

The background of the slide is a blurred image of a financial market display. It features various stock indices and their values, such as 'OMX COPENHAGEN 25 INDEX' with a value of 1172.94, 'OMX RIGA GI' with 984.13, and 'OMX ICELAND 8' with 6230.9. There are also line charts showing market trends. The overall color scheme is dominated by blue and red, typical of financial data visualizations.

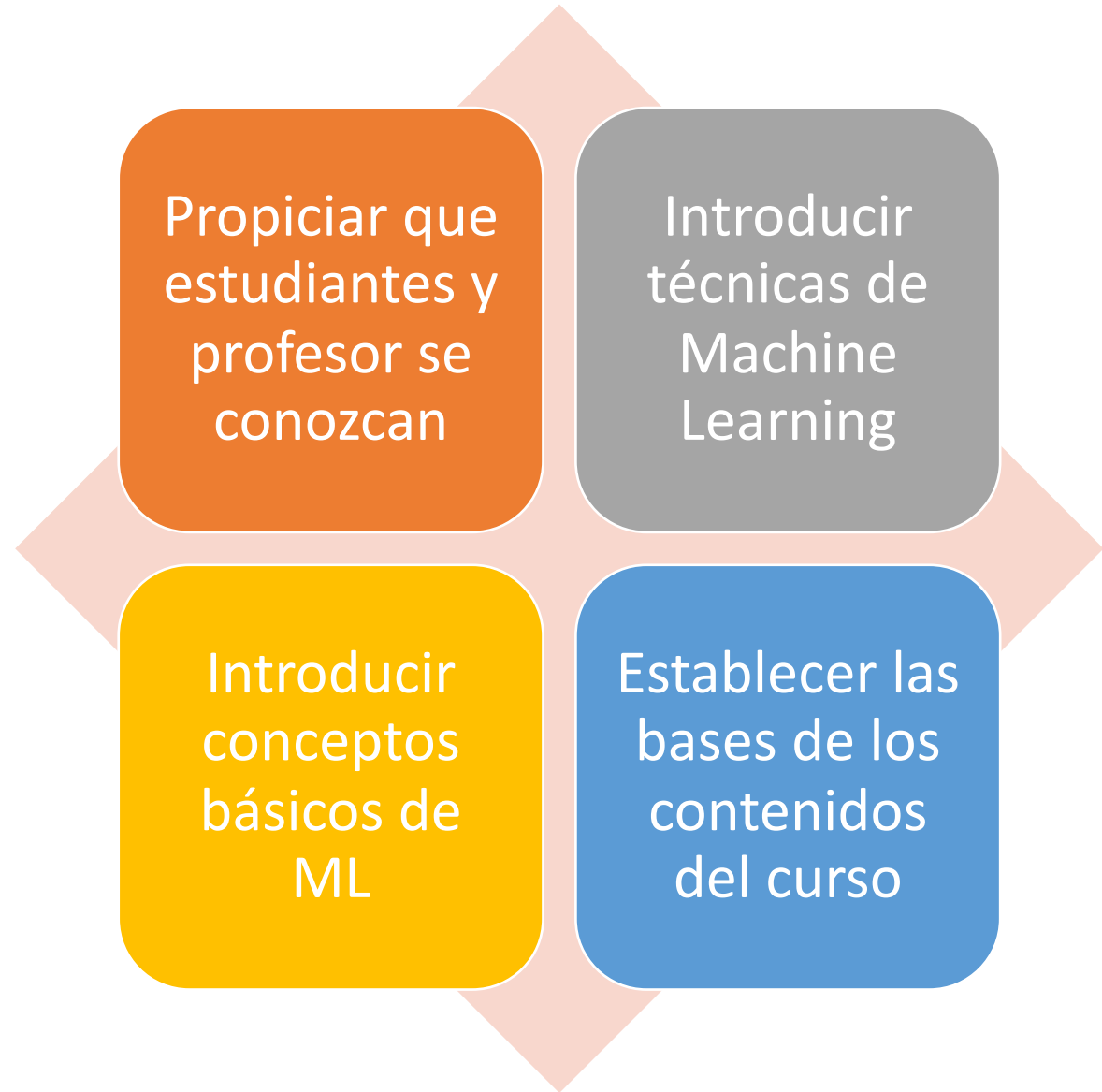
Inteligencia Artificial LFIS419

Clase 4: Cpnceptos fundamentales para machine learning

Profesor: Jorge Arevalo (jorge.arevalo@uv.cl jab@meteo.uv.cl)

Martes 4 de abril de 2023

Objetivos de la sesión



Modelo

- Estructura matemática típicamente representada mediante código computacional capaz de estimar el valor de una variable (categórica o continua) a partir de predictores.
- Todos los modelos tienen error (las observaciones también)

Error (residuos, pérdida, etc.) de un modelo

- Desviación del valor estimado respecto del esperado.
- Se puede medir de múltiples maneras (métricas).
- Típicamente se calcula sobre un conjunto de estimaciones/observaciones
- El error depende del modelo, pero también de los datos de entrada y de los parámetros del modelo.

