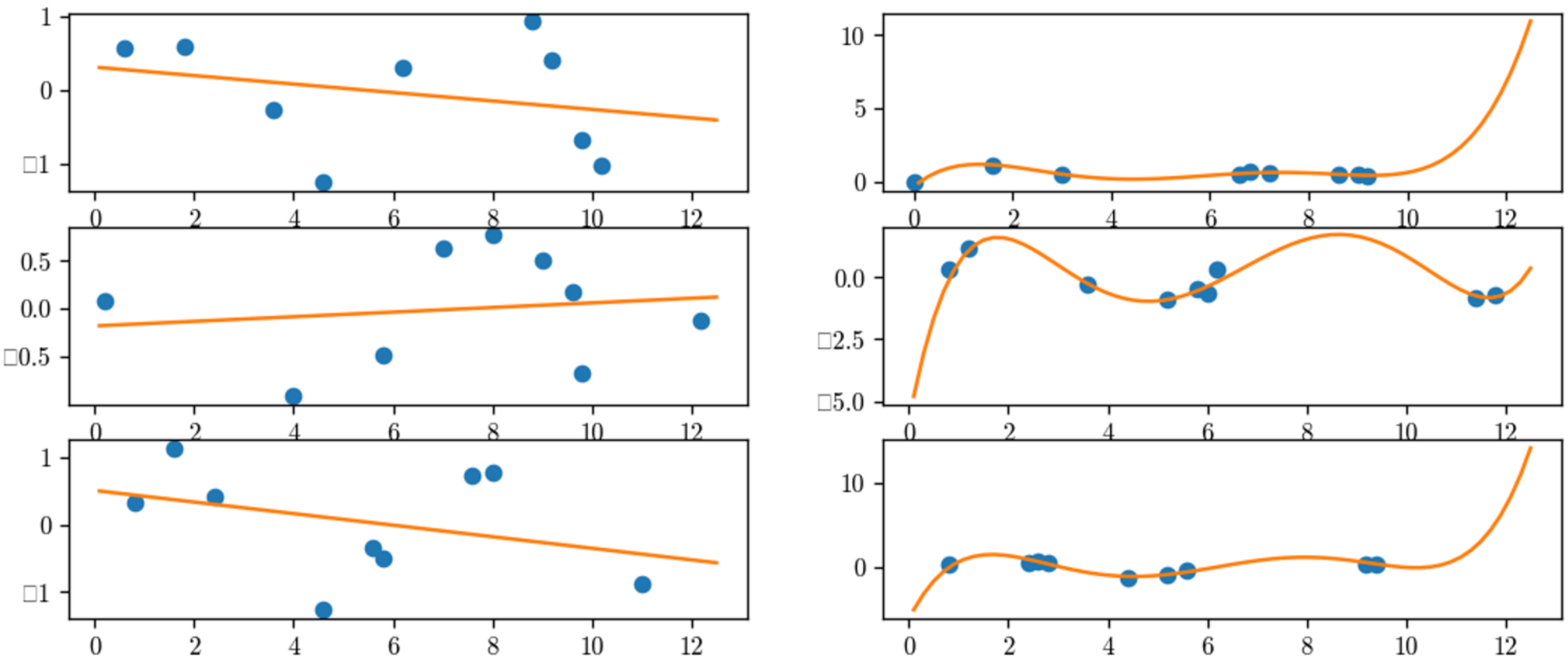
Estimate this


noise
x_i
f
+
y_i
$y_i = \mathbf{f}(\mathbf{x_i}) + \mathbf{e_i}$
Sampled Data: (x_i, y_i)
Sampled Data
Hidden Distribution
Given this
Estimate this

