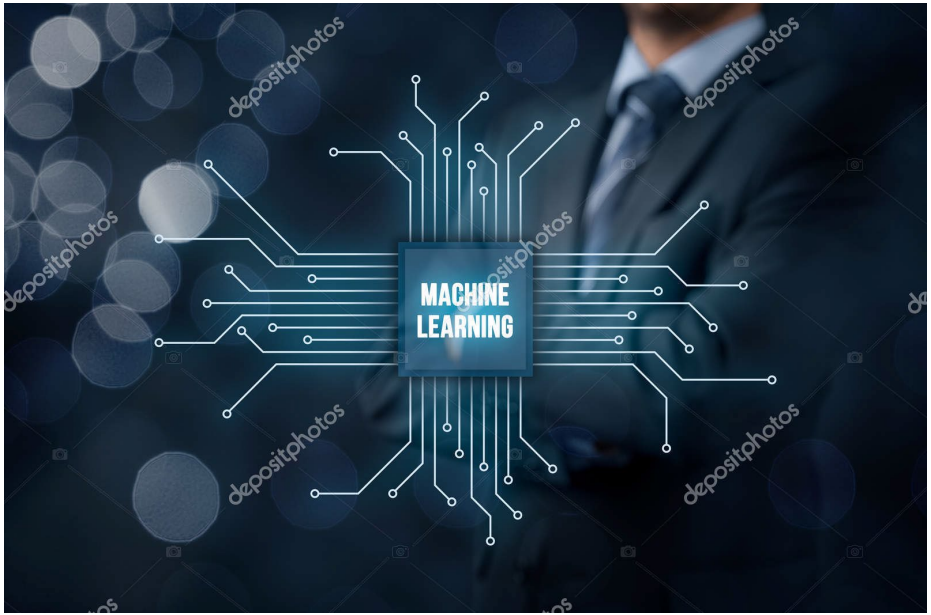
Machine Learning

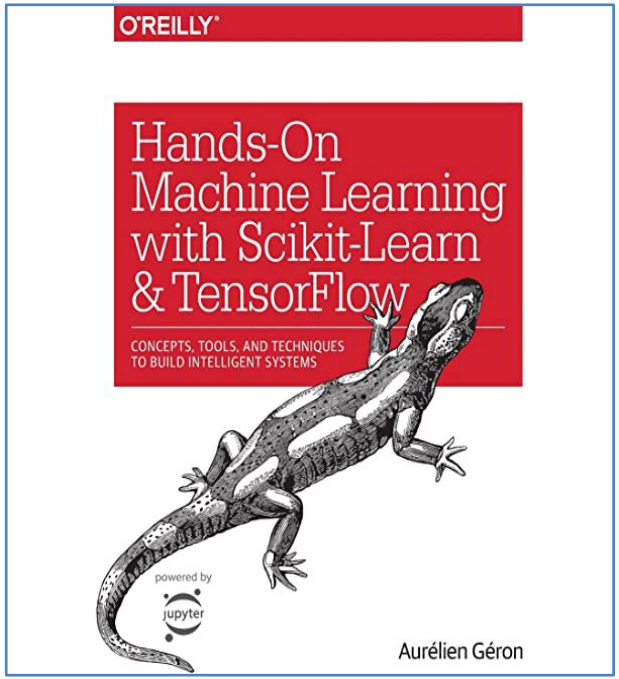
Introduced by

Dr. Ebtsam Adel

Lecturer of Information Systems,
Information Systems department,
Faculty of computers and information,
Damanhour university



Materials



Aurélien Géron

Introduced by

Dr. Ebtsam Adel

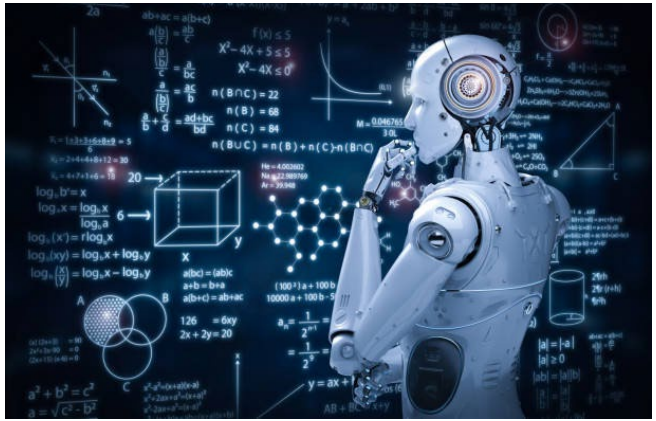

QUIZ!

