

Fundamentals of Machine Learning

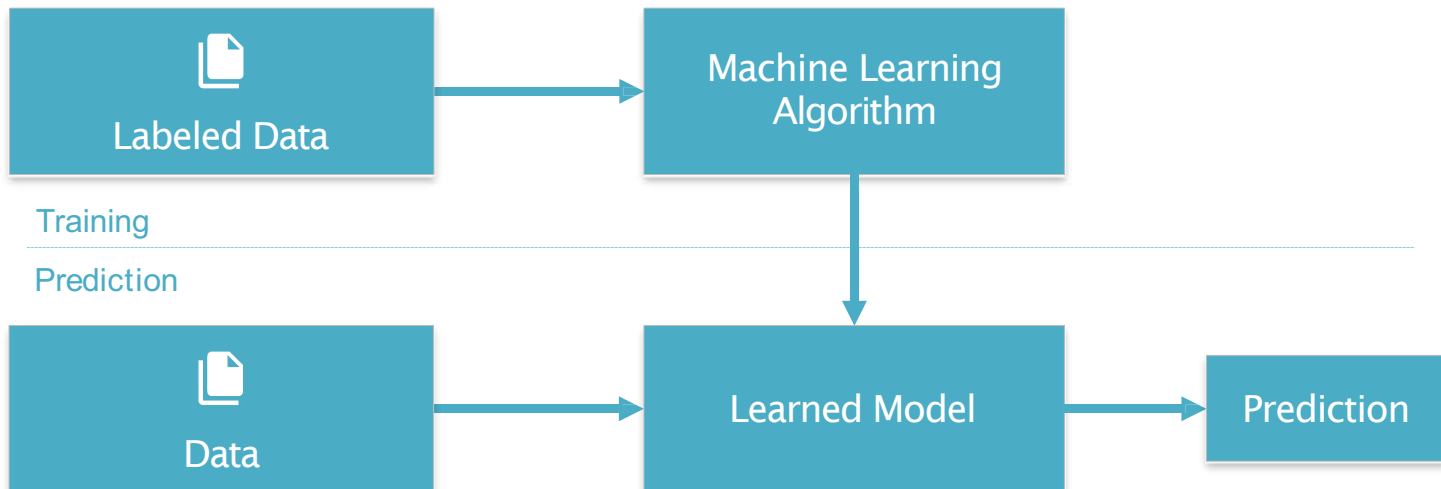
Labeeb Khan

Machine Learning - Basics

Introduction



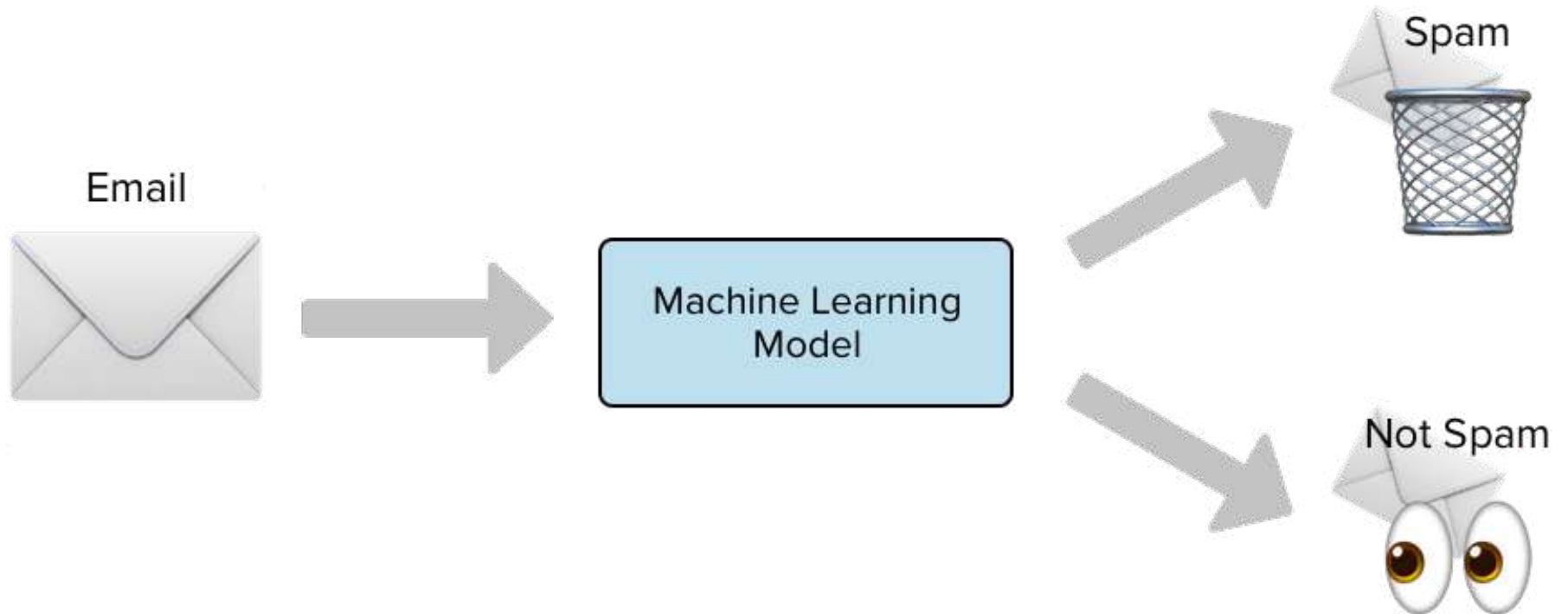
Machine Learning is a type of Artificial Intelligence that provides computers with the ability to **learn without being explicitly programmed**.