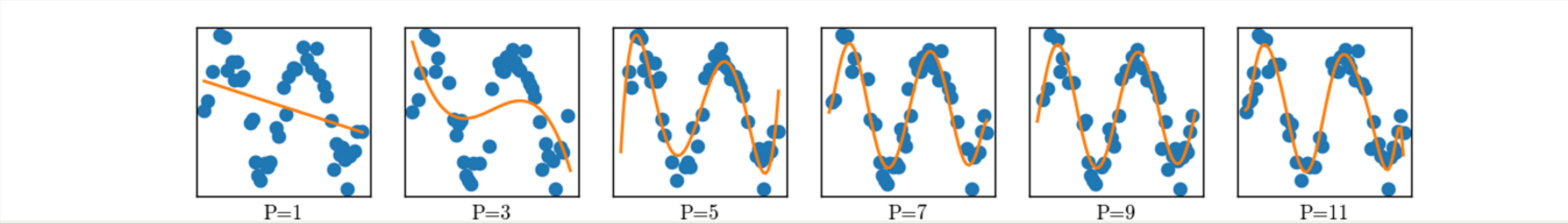
$$bias = \mathbb{E}[f'(x)] - f(x)$$

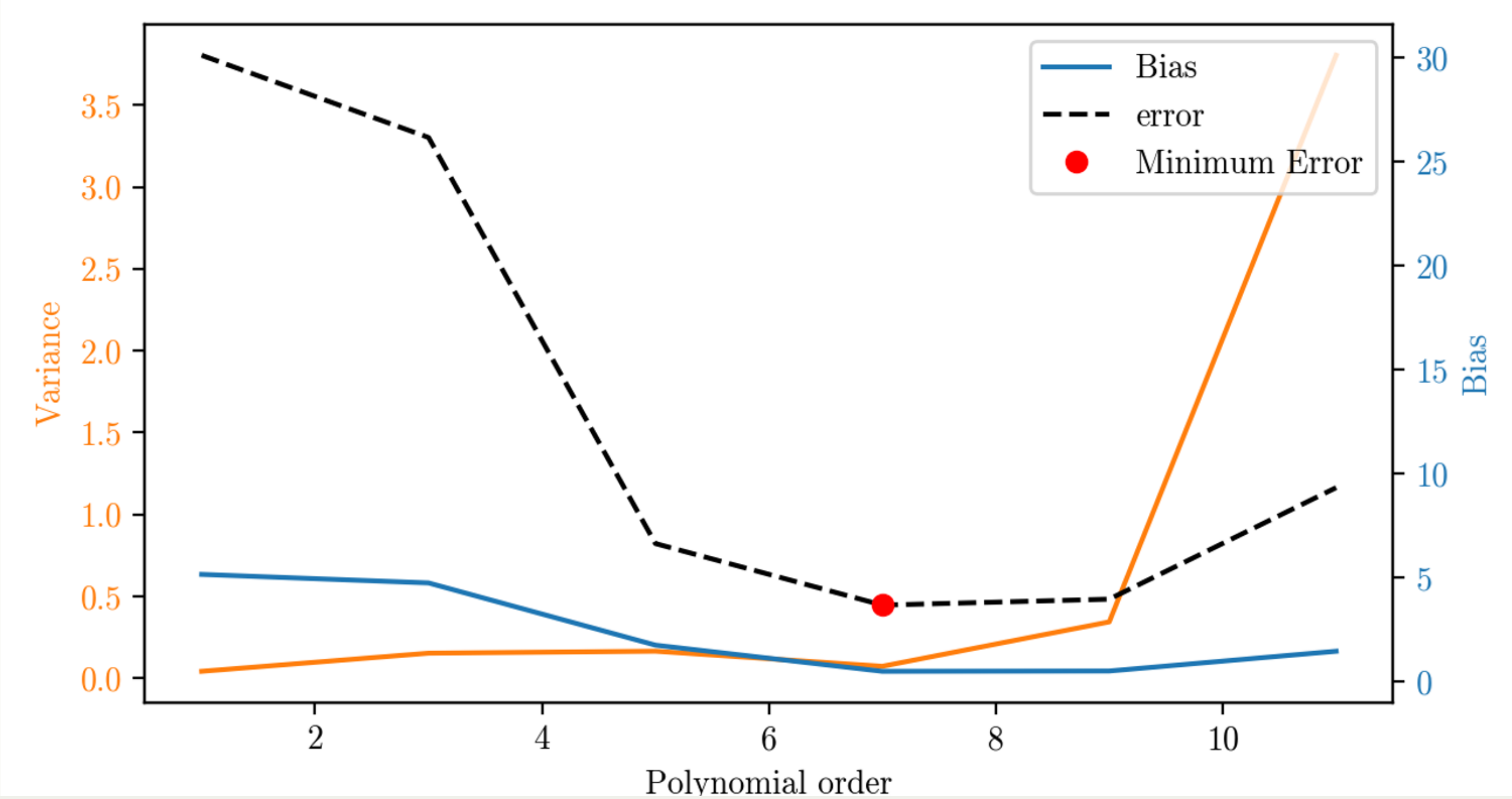


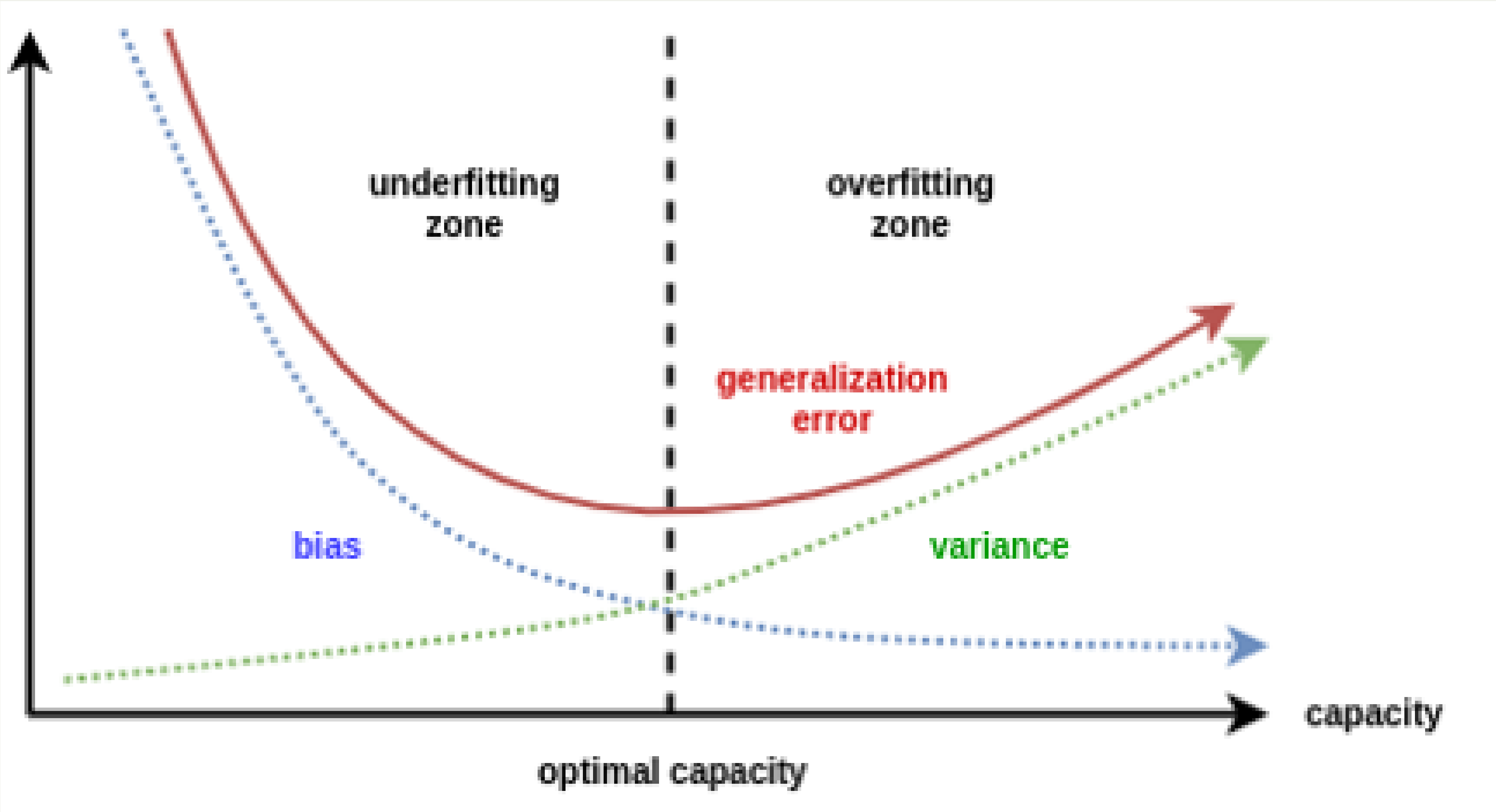
$$variance = \mathbb{E}\Big[\big(f'(x) - \mathbb{E}[f'(x)]\big)^2\Big]$$

