Institución Universitaria
Reacreditada en Alta Calidad
80 Años

Supervised Learning: Regression

Outline

1. Training, Validation and Testing phases
2. Regression Definition
3. Error measurements
4. Overtraining, Undertraining and Regularization

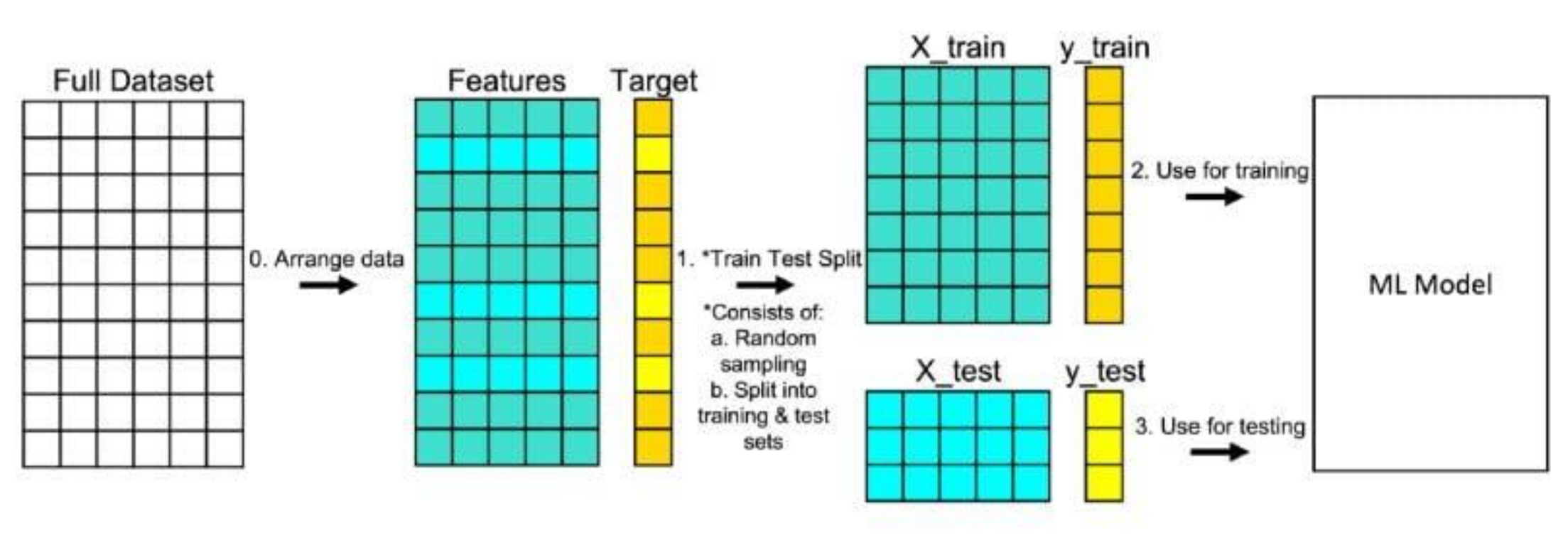
Training, Validation and Testing phases

Train test split is a model validation procedure that allows you to simulate how a model would perform on new/unseen data.

Hyperparameter tuning: model's parameters are tuned according to an optimizacion task during the training phase.

Inference (Testing): After finding a good set of hyperparameters, the trained model is tested on unseen data, in order to check its performance.