Métricas: MSE

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

Error cuadrático medio (MSE)

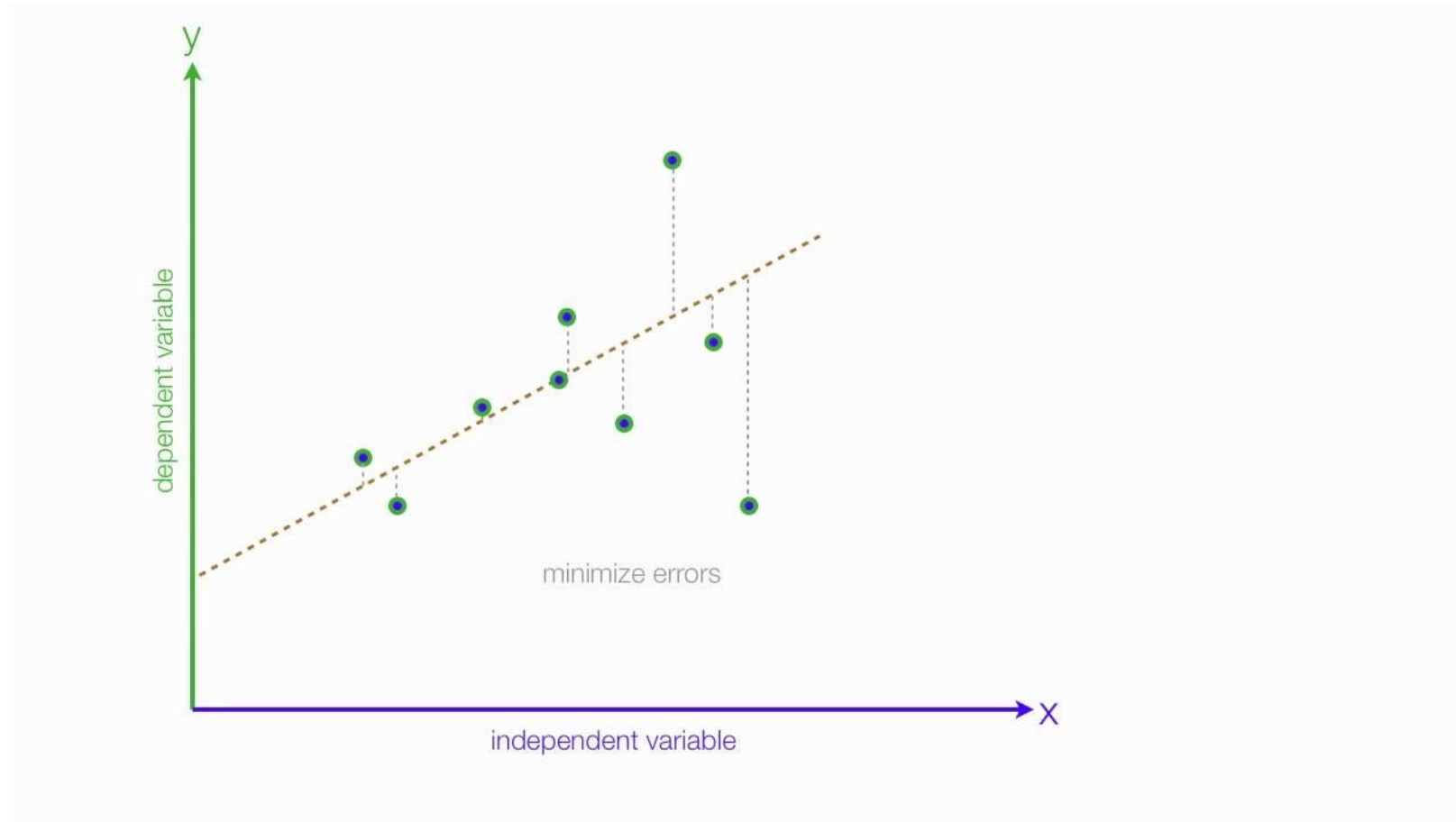
donde y_i es el resultado real esperado y \hat{y}_i es la predicción del modelo.