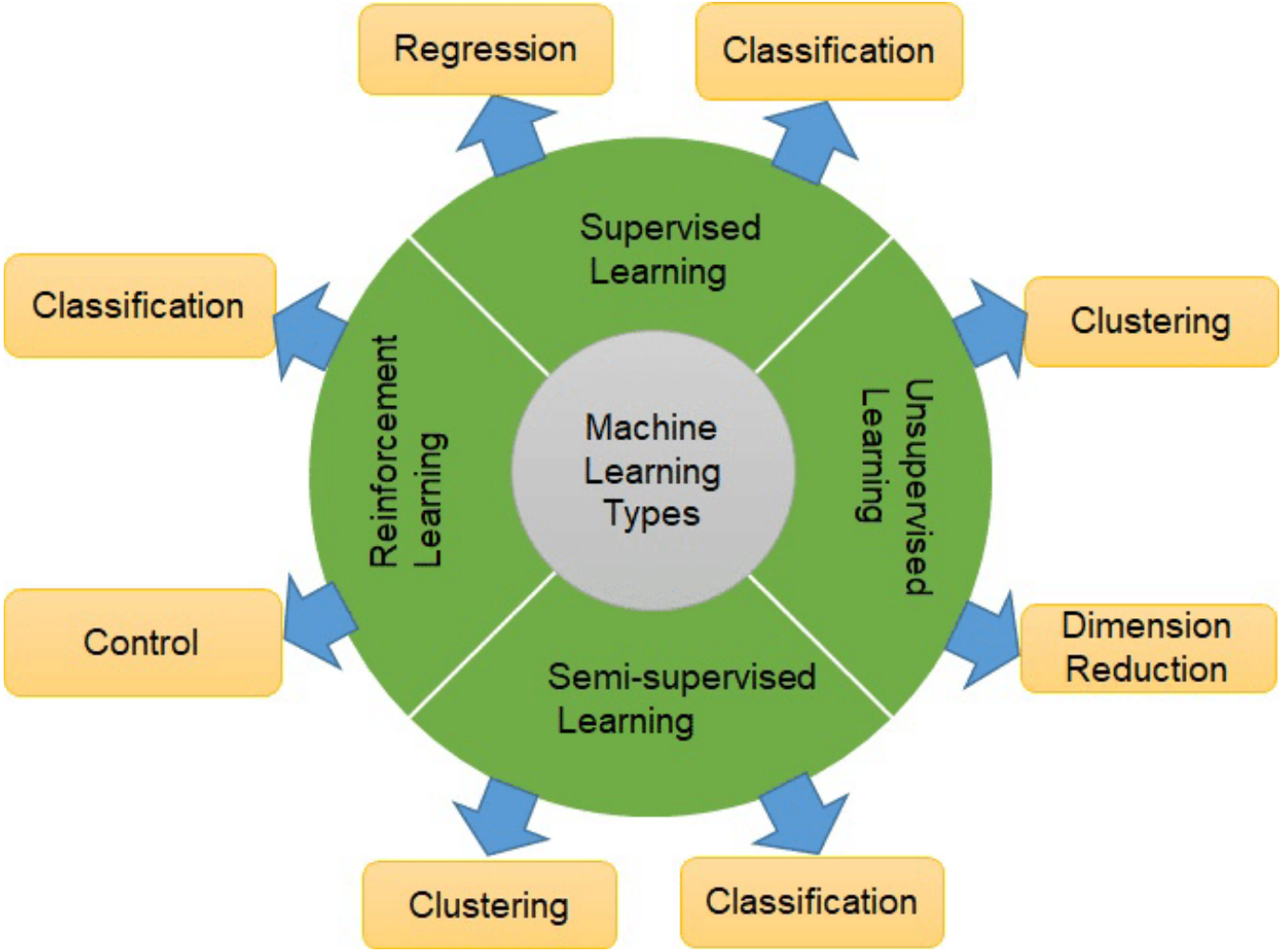
1. Which algorithm is commonly used in machine learning? a. Linear Regression b. Decision Trees c. Neural Networks d. All of the above	2. **What is the primary goal of machine learning?** a. To make predictions b. To build algorithms c. To analyze data d. To create software
3. What does supervised learning involve? a. Learning from labeled data b. Learning without any labels c. Learning through trial and error d. Learning from unstructured data	4. **What is 'training' in the context of machine learning?** a. Testing the model b. Improving the algorithm c. Feeding data to the model d. Analyzing results

End-to-End Machine Learning Project

How does Machine Learning Work?