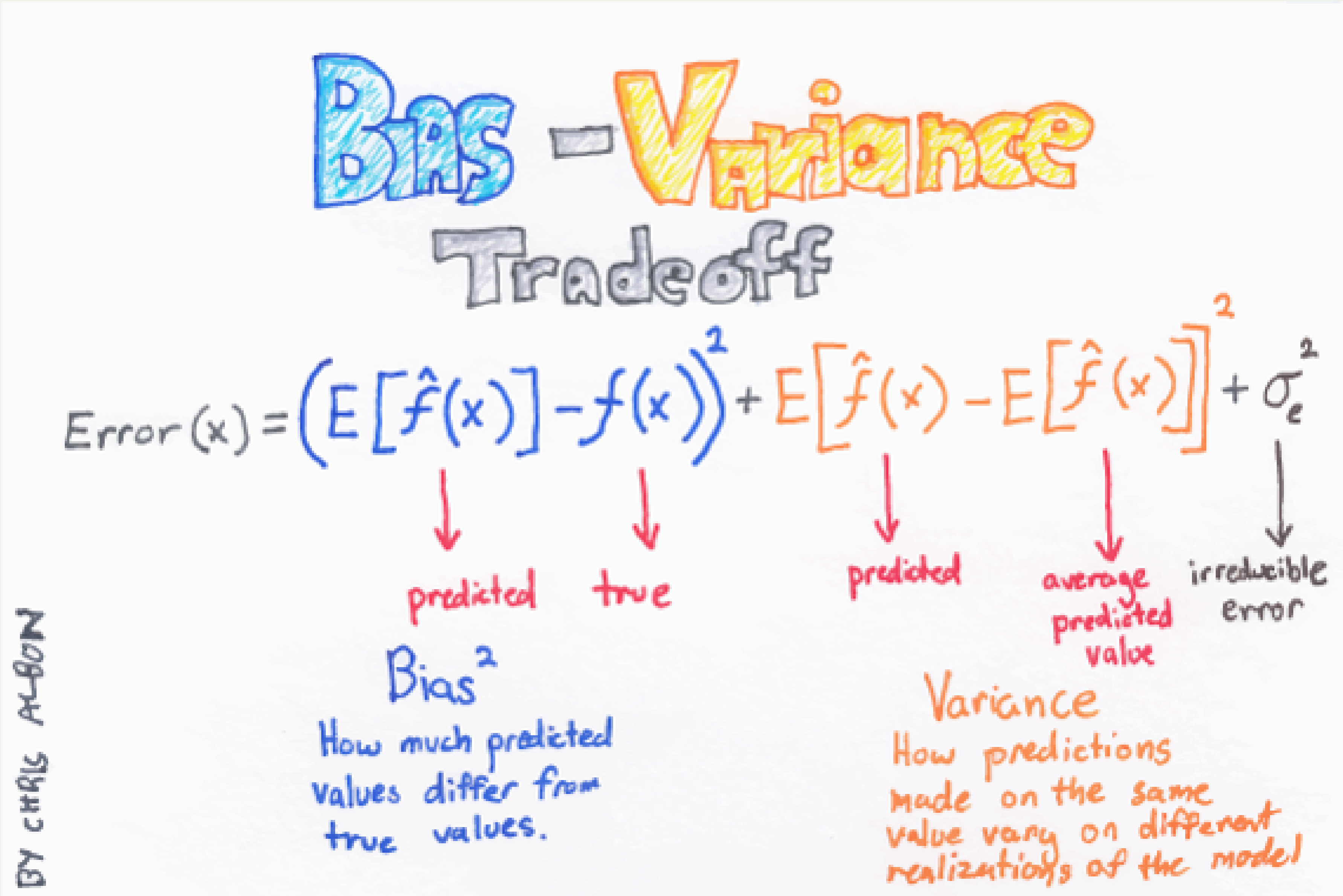


P=1 P=5

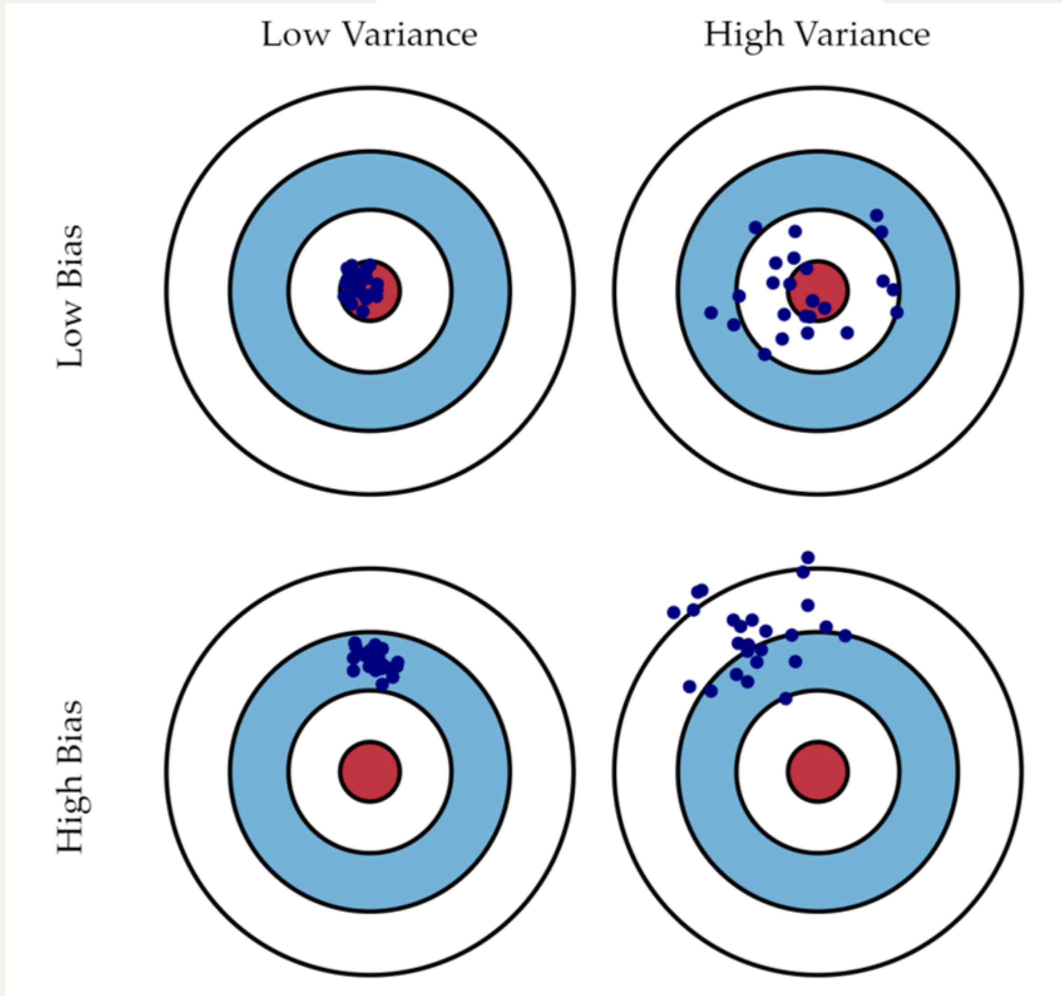

P=1
P=3
P=5
P=7
P=9
P=11


Bias
error
Minimum Error
Variance
Bias
Polynomial order
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
5
10
15
20
25
30
2
4
6
8
10


underfitting
zone
overfitting
zone
generalization
error
bias
variance
capacity
optimal capacity


Bias - Variance
Tradeoff
$Error(x) = \left(E[\hat{f}(x)] - f(x)\right)^2 + E\left[\hat{f}(x) - E[\hat{f}(x)]\right]^2 + \sigma_e^2$
predicted
true
predicted
average predicted value
irreducible error
Bias^2
How much predicted values differ from true values.
Variance
How predictions made on the same value vary on different realizations of the model
BY CHRIS ALBON