Métricas: MAE

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

Error absoluto medio (MAE)

MAE vs RMSE



Métricas: coeficiente de determinación

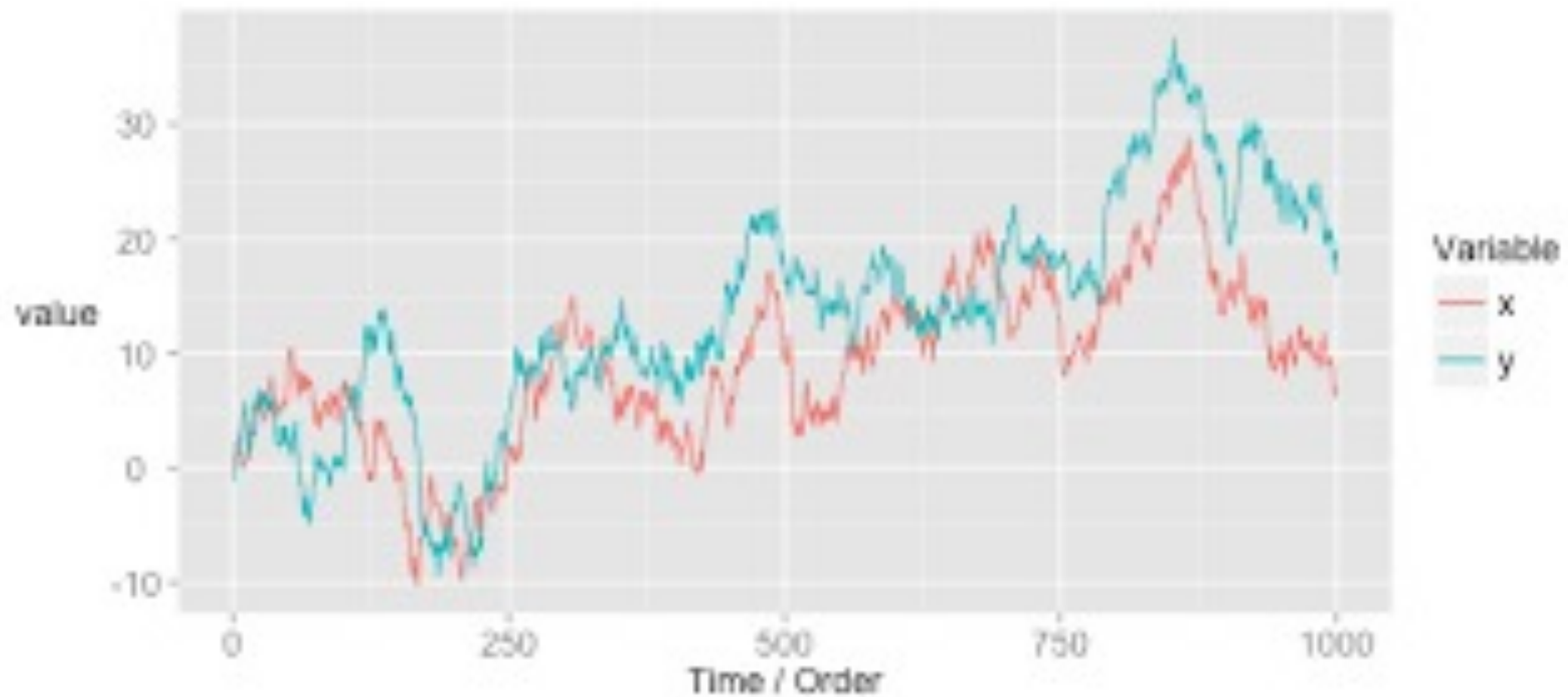
$$R^2 = 1 - \frac{\text{MSE}(\text{model})}{\text{MSE}(\text{baseline})}$$

Métricas de Regresión en aprendizaje automatico: R al cuadrado (R^2)

$$\text{MSE}(\text{baseline}) = \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2$$

Métricas de Regresión en aprendizaje automatico: MSE línea base

Correlación



Performance

$$\text{Precision} = \frac{\text{True}_{\text{positive}}}{\text{True}_{\text{positive}} + \text{False}_{\text{positive}}}$$

$$\text{Recall} = \frac{\text{True}_{\text{positive}}}{\text{True}_{\text{positive}} + \text{False}_{\text{negative}}}$$