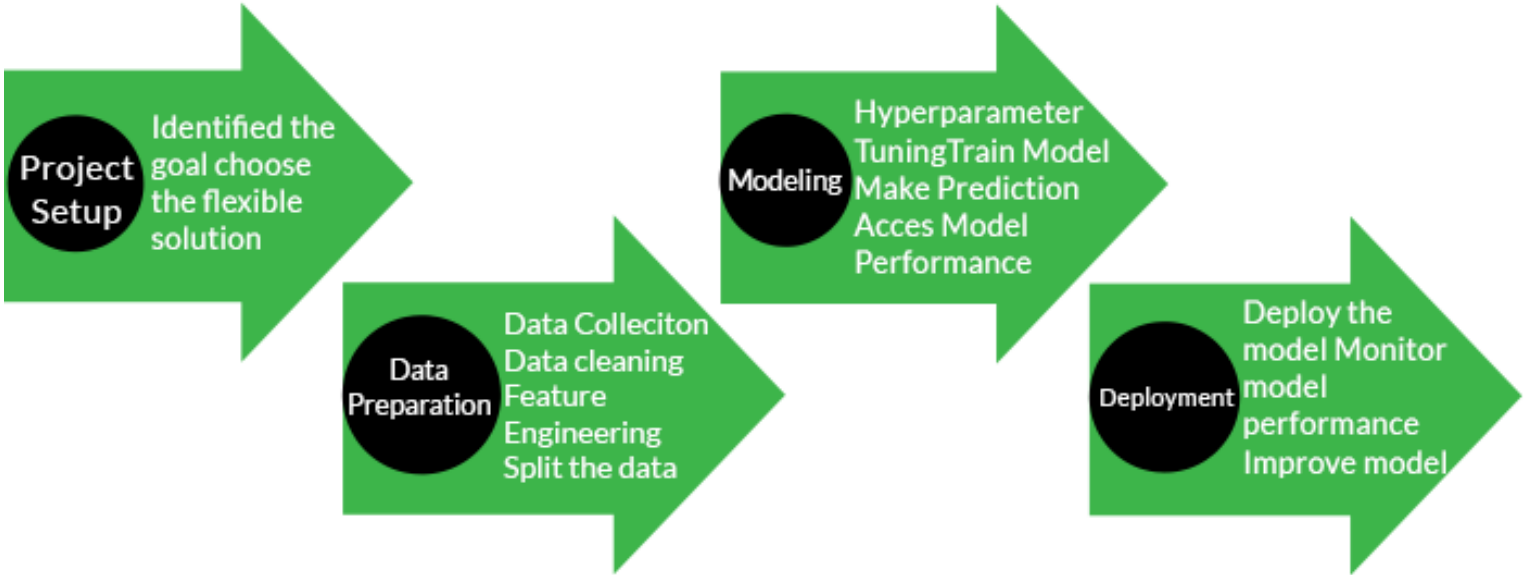
- **Machine Learning** is a branch of artificial intelligence that works on algorithm developments and statistical models that allow computers to learn from data and make predictions or decisions without being explicitly programmed.




Regression
Classification
Supervised Learning
Classification
Clustering
Reinforcement Learning
Machine Learning Types
Unsupervised Learning
Control
Dimension Reduction
Semi-supervised Learning
Clustering
Classification

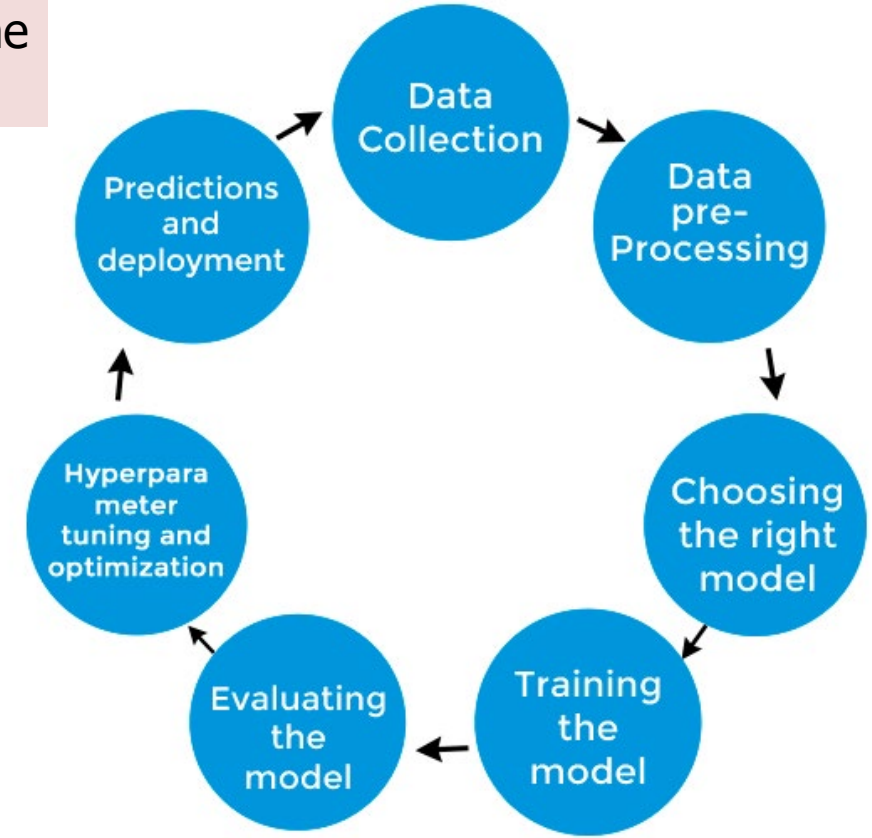
How does Machine Learning Work?