Provides **various techniques** that can learn from and make predictions on data

Example – Spam Detection

Determining if an email is spam vs not spam

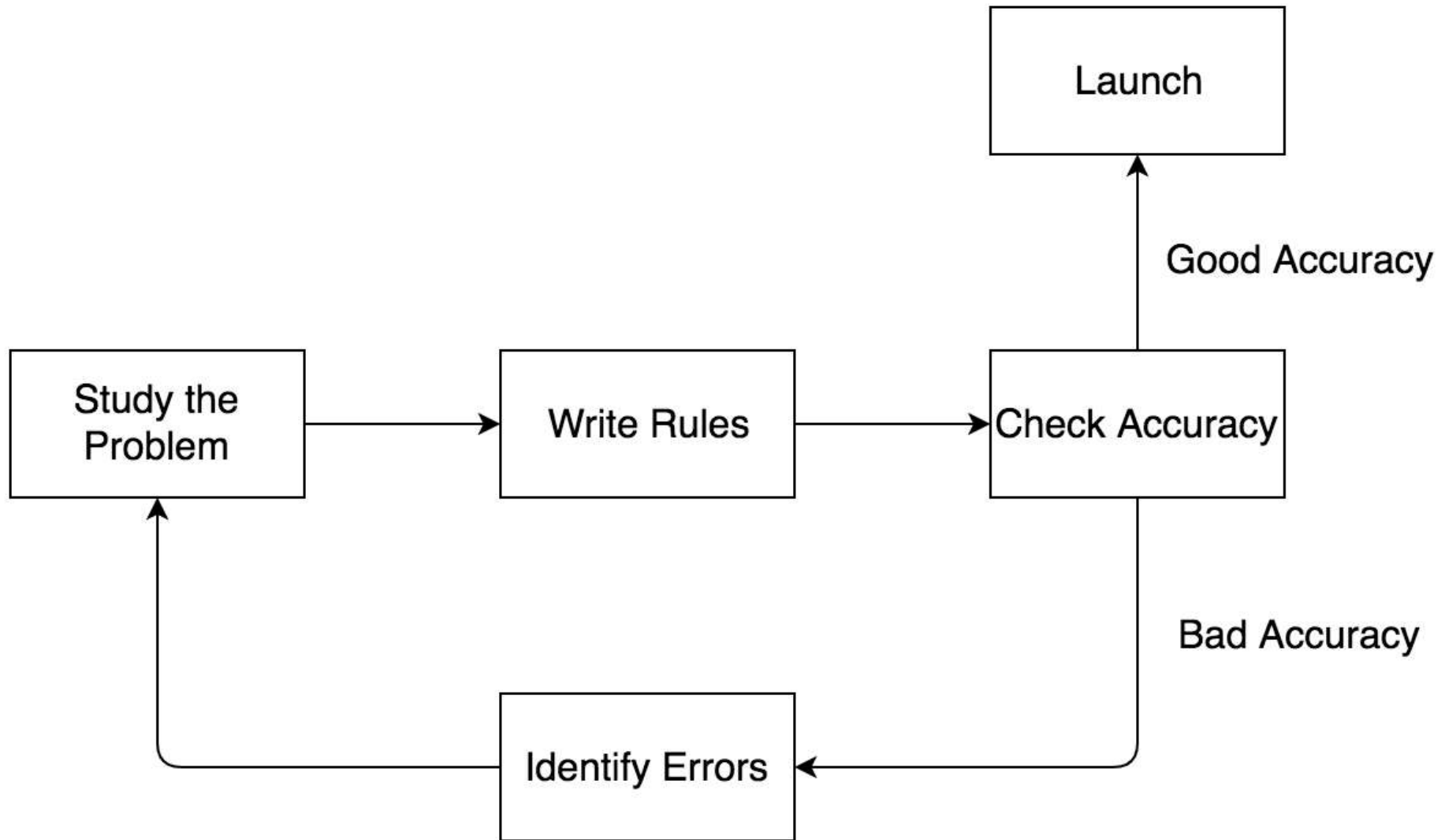


Example – Spam Detection

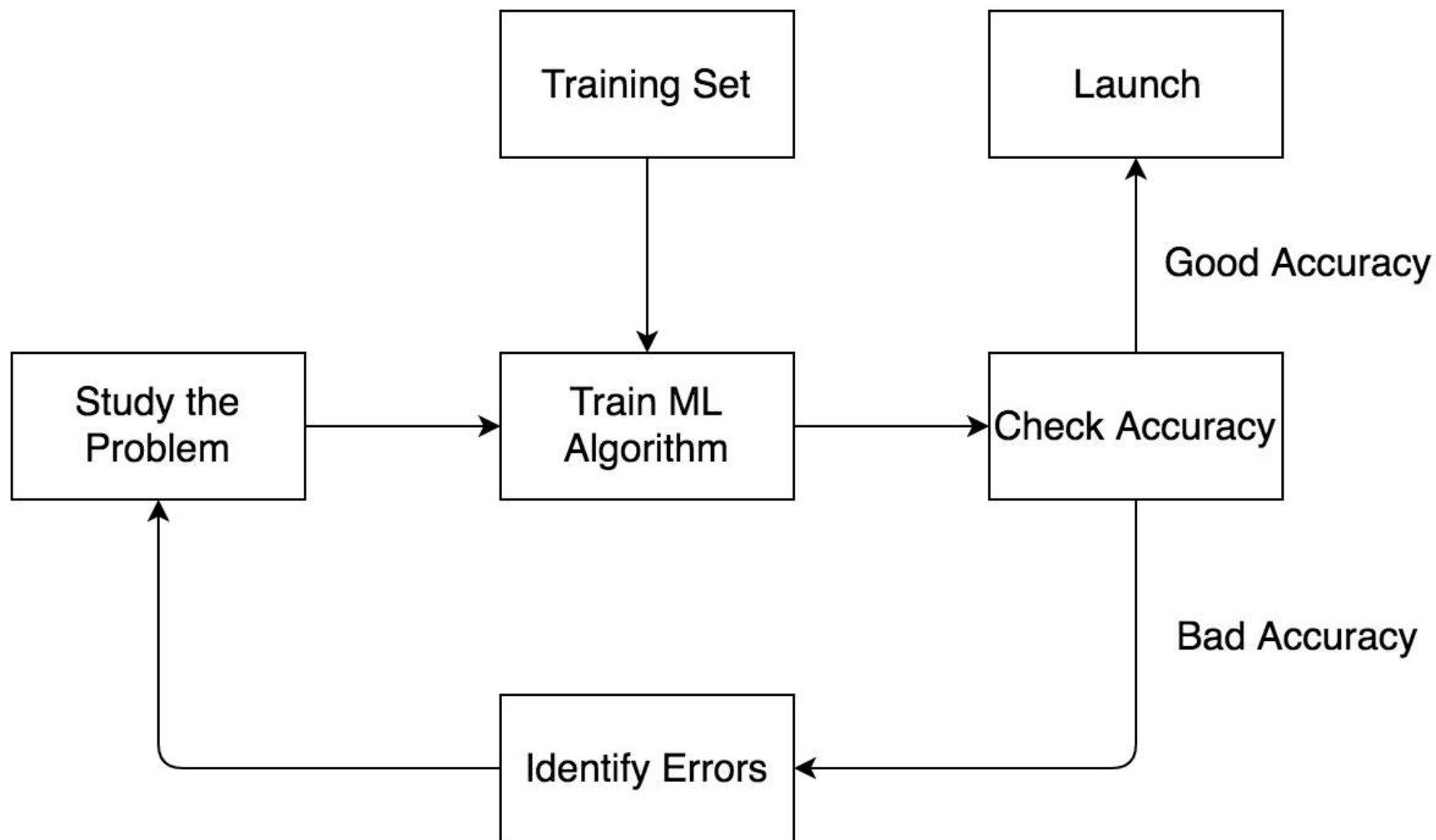
Determining if an email is spam vs not spam