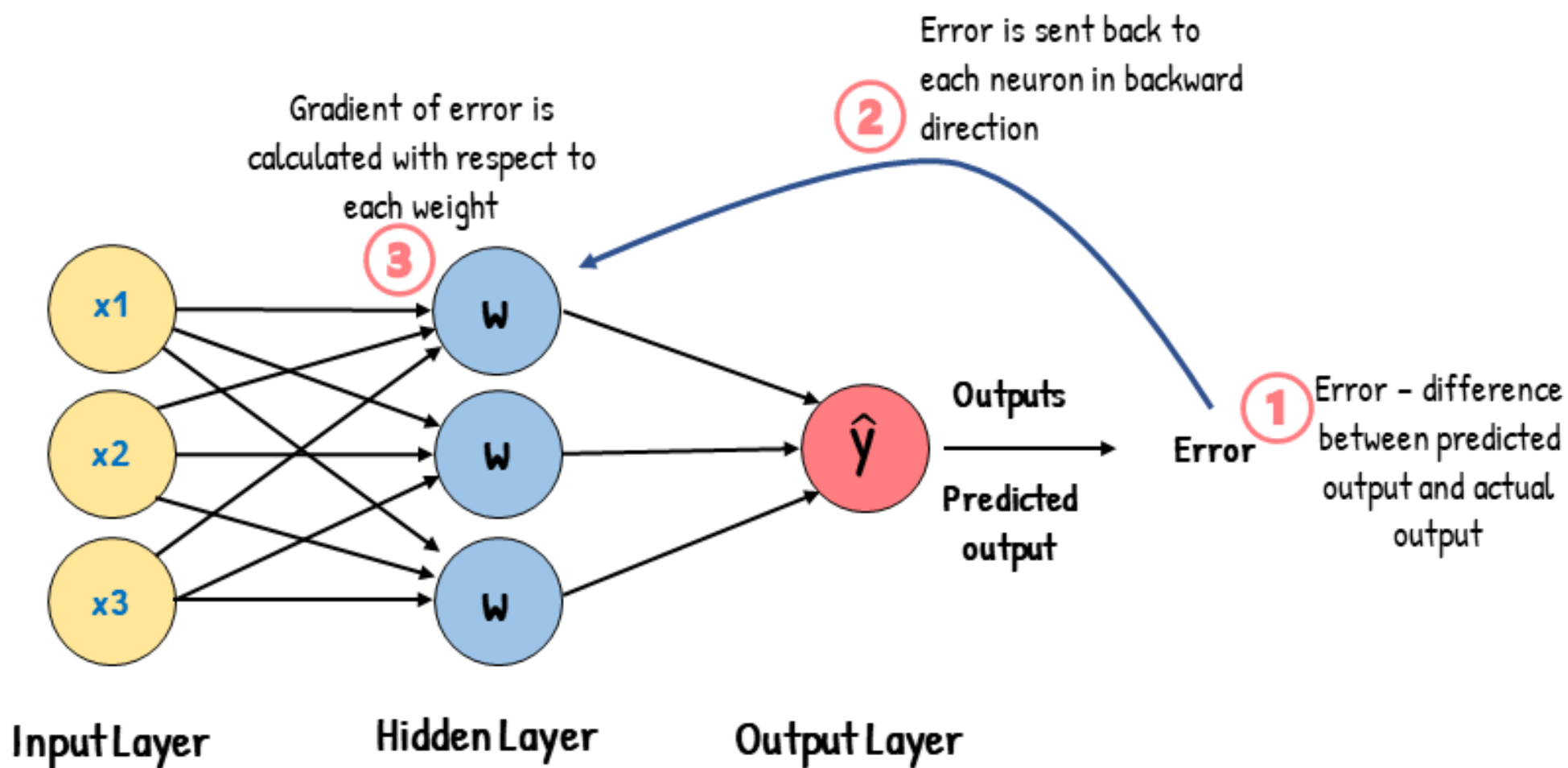
$$F_1 = \frac{2}{\frac{1}{\text{precision}} + \frac{1}{\text{recall}}}$$

$$\text{Accuracy} = \frac{\text{True}_{\text{positive}} + \text{True}_{\text{negative}}}{\text{True}_{\text{positive}} + \text{True}_{\text{negative}} + \text{False}_{\text{positive}} + \text{False}_{\text{negative}}}$$

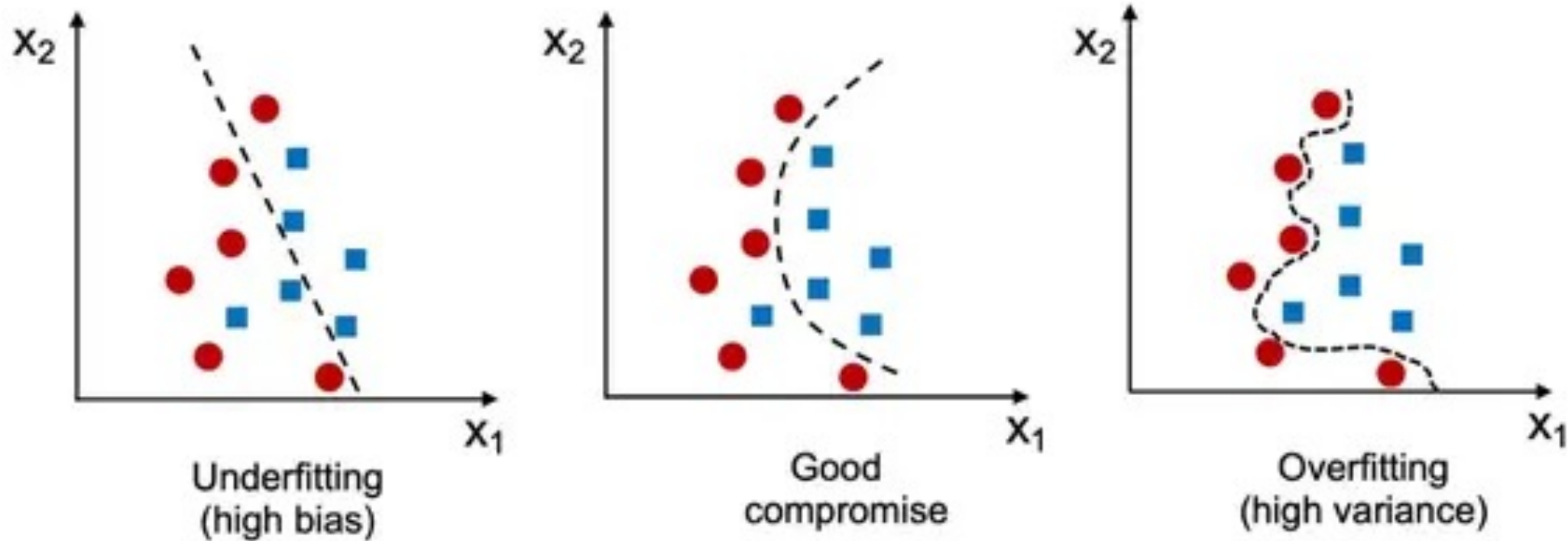
		Predicted	
		Positive	Negative
Ground-Truth	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