- Machine Learning process **includes** Project Setup, Data Preparation, Modeling and Deployment.



How does Machine Learning Work?

A detailed sequential process of Machine Learning.



How does Machine Learning Work?

1.Data collection: Data collection is an initial step in the process of machine learning.

- Data is a fundamental part of machine learning, the quality and quantity of your data can have direct consequences for model performance.
- Different sources such as databases, text files, pictures, sound files, or web scraping may be used for data collection.
- Data needs to be prepared for machine learning once it has been collected. This process is to organize the data in an appropriate format, such as a CSV file or database, and make sure that they are useful for solving your problem.

How does Machine Learning Work?

2. Data pre-processing: Pre-processing of data is a key step in the process of machine learning.

- It involves deleting duplicate data, fixing errors, managing missing data either by eliminating or filling it in, and adjusting and formatting the data.
- Pre-processing improves the quality of your data and ensures that your machine-learning model can read it right.

- The accuracy of your model may be significantly improved by this step.

How does Machine Learning Work?

3. Choosing the right model: The next step is to select a machine learning model; once data is prepared then we apply it to ML Models like Linear regression, decision trees, and Neural Networks that may be selected to implement.

- The selection of the model generally depends on what kind of data you're dealing with and your problem. The size and type of data, complexity.

How does Machine Learning Work?

4. Training the model: The next step is to train it with the data that has been prepared after you have chosen a model.

- Training is about connecting the data to the model and enabling it to adjust its parameters to predict output more accurately.

- Overfitting and Underfitting must be avoided during the training.

How does Machine Learning Work?

5. Evaluating the model: It is important to assess the model's performance before deployment as soon as a model has been trained.

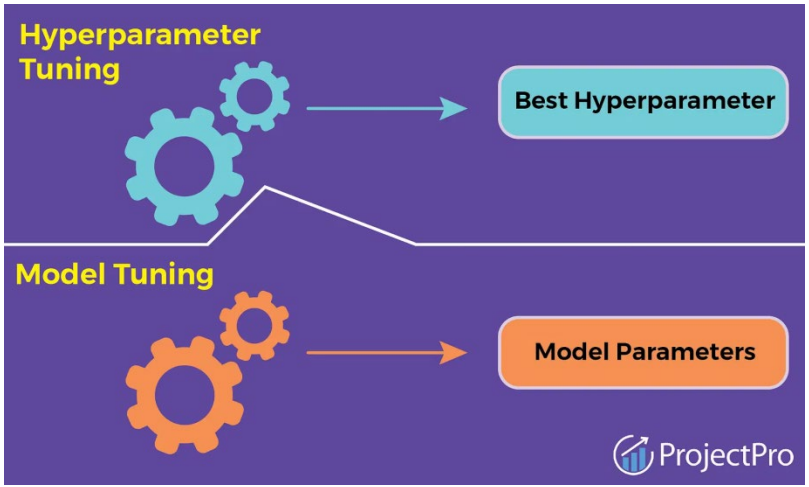
- This means that the model has to be tested on new data that they haven't been able to see during training.

- Accuracy in classifying problems, precision and recall for binary classification problems, as well as mean error squared with regression problems, are **common metrics** to evaluate the performance of a model.

How does Machine Learning Work?

6. Hyperparameter tuning and optimization: You may need to adjust its hyperparameter to make it more efficient after you've evaluated the model.

- where you try different combinations of parameters, and cross-validation, where you divide your data into subsets and train your model on each subset, to ensure that it performs well on different data sets, are techniques for hyperparameter tuning.



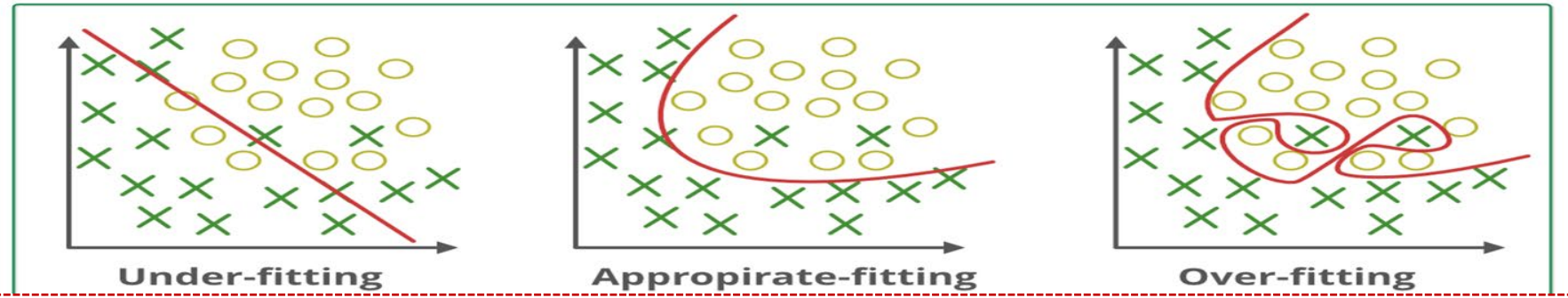
How does Machine Learning Work?