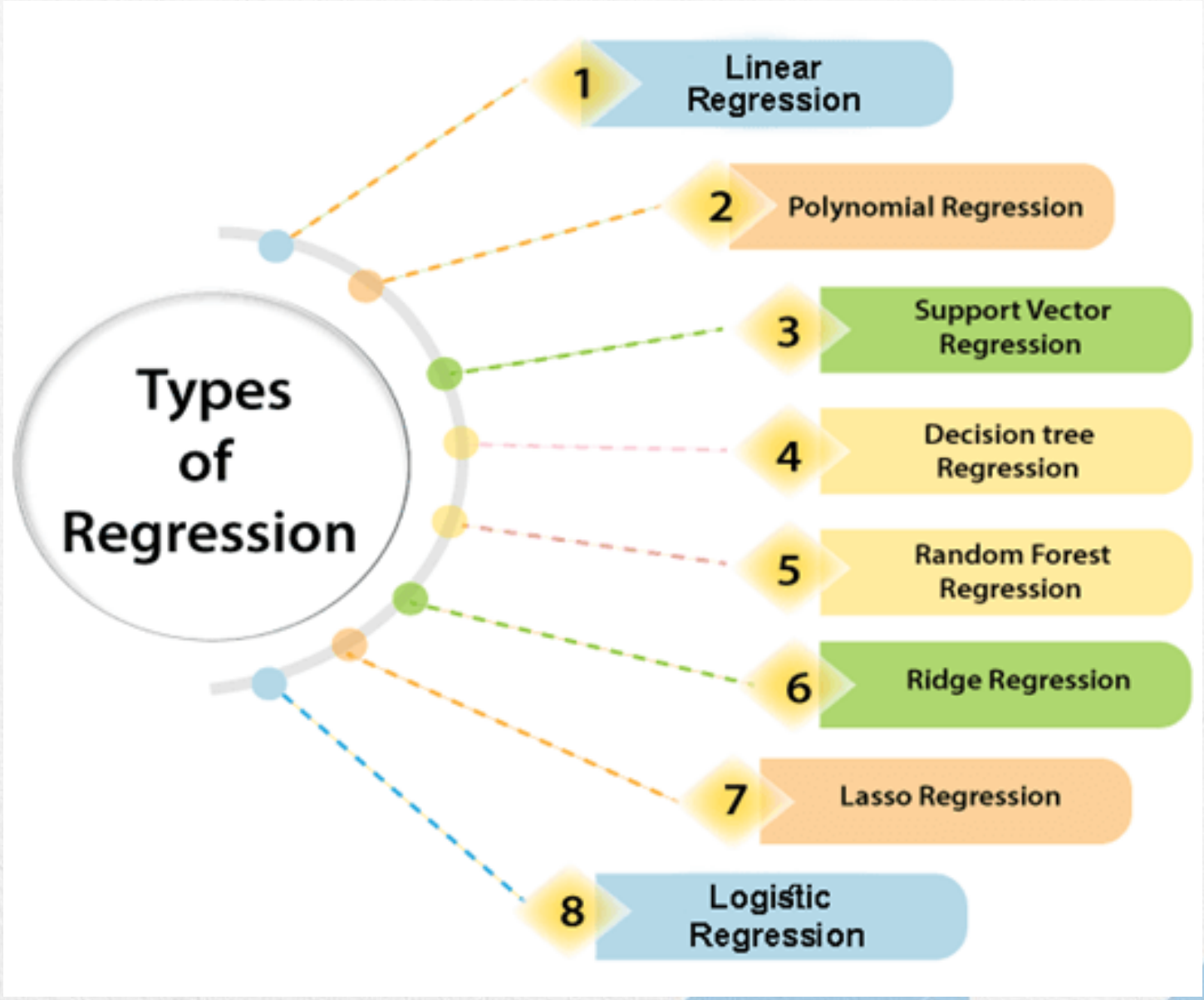
Training, Validation and Testing phases



Example of splitting the data into Training and Test sets.

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

Regression analysis



Outline

1. Training, Validation and Testing phases
2. Regression Definition
3. Error measurements
4. Overtraining, Undertraining and Regularization

Regression analysis

Regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables.