<https://blog.paperspace.com/deep-learning-metrics-precision-recall-accuracy/>

Backpropagation

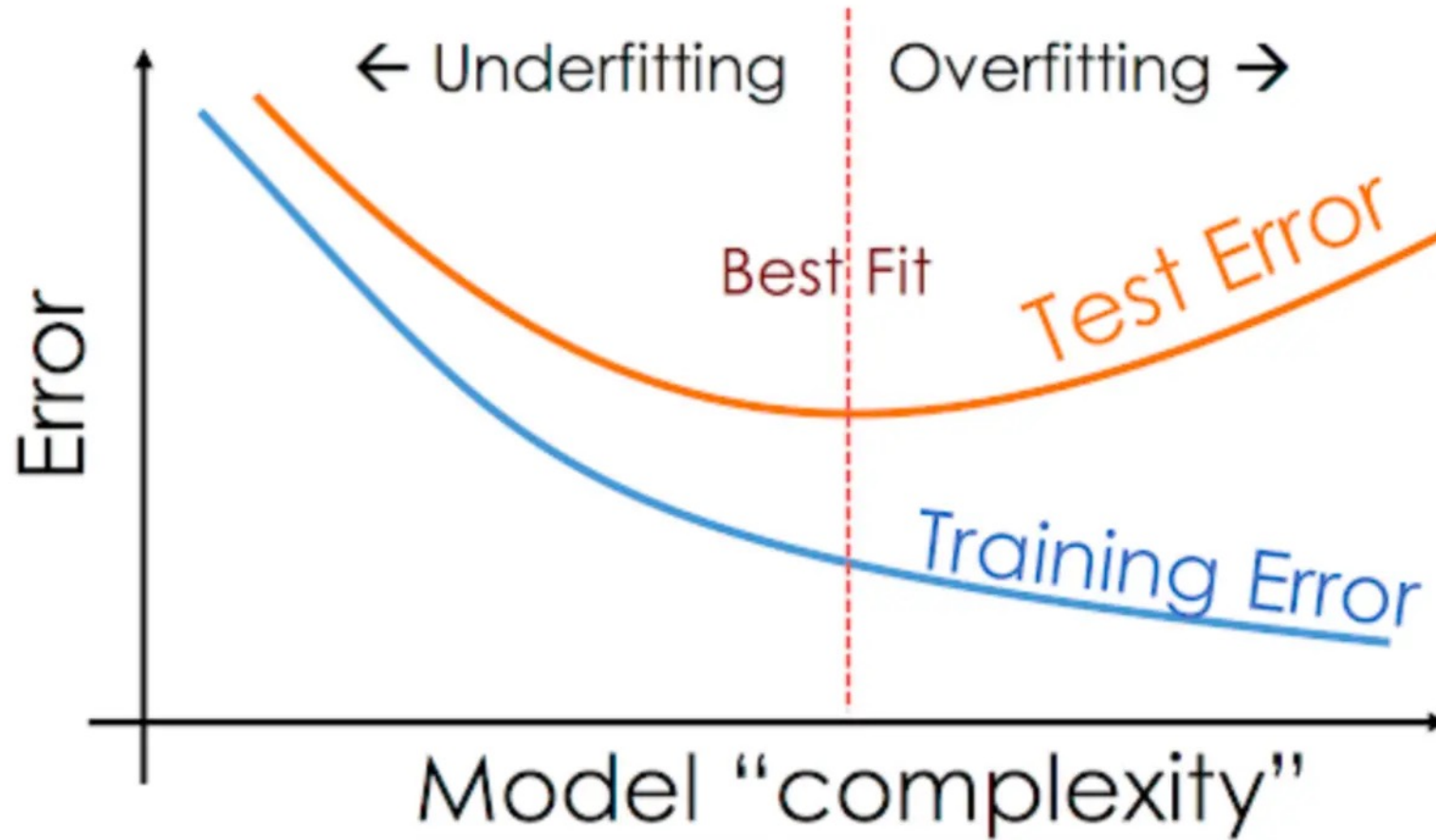


¿Qué es el sobreajuste?