7. Deployment: As soon as the model has been programmed and optimized, it will be ready to estimate new data.

- This is done by adding new data to the model and using its output for decision-making or other analysis.

- The deployment of this model involves its integration into a production environment where it is capable of processing real-world data and providing timely information.

Overfitting the Training Data



$\theta_0 + \theta_1 x$

High Bias
(Underfitting)

$\theta_0 + \theta_1 x + \theta_2 x^2$

Low Bias, Low Variance
(Goodfitting)

$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$

High Variance
(Overfitting)

Underfitting

- Poor performance training data
- Poor performance testing data
- High Bias "توقع"
- Low variance

Overfitting

- Good performance training data
- Poor performance testing data
- Low Bias, High variance "تفاوت"
- Memorizing, not generalizing.

❖ **Variance:** the difference in fits between datasets .

Overfitting

Overfitting the Training Data

In Machine Learning this is called **overfitting**: it means that the model performs well on the training data, but it does not **generalize** well.

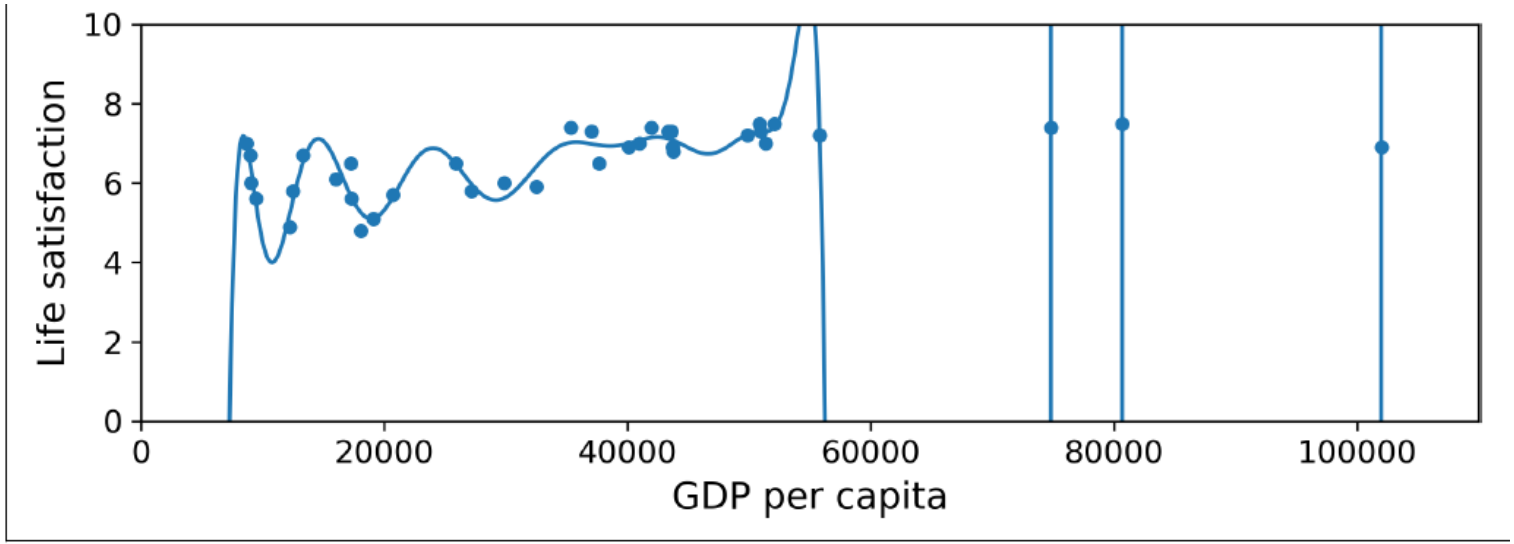