| Email message | Class |
|--|----------|
| Buy these pills | Spam |
| Win cash prizes | Spam |
| Dear Mr. Atreides, please find attached... | Not Spam |

Traditional Approach



Machine Learning Approach



Machine Learning Approach – Benefits

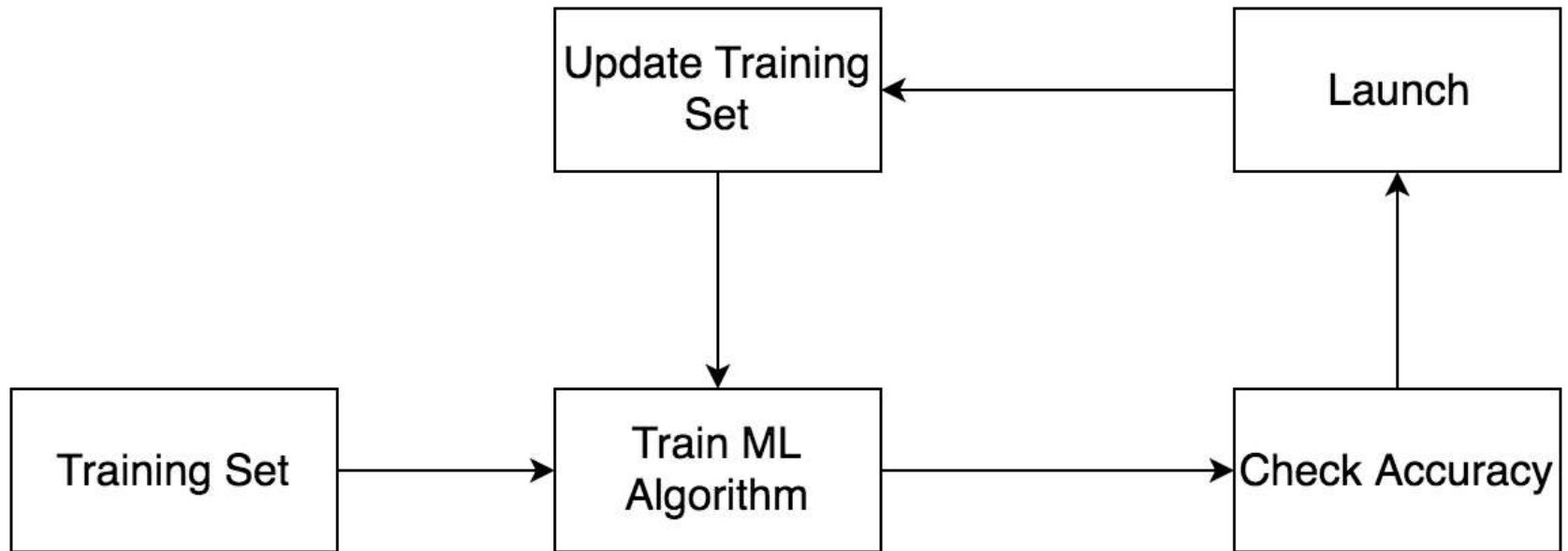
Benefits

- Program is shorter to write
- Easier to maintain
- More accurate
- Adaptable to new instances