$$E\left[\left(\hat{f}(x) - E[\hat{f}(x)]\right)^2\right]$$

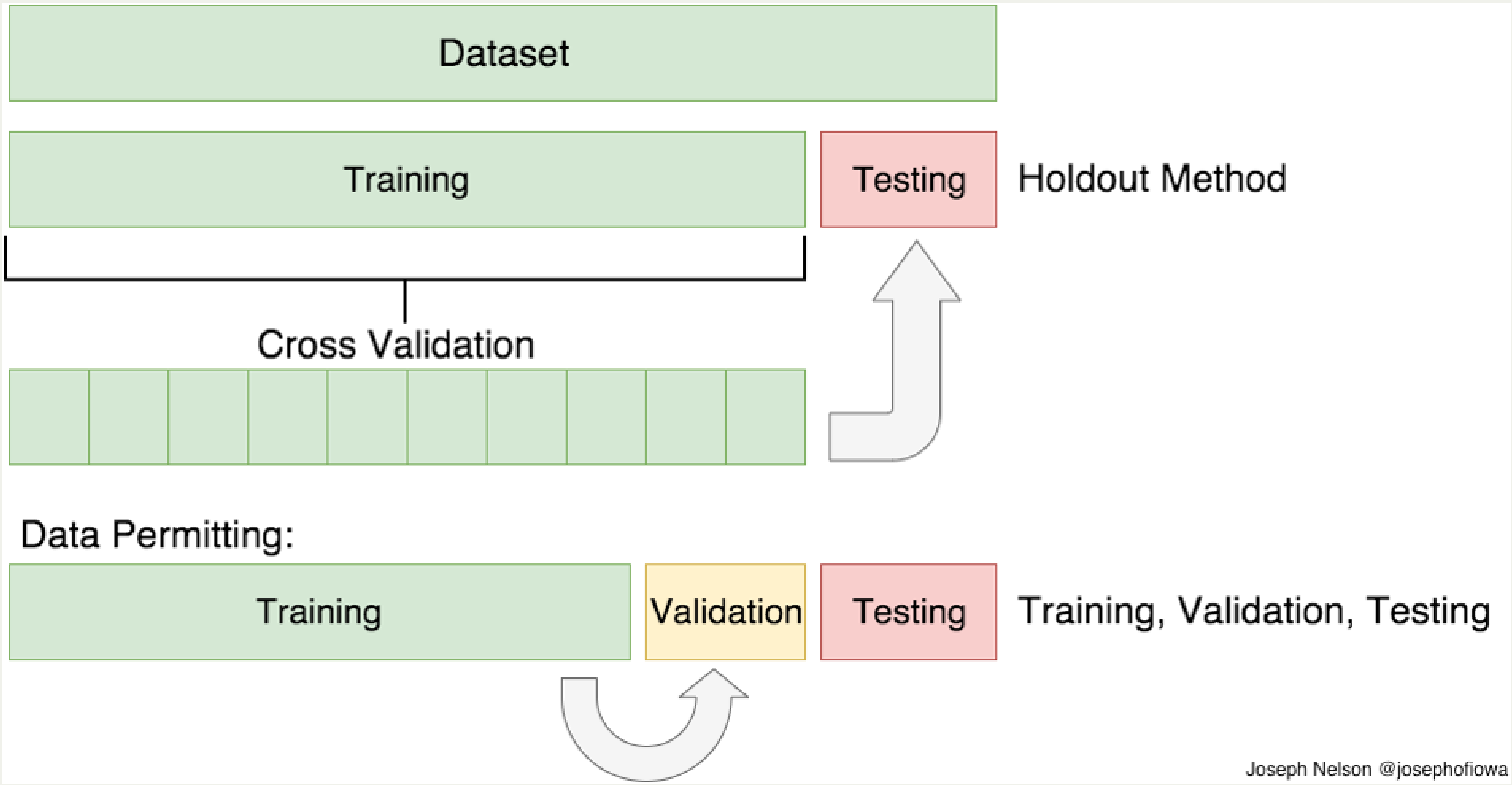


ACCURACY

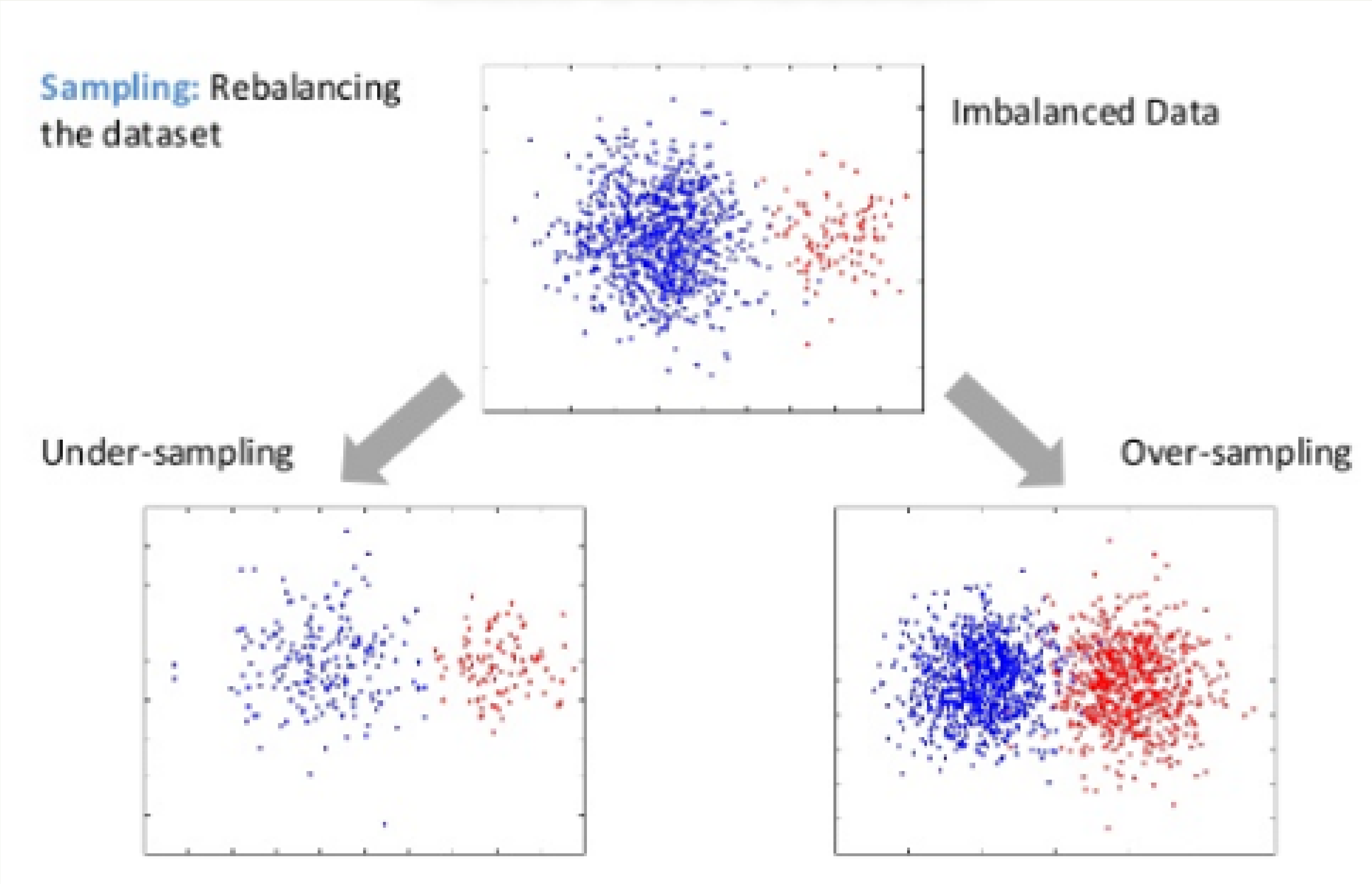
PRECISION

$$\left(E[\hat{f}(x)] - f(x)\right)^2$$

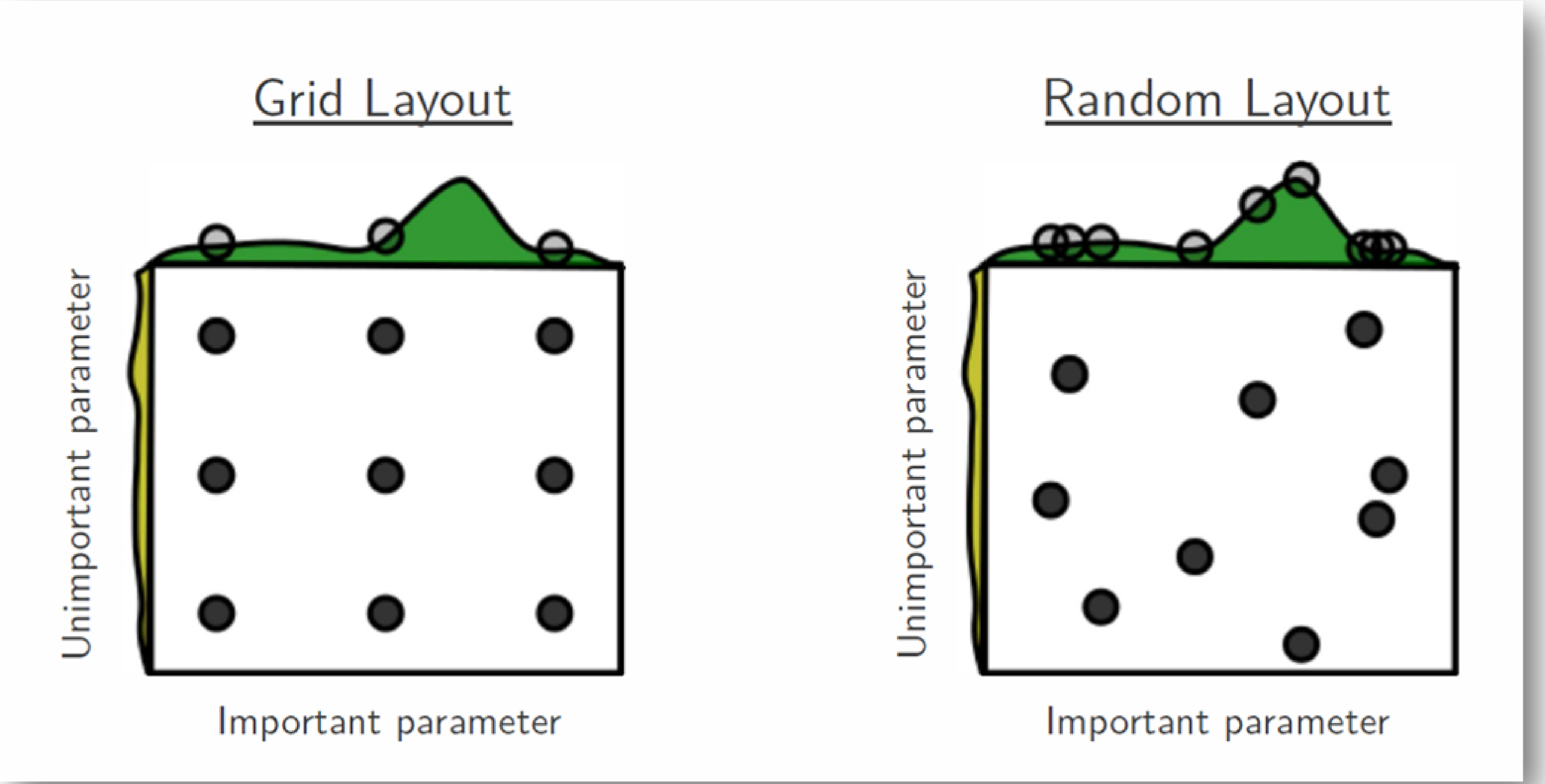
Entrenamiento & Validación



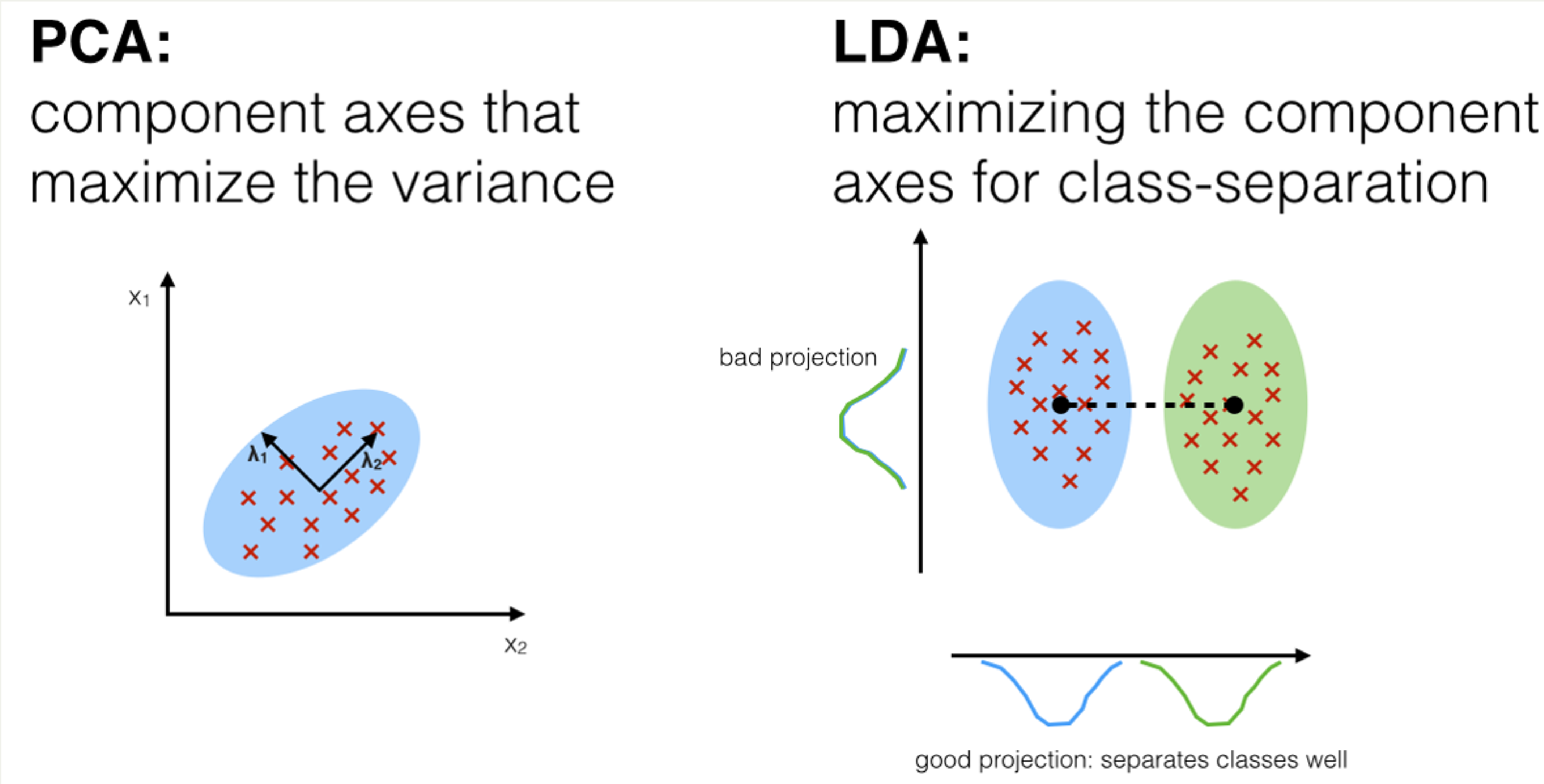
Datos Desbalanceados



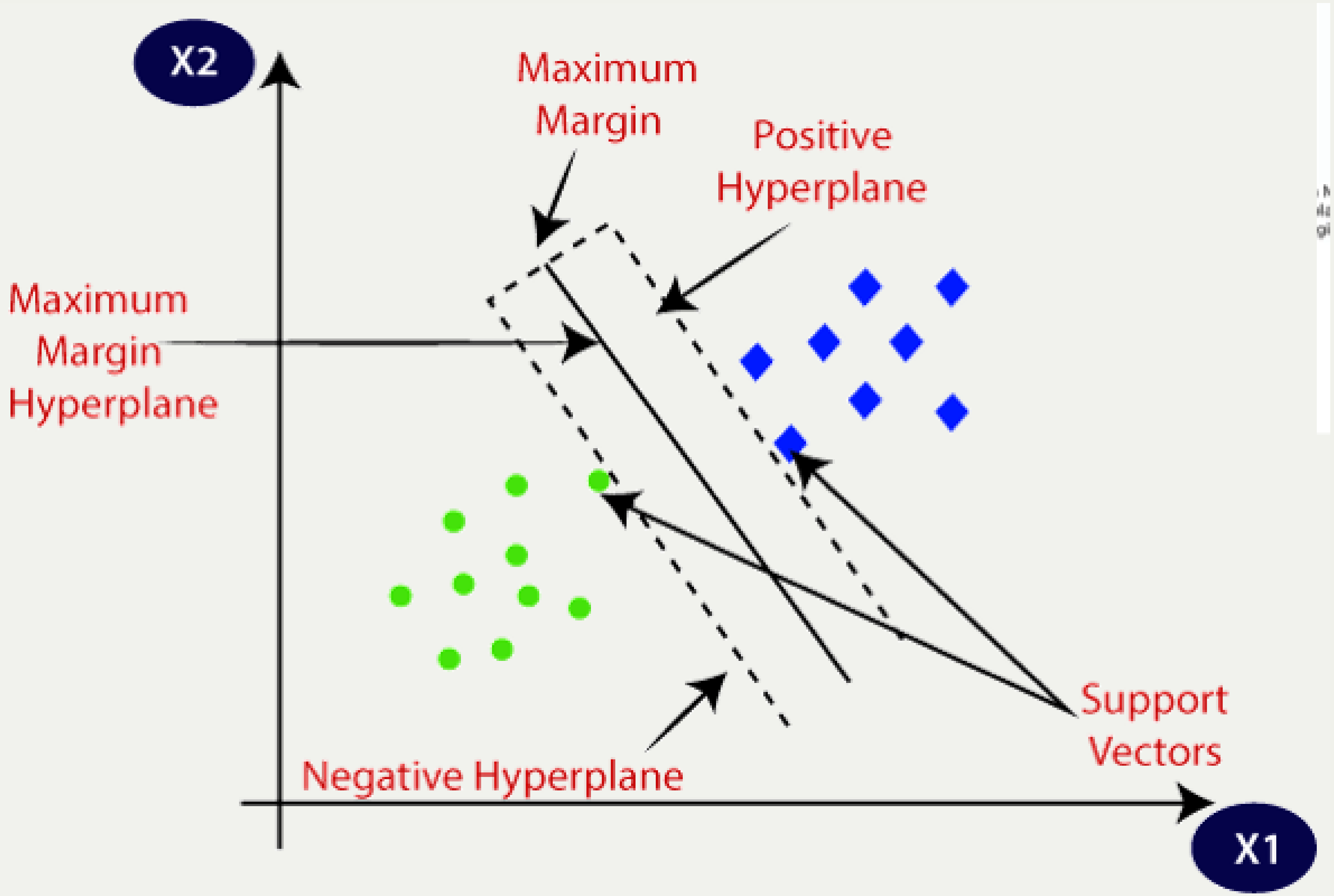
Optimización del modelo



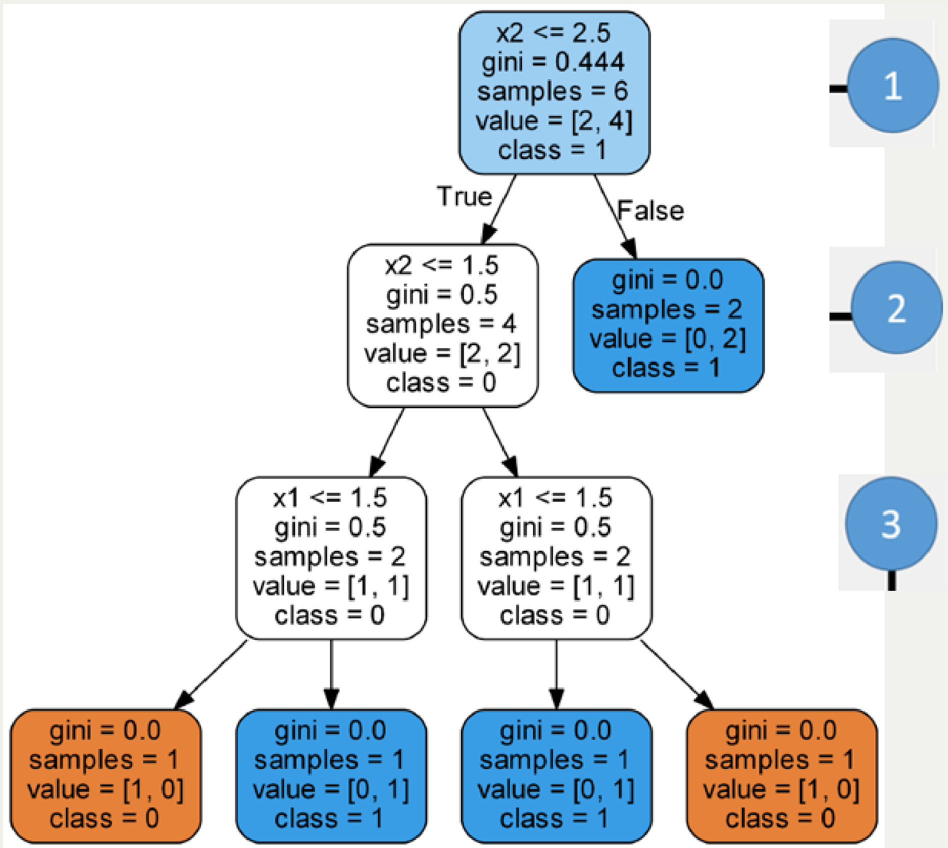
Análisis Discriminante Lineal