Figure 1-22. Overfitting the training data

More spread

Overfitting

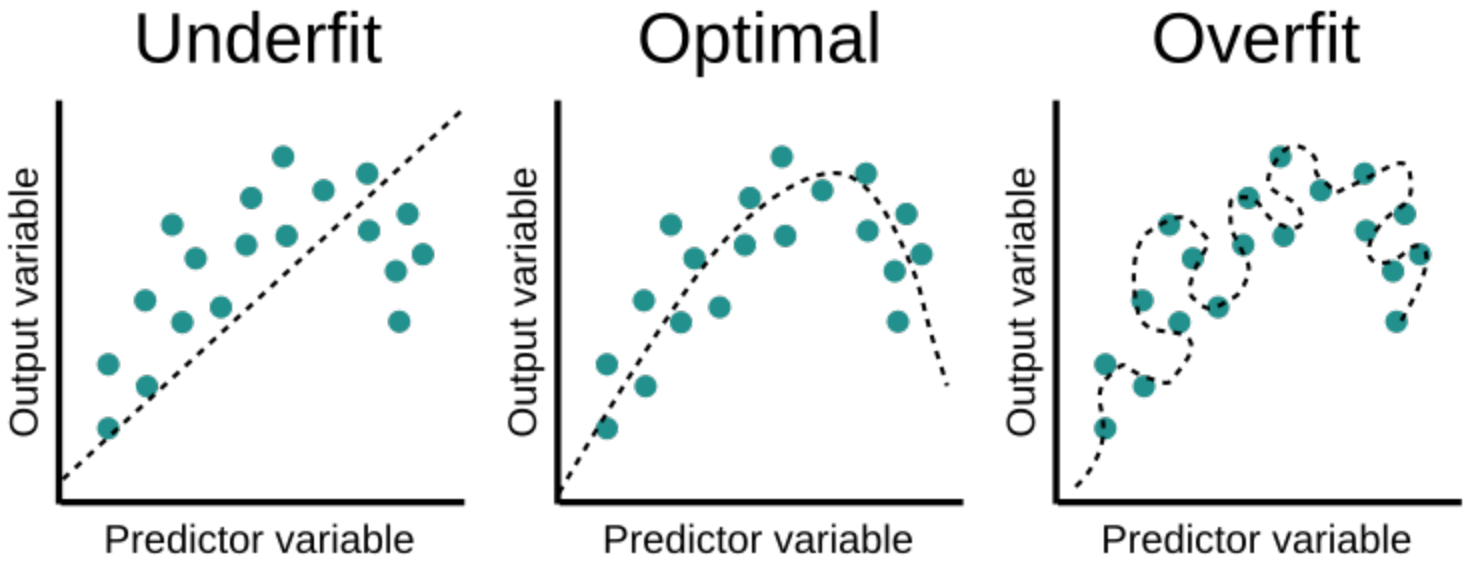
- **Overfitting** occurs when a model learns the noise in the training data, rather than the underlying patterns. This causes the model to perform well on the training data, but poorly on new data.

- Essentially, the model becomes too specialized to the training data, and is unable to generalize to new data.

Underfitting

Underfitting the Training Data

- **Underfitting** is the opposite of overfitting: it occurs when your model is **too** simple to learn the underlying structure of the data.



Underfitting & Overfitting

What is Bias?

While making predictions, a difference occurs between prediction values made by the model and actual values/expected values, and this difference is known as bias errors or Errors due to bias.

- **High bias** means our model is not complex enough to capture the underlying patterns in the data, resulting in poor generalization (low accuracy).
- **Low bias** means our model has overfit the training data, learning the noise rather than the true patterns.

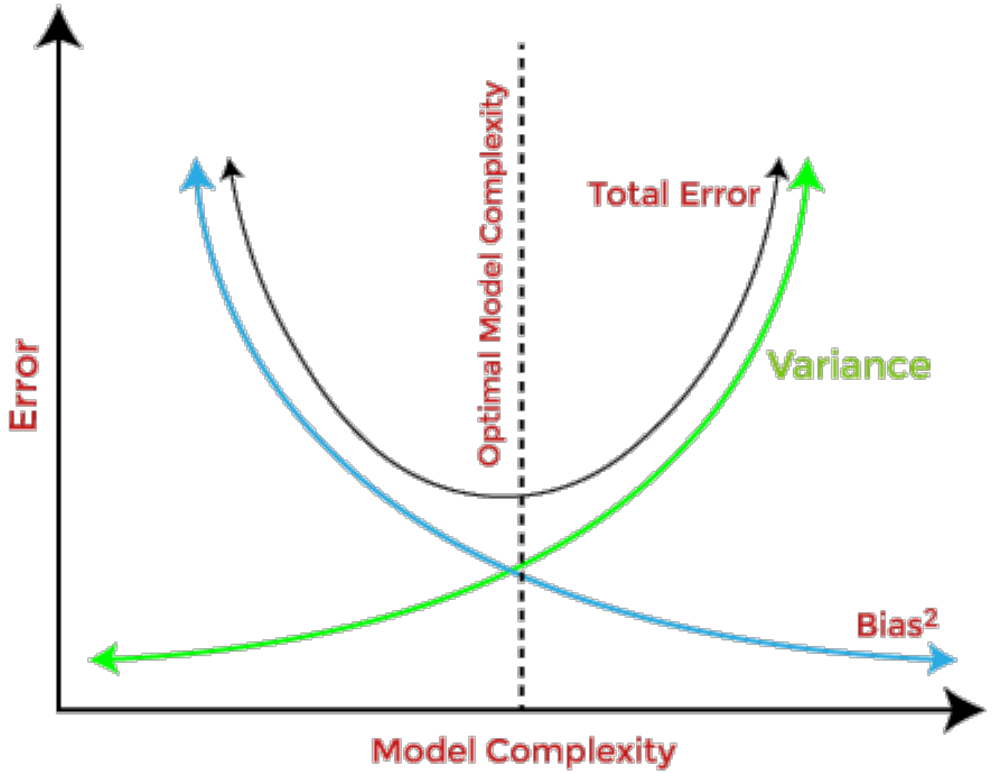
Under-fitting & over-fitting

What is Variance?