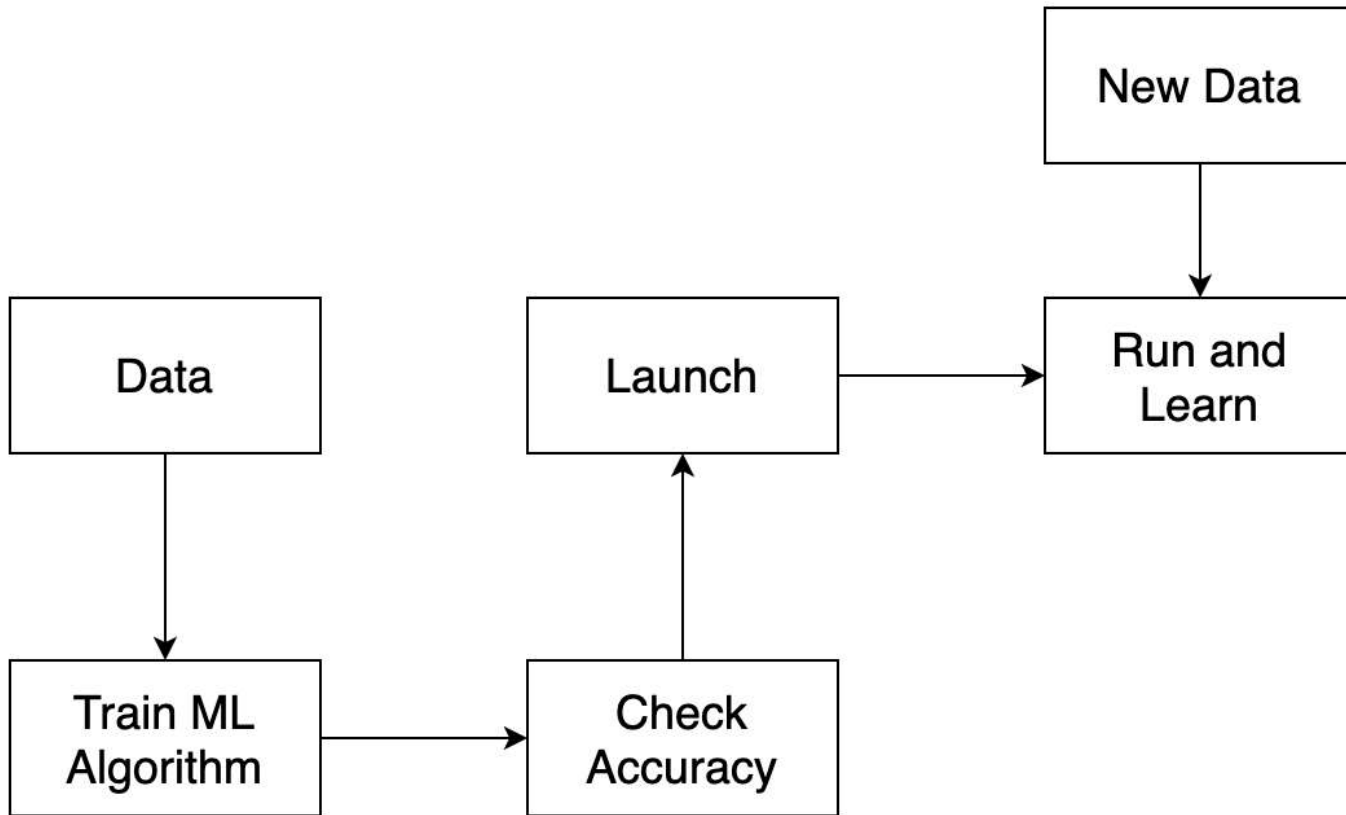
Machine Learning Approach (Offline / Batch Learning)

Adaptive to changes in the target or features (*concept drift & data drift*)