Simple regression

$$f(x) = y$$

Multiple regression

$$f(x_1, x_2, \dots, x_n) = y$$

$$f(\boldsymbol{x}_i) = y_i$$

Outline

1. Training, Validation and Testing phases
2. Regression Definition
3. Error measurements
4. Overtraining, Undertraining and Regularization

Error measurements

Most typical objective functions for regression are, the root mean squared error (RMSE) and the mean absolute error (MAE).

RMSE

$$f(x_i, \theta\ ) = \frac{1}{N}\sum_{i=1}^{N}(\widehat{y}_i(x_i, \theta) - y_i(x_i))^2$$

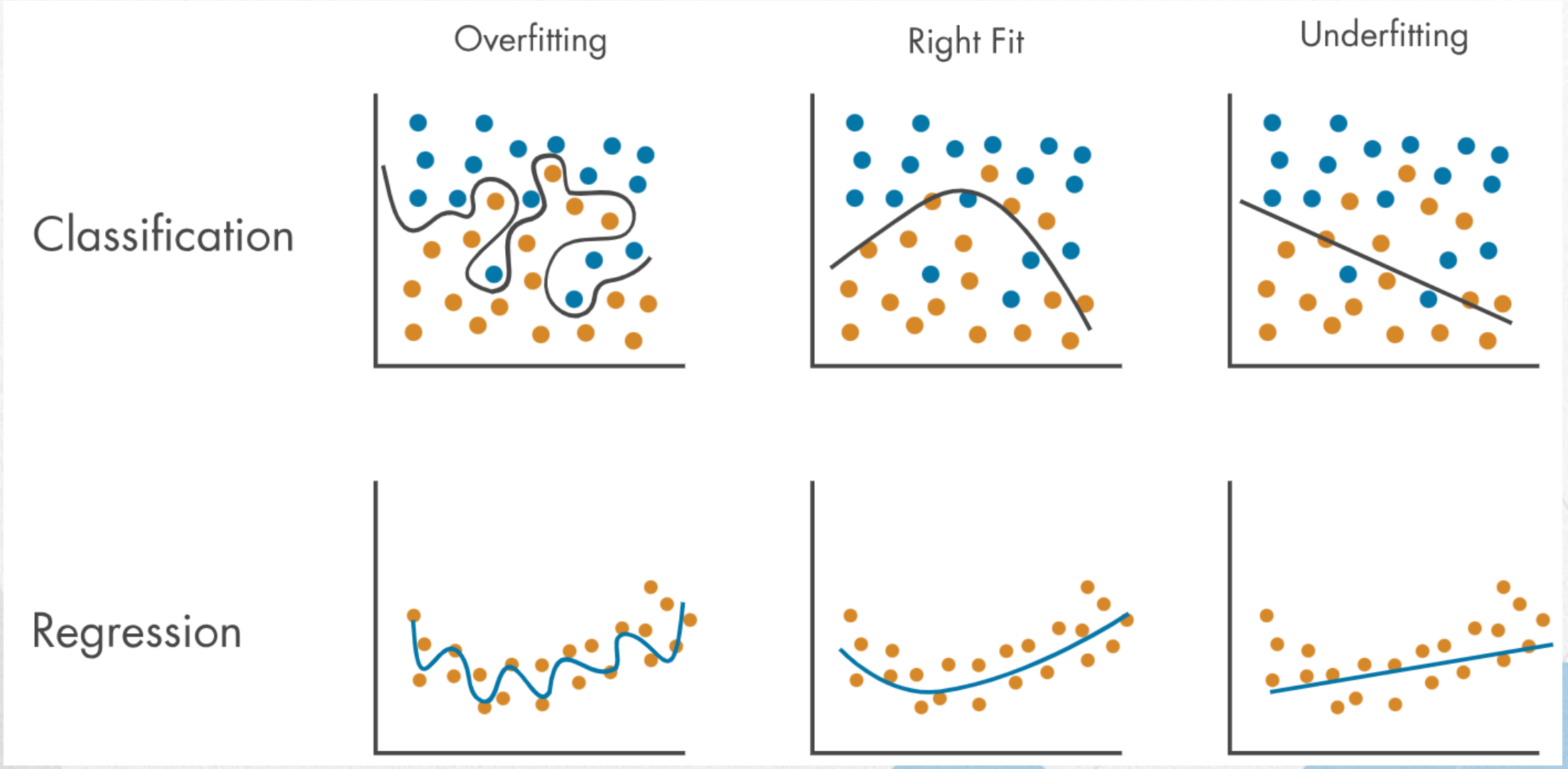
MAE

$$f(\theta) = \frac{1}{N}\sum_{i=1}^{N}|\widehat{y}_i(\theta) - y_i|$$

Outline

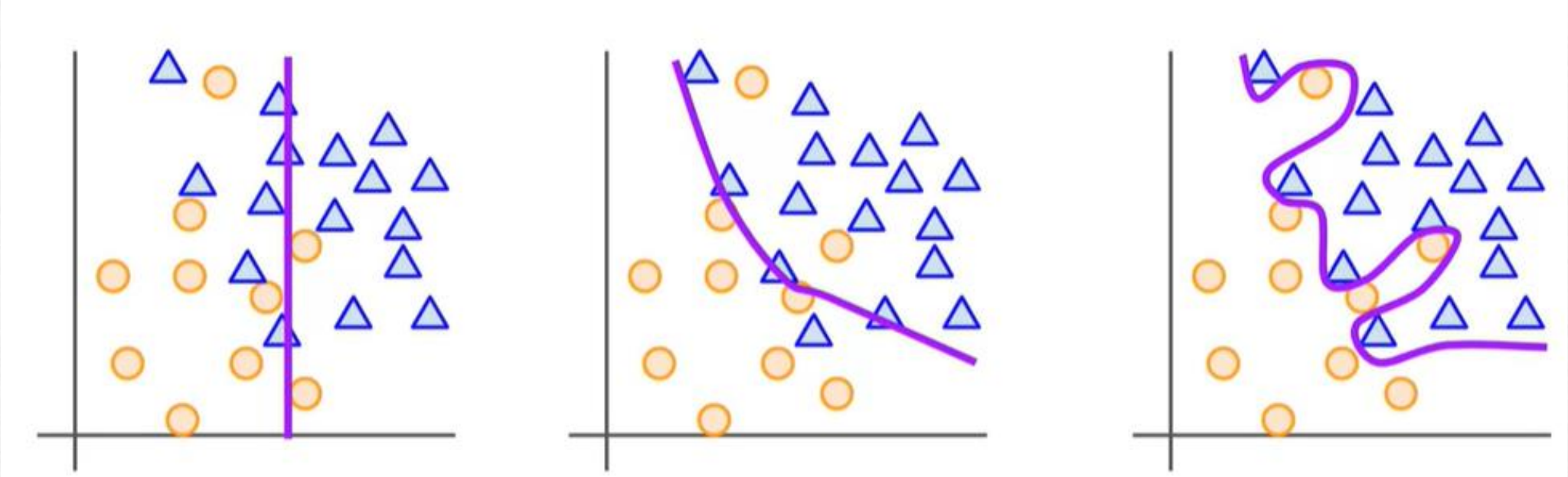
1. Training, Validation and Testing phases
2. Regression Definition
3. Error measurements
4. Overtraining, Undertraining and Regularization