- In machine learning variance is the amount by which the performance of a predictive model changes when it is trained on different subsets of the training data.

- More specifically, variance is the variability of the model that how much it is **sensitive to another subset** of the training dataset. i.e. how much it can adjust on the new subset of the training dataset.

Underfitting & overfitting



Overfitting & Underfitting **solutions**

Overfitting happens when the model is too complex relative to the amount and noisiness of the training data.
The possible **solutions** are:

- To simplify the model by selecting **fewer parameters**
- To gather more training data
- To reduce the noise in the training data.

The main options to fix **Underfitting** problem are:

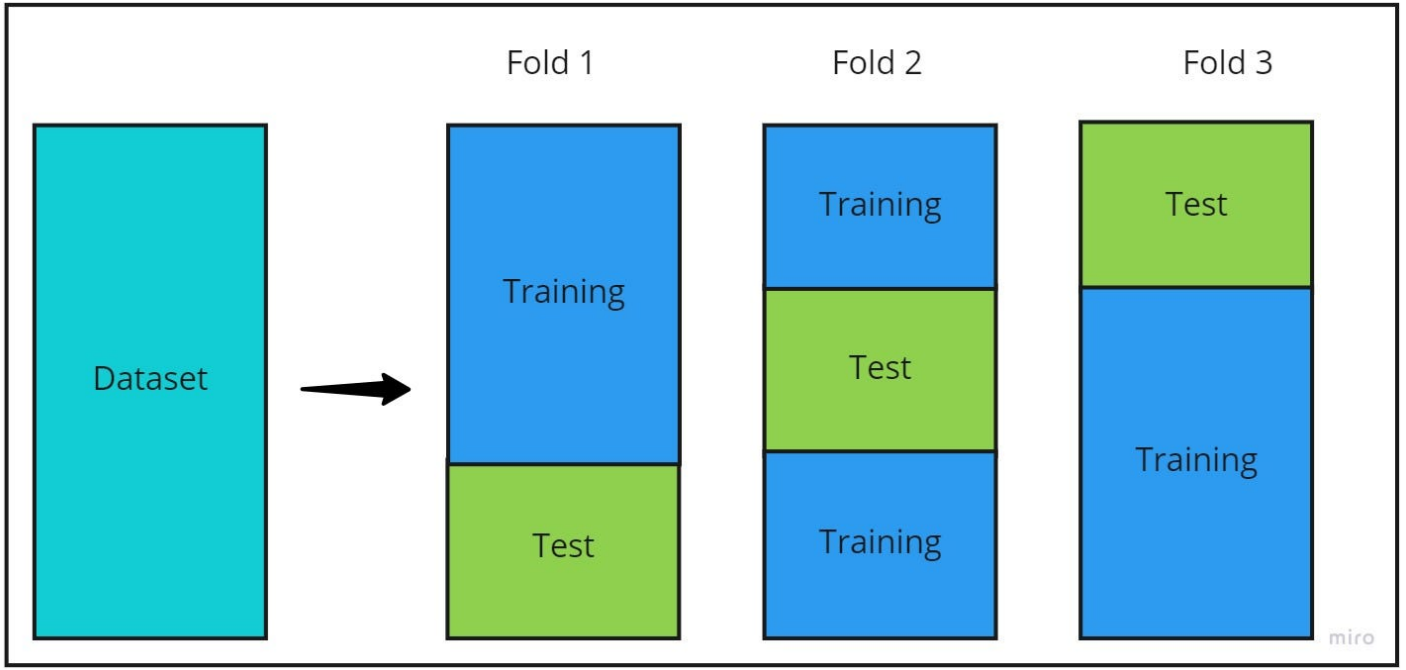
- Selecting a more powerful model, **with more parameters**
- Feeding better features to the learning algorithm (feature engineering)
- Reducing the constraints on the model.

Techniques to Prevent Overfitting

There are several techniques that can be used to prevent overfitting in machine learning:

- ✓ **Cross-validation**: Cross-validation is a technique used to evaluate a model's performance on new, unseen data. It involves dividing the data into several subsets, and using each subset in turn as a validation set, while training on the remaining data.

Techniques to Prevent Overfitting



Generalization

Generalization

- **Generalization** in machine learning refers to the ability of a trained model to accurately make predictions on new, unseen data.

- The purpose of generalization is to prepare the model to understand the patterns and relationships within its training data and apply them to previously unseen examples from within the same distribution as the training set.

Techniques for Generalization

Finding the **optimal balance** between Underfitting and overfitting is crucial for achieving the best performance of a machine learning model.