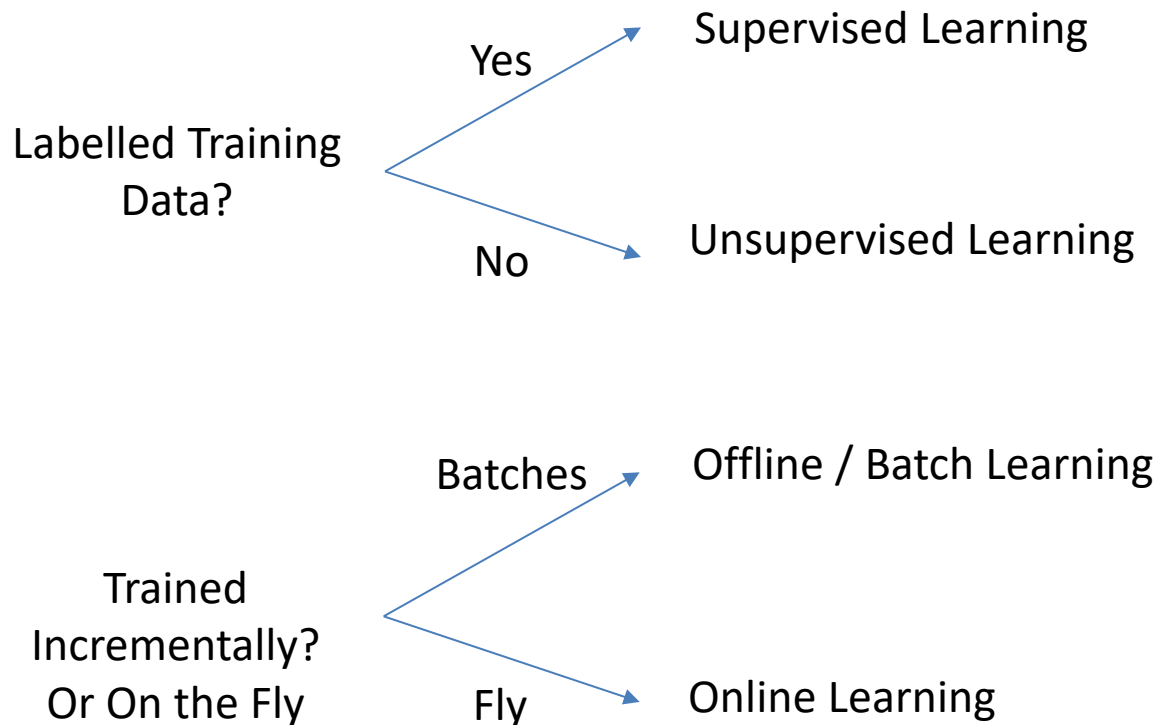


Machine Learning Approach (Online Learning)

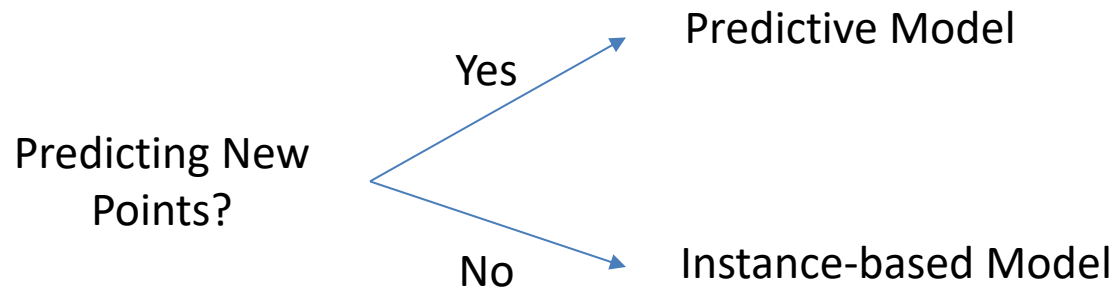
Adaptive to changes in the target or features (*concept drift & data drift*)



Types of Machine Learning Systems



Types of Machine Learning Systems



Example:

An (instance-based) spam filter (supervised learning) trained on the fly (online learning)

Stages of a Machine Learning Project

```
graph TD; A((Business understanding)) --- B[Exploratory Data Analysis (EDA)]; B --- C[Data Preparation]; C --- D[Modelling]; D --- E[Evaluation]; E --- F[Deployment];
```

Business understanding

Exploratory Data Analysis (EDA)