<https://vitalflux.com/overfitting-underfitting-concepts-interview-questions/>

Sobreajuste

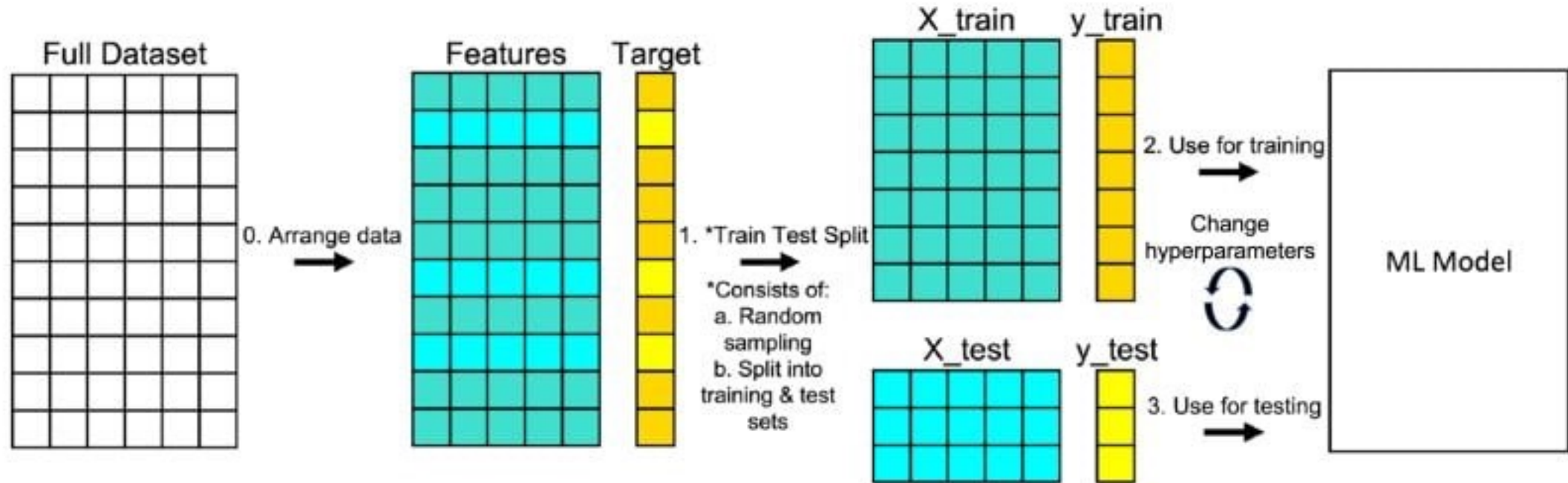


<https://vitalflux.com/overfitting-underfitting-concepts-interview-questions/>

Validation techniques

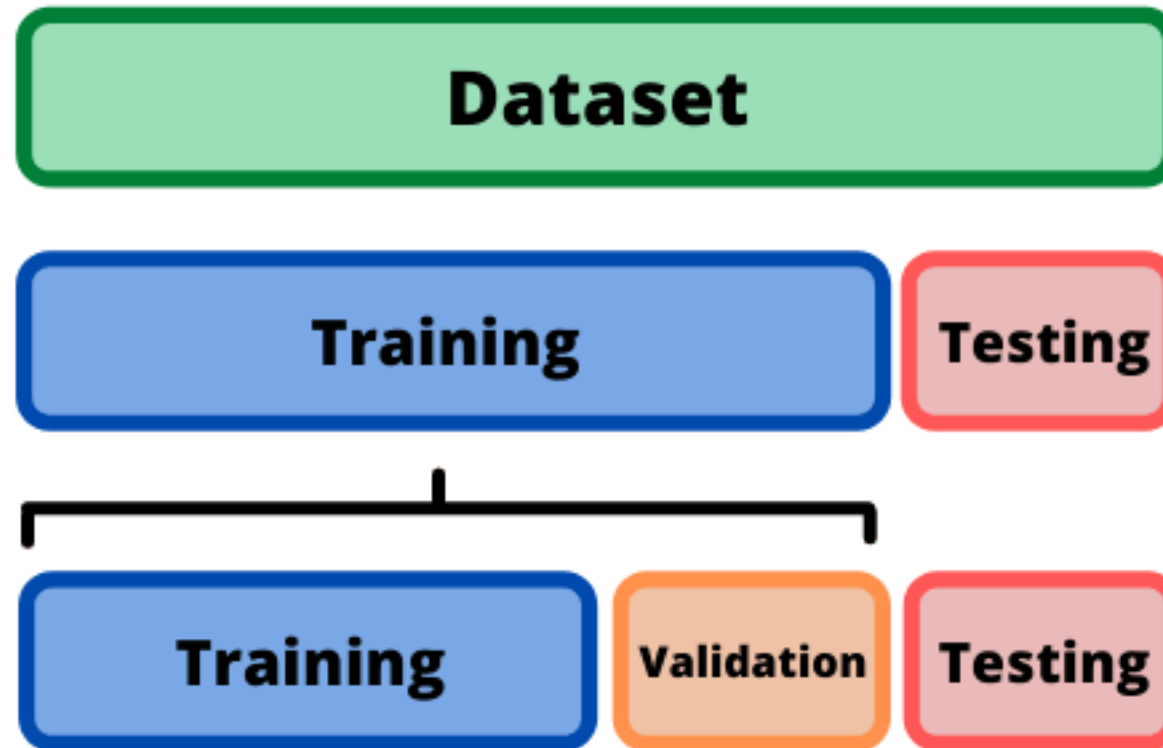
- Train and Test Split
- Train, Validation and Test Split.
- K-Fold Cross-Validation
- Nested Cross-Validation
- Random Subsampling
- Bootstrapping