- **Regularization**: Techniques like L1 and L2 regularization (also known as ridge and lasso regression) help to control model complexity and prevent overfitting.

- **Cross-Validation:** It involves splitting the available data into multiple subsets, training the model on a portion of the data, and evaluating its performance on the remaining test set.

Techniques for Generalization

- **Data Augmentation:**This technique involves artificially increasing the size of the training dataset by introducing variations or modifications to the existing data.

- **Feature Engineering:** By selecting and engineering relevant features, data scientists can provide the model with more informative representations.

Testing and Validating

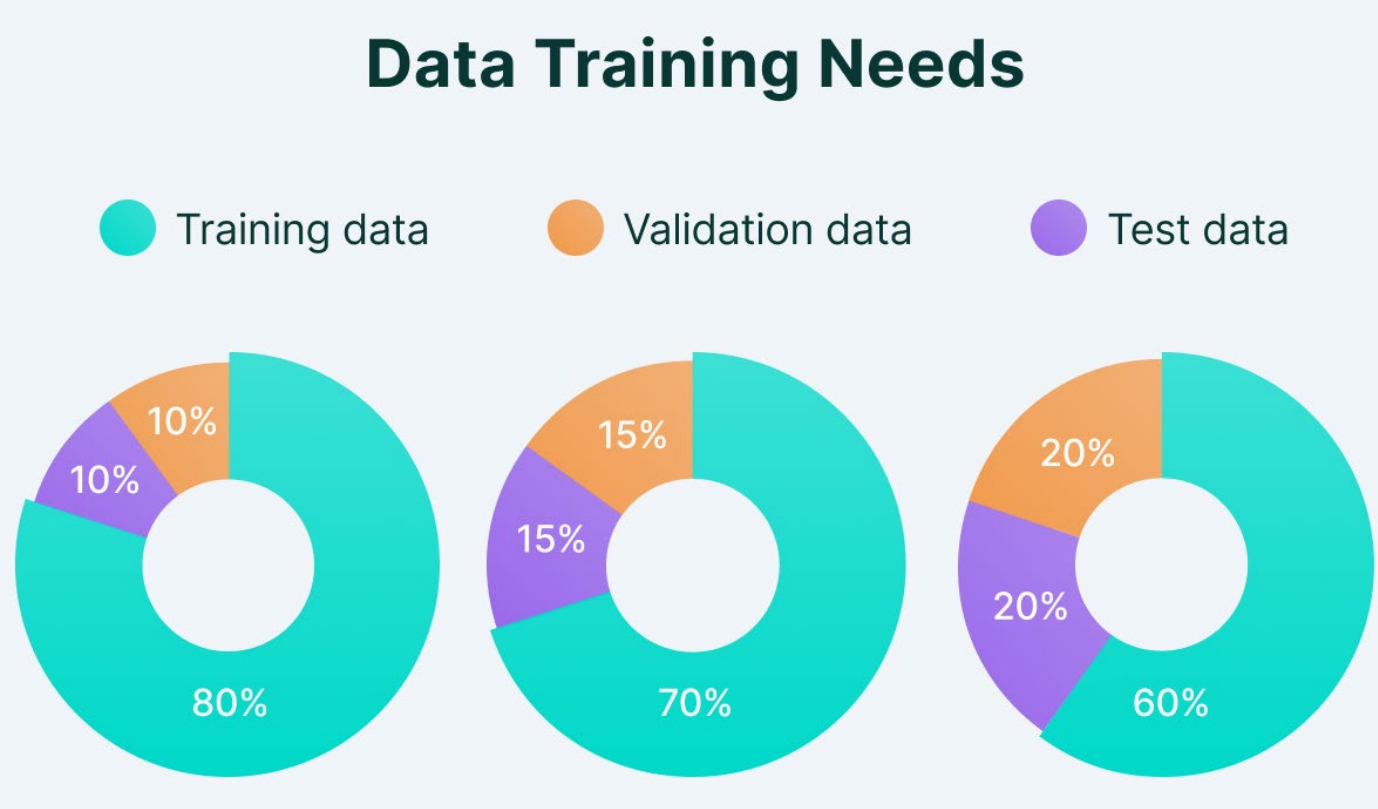
Testing and Validating

Testing and Validating

- A better option is to split your data into two sets: the **training set** and the **test set**. As these names imply, you train your model using the training set, and you test it using the test set.

- The error rate on new cases is called the generalization error (or out-of sample error), and by evaluating your model on the test set, you get an estimate of this error. This value tells you how well your model will perform on instances it has never seen before.

Testing and Validating



Practical Part

Practical part

- Use the Sequential model in Keras to define the model architecture, and add L2 regularization to the first two layers using the kernel_regularizer argument.

- from keras.models import Sequential
- from keras.layers import Dense
- from keras.callbacks import EarlyStopping
- from keras import regularizers


Thank You!
Any questions?