Data Preparation

Modelling

Evaluation

Deployment

Stages of a Machine Learning Project

Business understanding

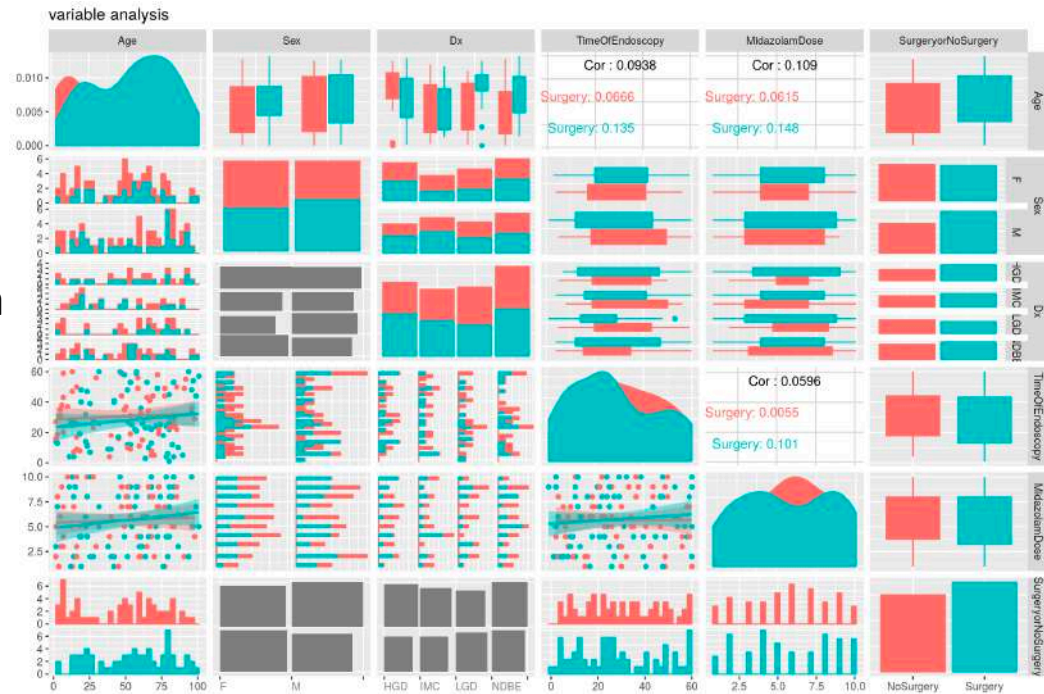
- Understand the business objectives
- Define the problem and the goals
- Create a project plan
- Ask relevant questions



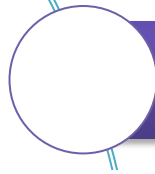
Stages of a Machine Learning Project

Exploratory Data Analysis (EDA)

- Identify data sources
- Get familiarized with the data
- Discover initial insights from the data
- Look for correlations in the data
- Form a hypothesis using the data
- Define an architecture for ETL