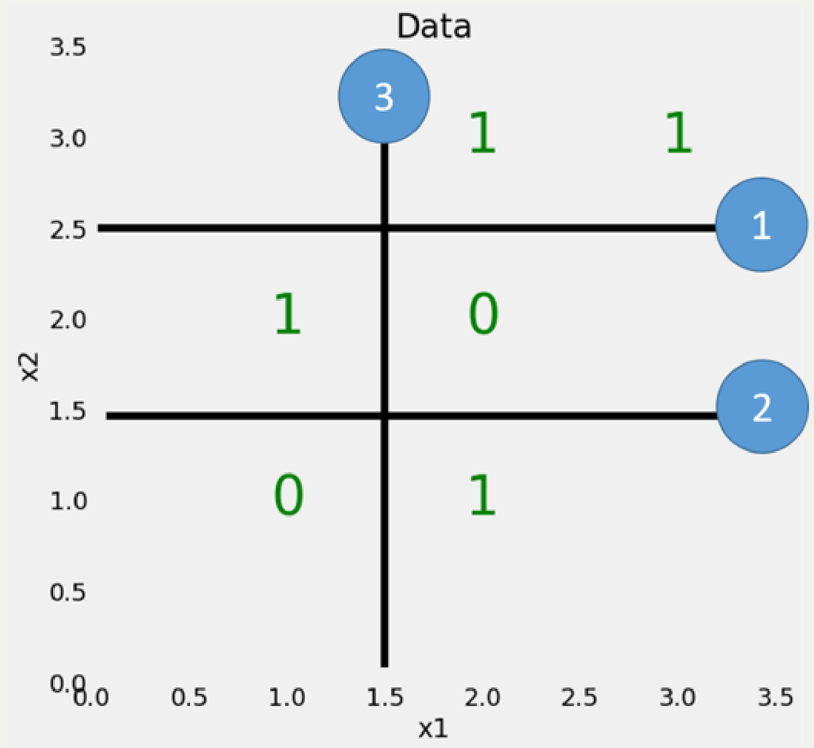


Support Vector Machine



Random forest





Credit Rating	Liability		
	Normal	High	Total
Excellent	3	1	4
Good	4	2	6
Poor	0	4	4
Total	7	7	14

Entropy = -p log2(p) – q log2(q),

Probabilidad del Si

Probabilidad del No

$$E(\,Liability) = -\frac{7}{14}\log_2\left(\frac{7}{14}\right) - \frac{7}{14}\log_2\left(\frac{7}{14}\right)$$

$$= -\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)$$

$$= 1$$

$$E(Liability \mid CR = Excellent) = -\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right) \approx 0.811$$

$$E(Liability \mid CR = Good) = -\frac{4}{6}\log_2\left(\frac{4}{6}\right) - \frac{2}{6}\log_2\left(\frac{2}{6}\right) \approx 0.918$$

$$E(Liability \mid CR = Poor) = -0\log_2(0) - \frac{4}{4}\log_2\left(\frac{4}{4}\right) = 0$$

Weighted Average:

$$E(Liability \mid CR) = \frac{4}{14}\times 0.811 + \frac{6}{14}\times 0.918 + \frac{4}{14}\times 0$$

Information Gain:

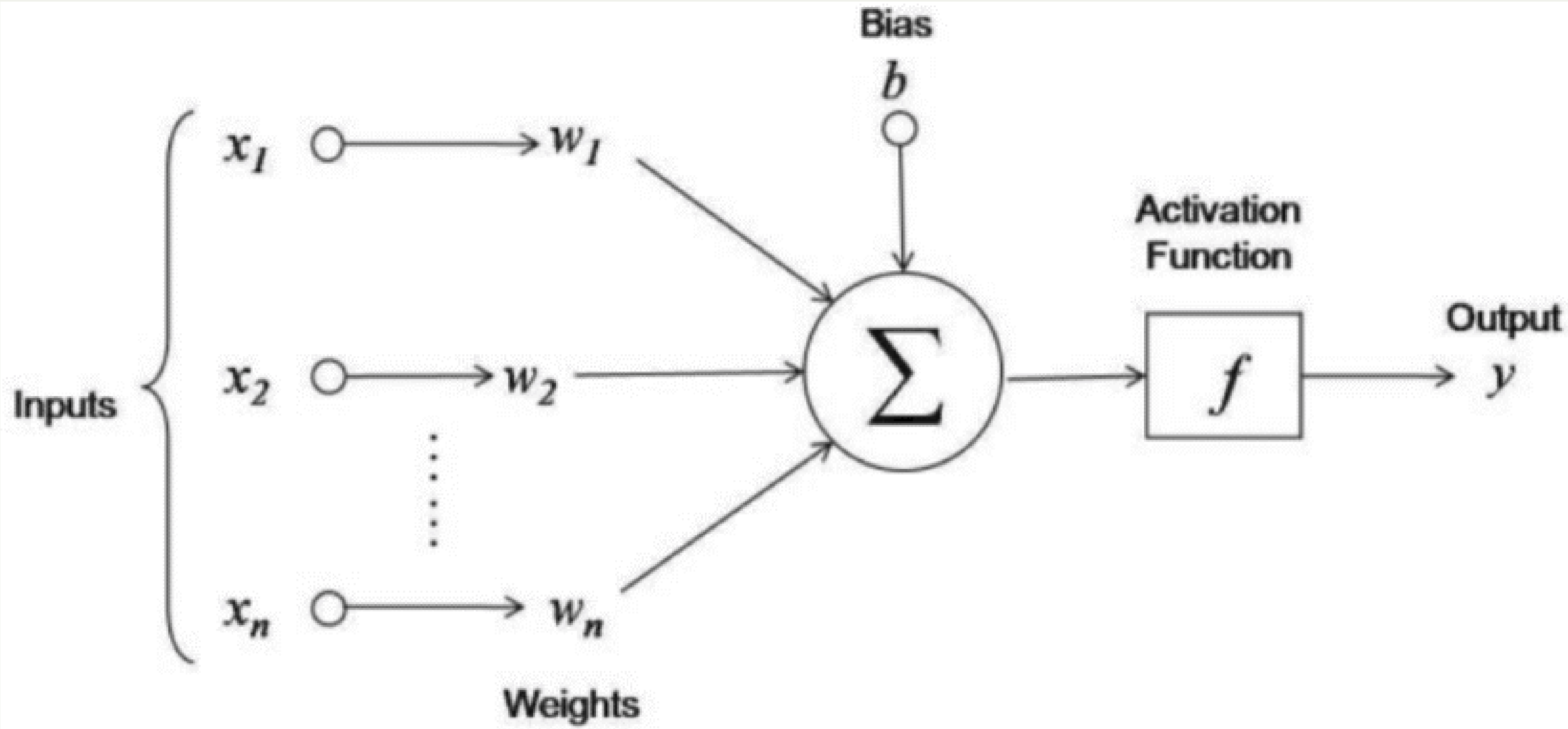
$$= 0.625$$

$$IG(Liability,\ CR) = E(Liability) - E(Liability \mid CR)$$

$$= 1 - 0.625$$

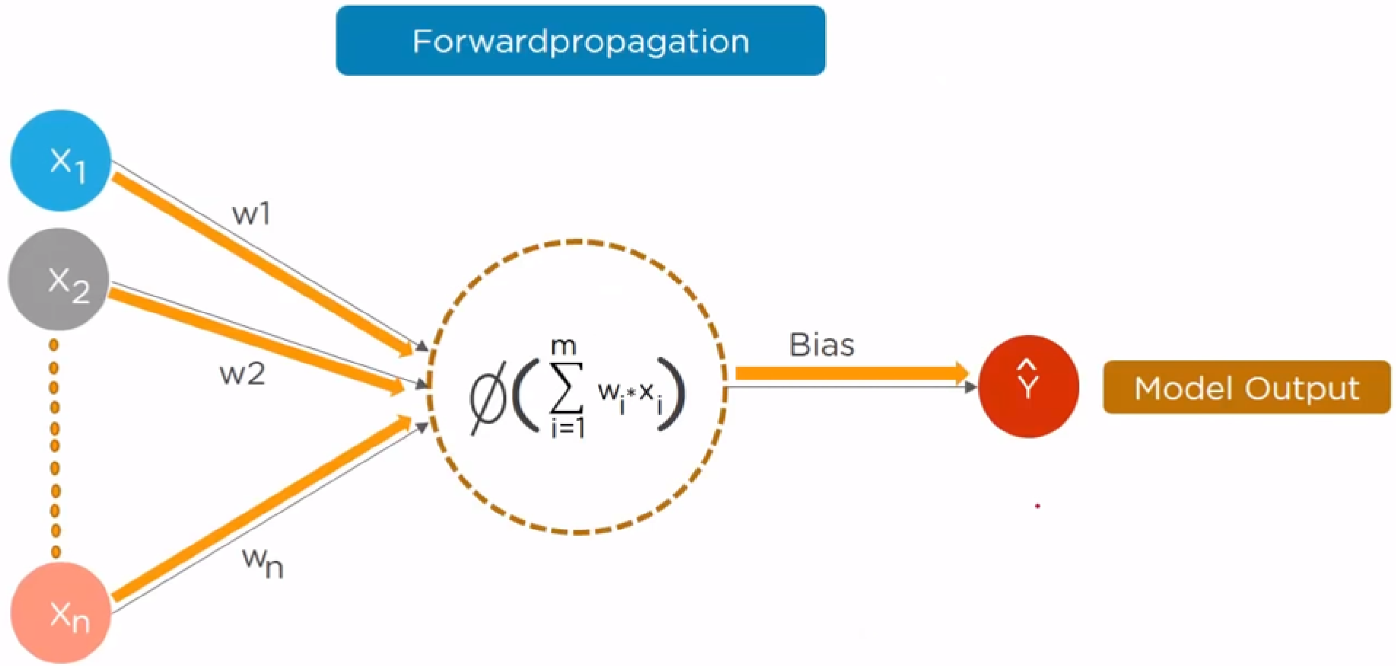
$$= 0.375$$

Redes Neuronales Artificiales



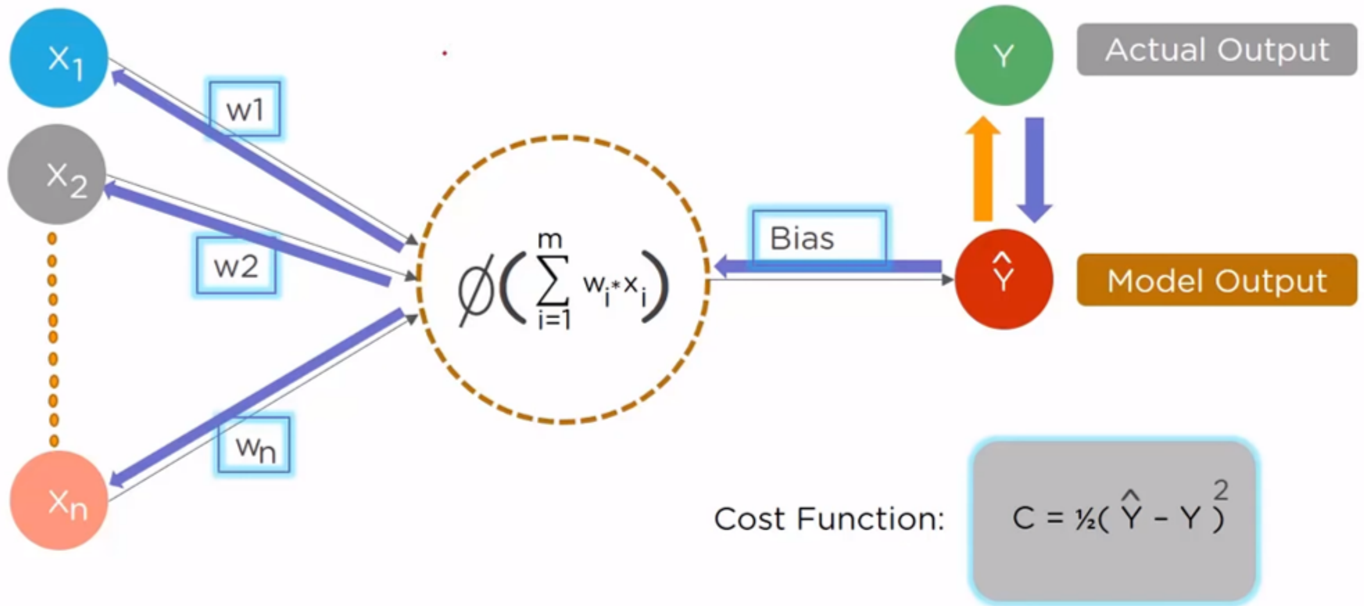
Working of a Neural Network

Lets consider a simple neural network



Cost Function

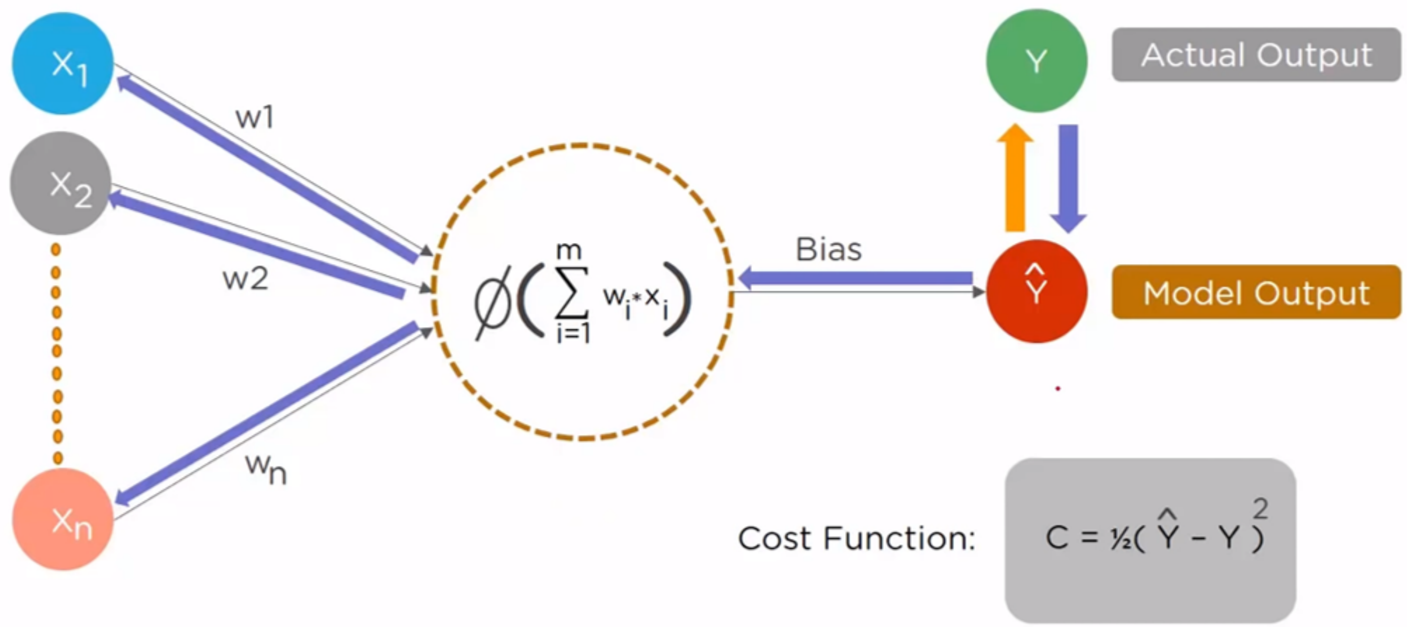
The *Cost* value is actually the difference between the neural nets predicted output and the actual output from a