Overfitting



https://www.mathworks.com/discovery/overfitting.html

Overfitting



Underfit: Model fails to capture trends in the data

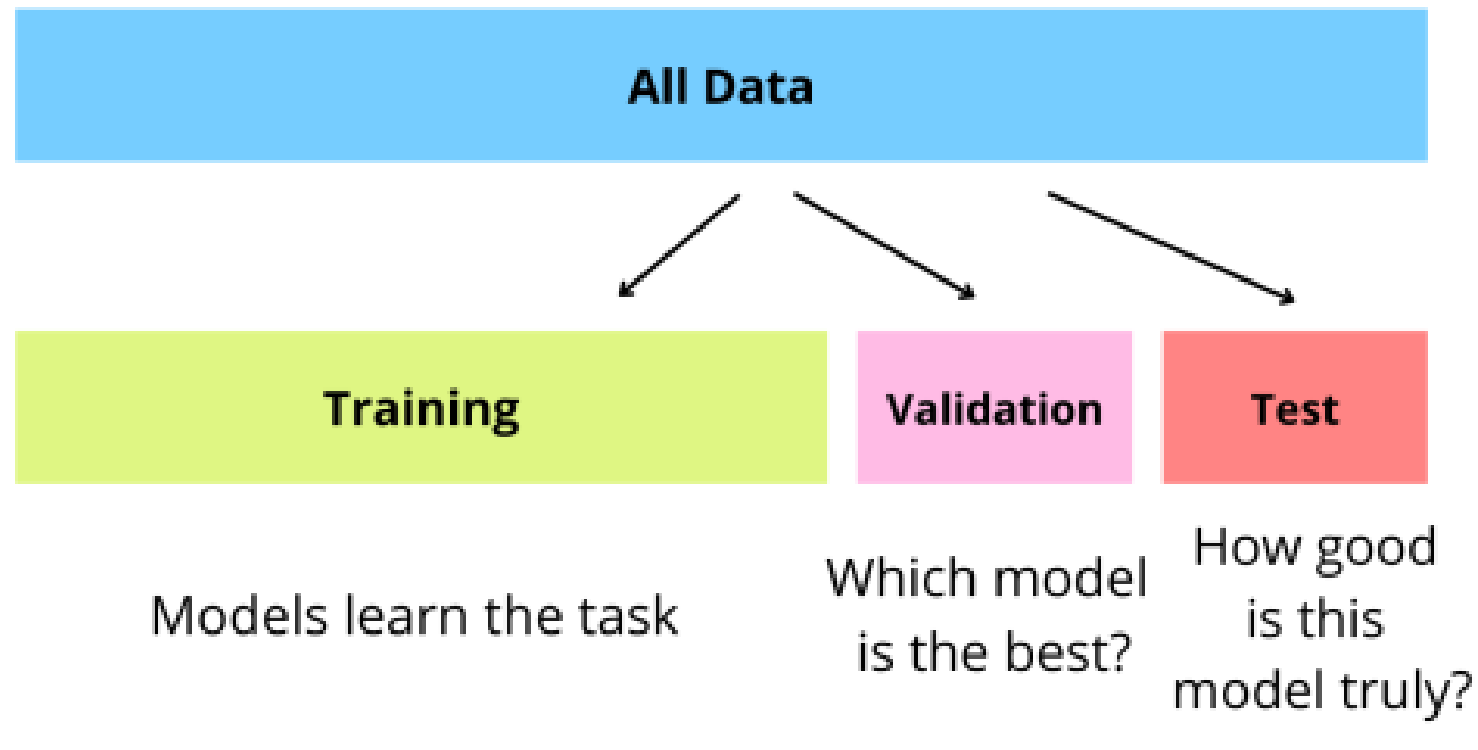
Good fit: Model captures trends and can generalize to unseen data