Train and Test Split



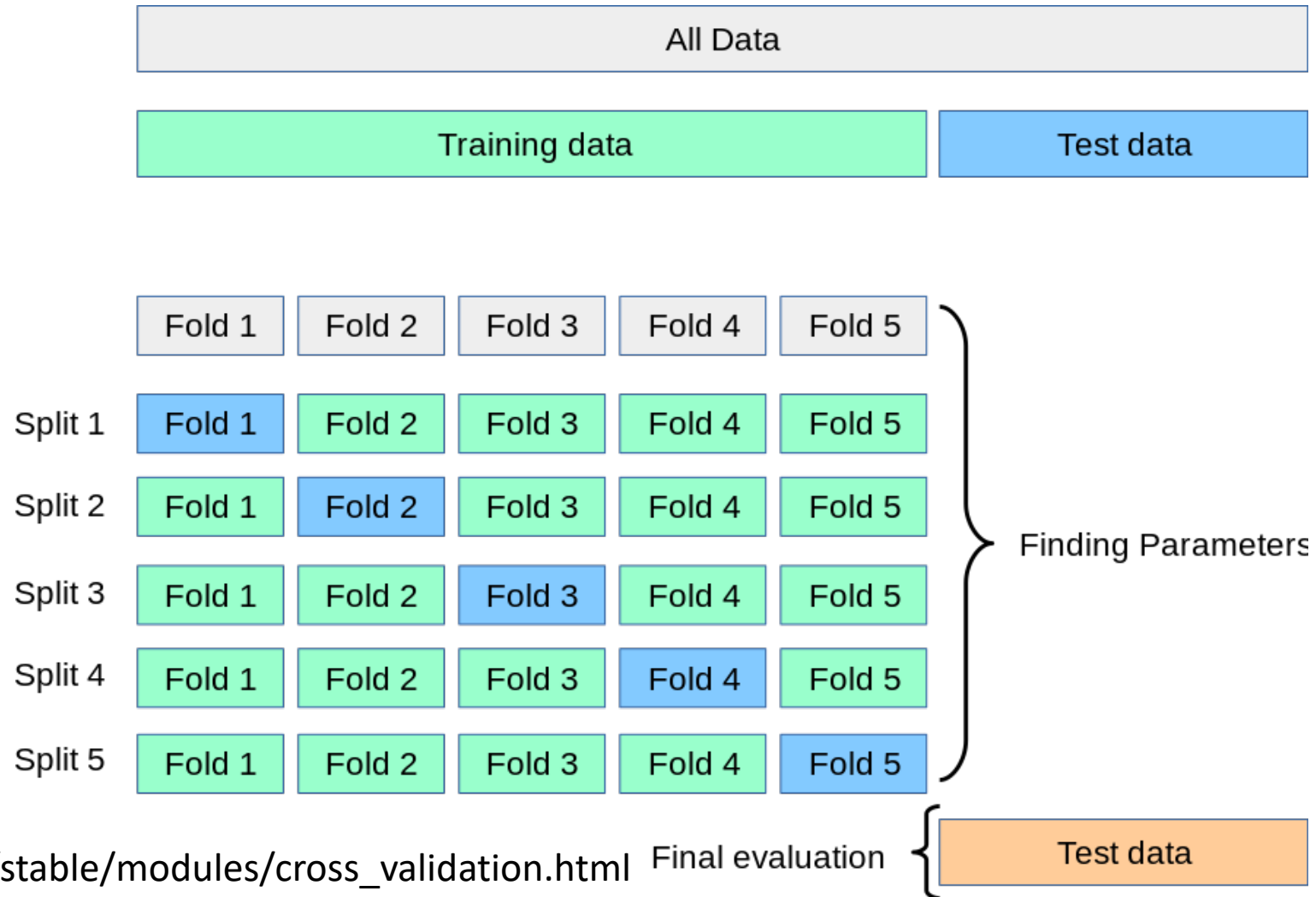
<https://builtin.com/data-science/train-test-split>

Train, Validation and Test Split.