set of labelled training data. The least cost value is obtained by making adjustments to the weights and biases iteratively throughout the training process.



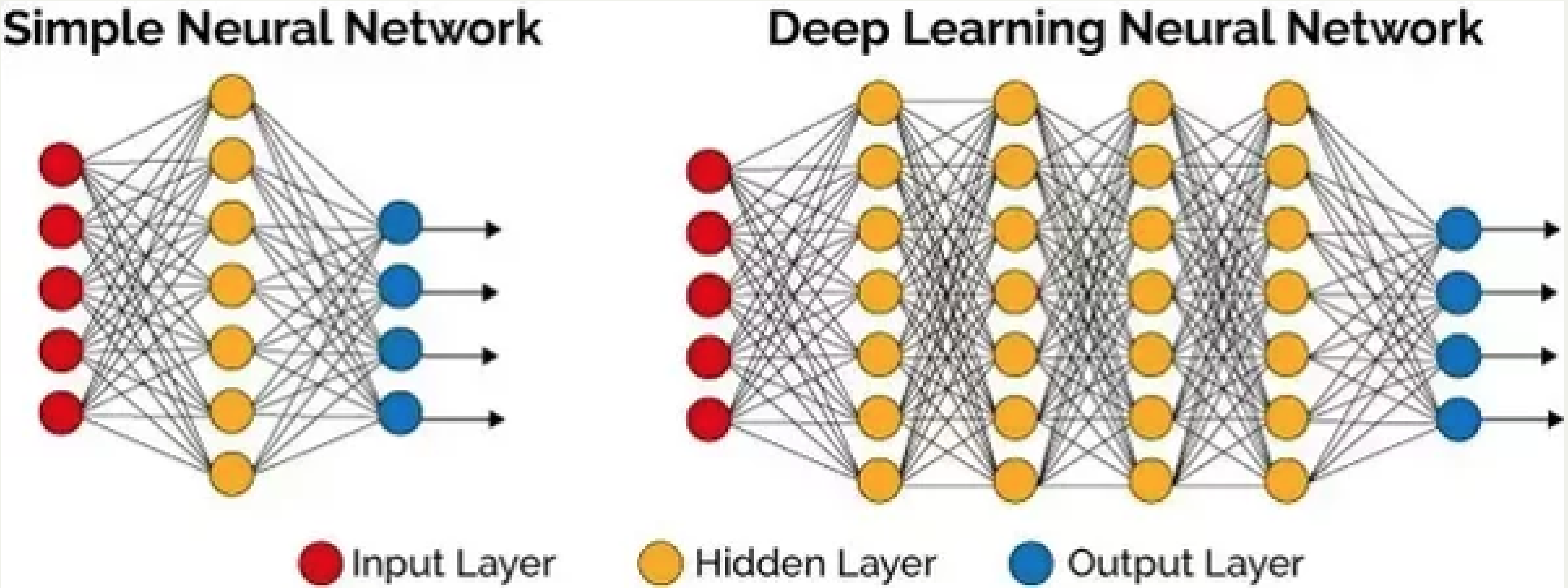
simplilearn

Working of a Neural Network

After training the Neural Network, it uses *Backpropagation* method to improve the performance of the network. Cost Function helps to reduce the error rate.



simplilearn


Simple Neural Network
Deep Learning Neural Network
Input Layer
Hidden Layer
Output Layer