Overfit: Model captures training data trends but fails on unseen data

https://www.coursera.org/learn/introduction-to-embedded-machine-learning/

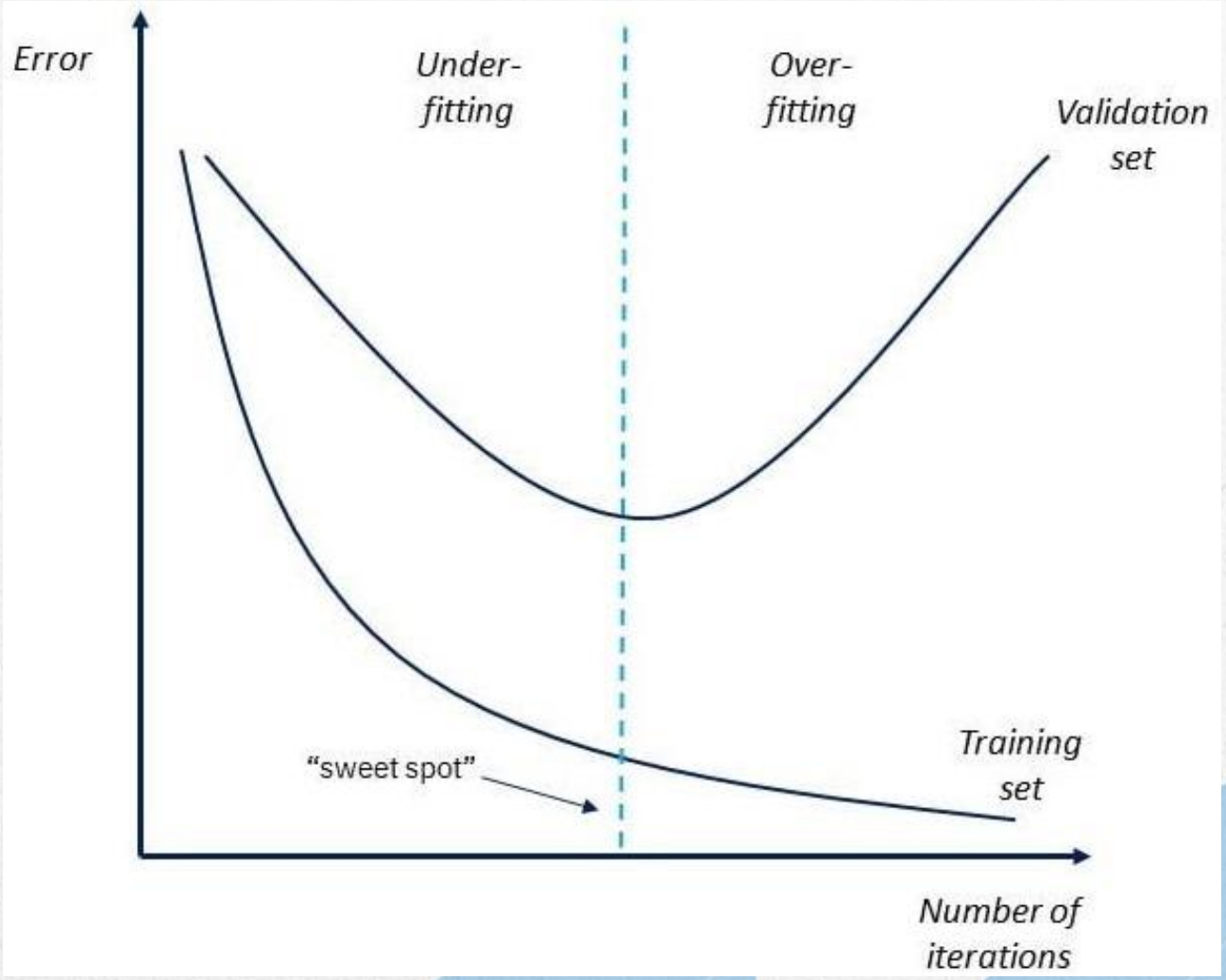
Training in Deep Learning

Validation data is used during training phase to avoid overfitting and perform early stopping.



Training in Deep Learning

Validation data is used during training phase to avoid overfitting and perform early stopping.



Institución Universitaria
Reacreditada en Alta Calidad
80 Años

¡MUCHAS GRACIAS!

VIGILADA MINEDUCACIÓN

www.itm.edu.co

Alcaldía de Medellín
Distrito de
Ciencia, Tecnología e Innovación