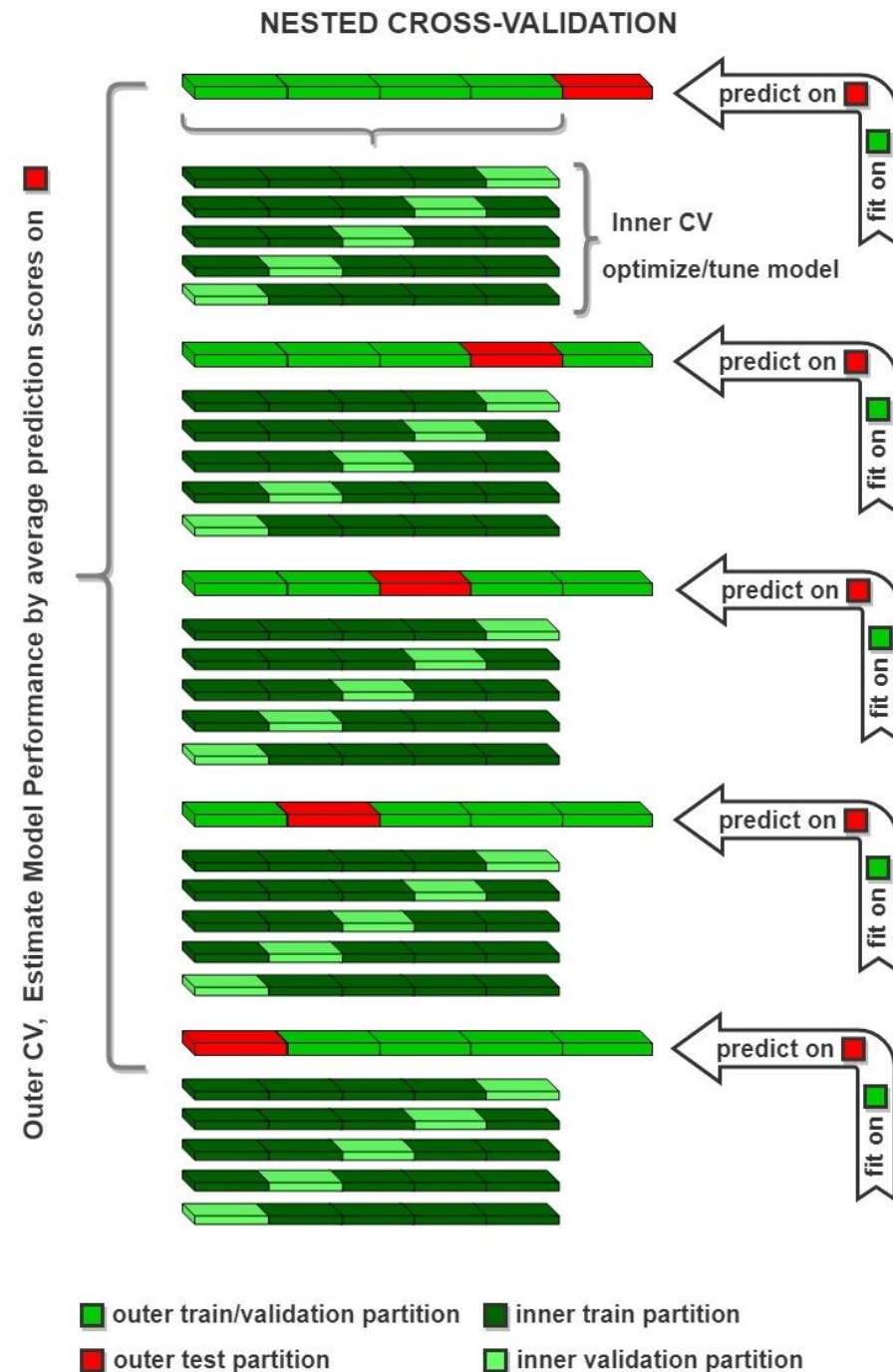
<https://sdsclub.com/how-to-train-and-test-data-like-a-pro/>

K-Fold Cross-Validation



Nested Cross-Validation

https://www.researchgate.net/figure/Overview-of-the-adopted-Nested-cross-validation-scheme-Model-optimization-is-performed_fig1_351514990



Random Subsampling

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Iteration 1															
Iteration 2															
Iteration 3															
Iteration 4															
Iteration 5															

<https://towardsdatascience.com/understanding-8-types-of-cross-validation-80c935a4976d>

Bootstrapping



This work by Sebastian Raschka is licensed under a
Creative Commons Attribution 4.0 International License.

http://rasbt.github.io/mlxtend/user_guide/evaluate/bootstrap_point632_score/