Stages of a Machine Learning Project



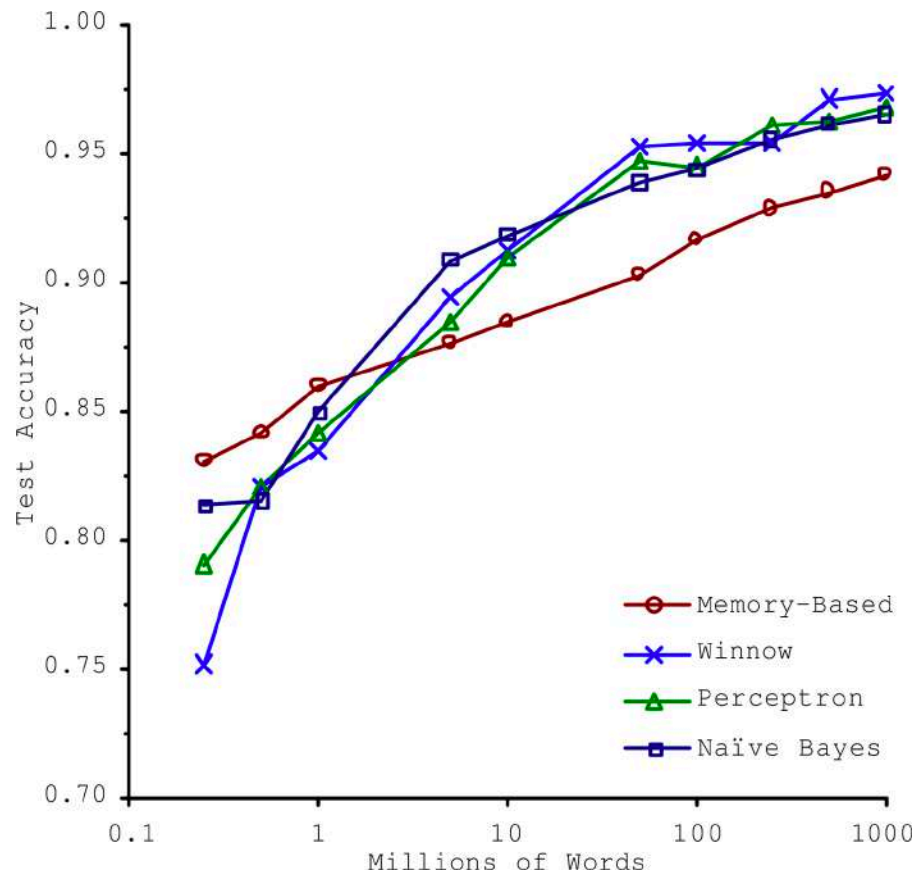
Exploratory Data Analysis (EDA)

Even the best learning algorithms on wrong data produce wrong results.

- Alessandro Negro, 2015

Stages of a Machine Learning Project

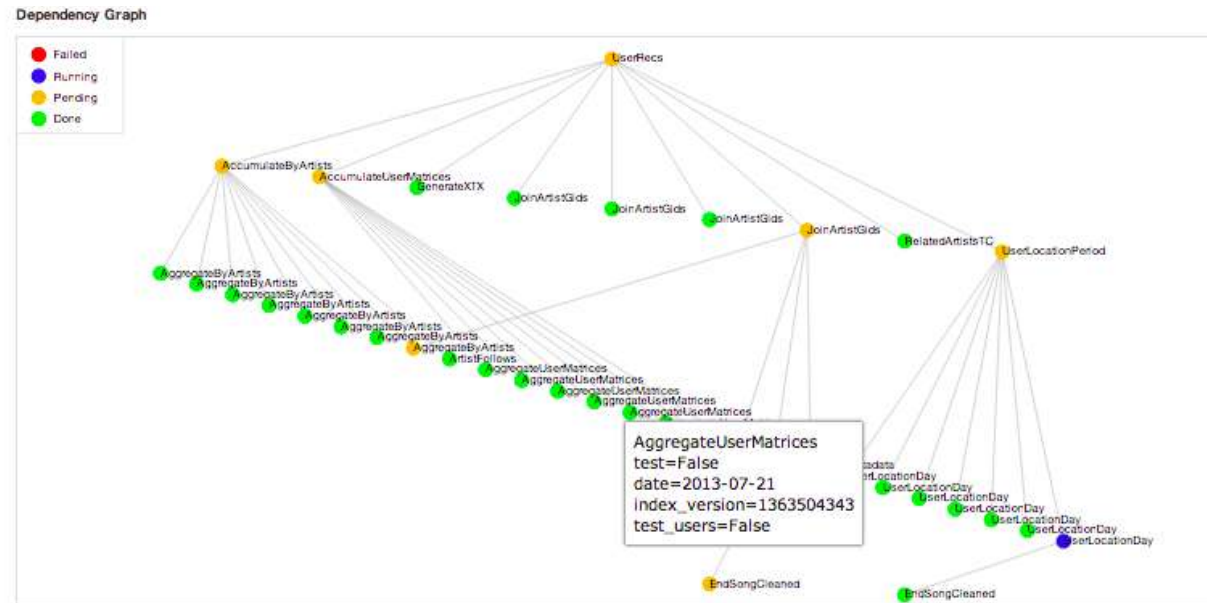
Exploratory Data Analysis (EDA)



Stages of a Machine Learning Project

Data Preparation & Cleaning

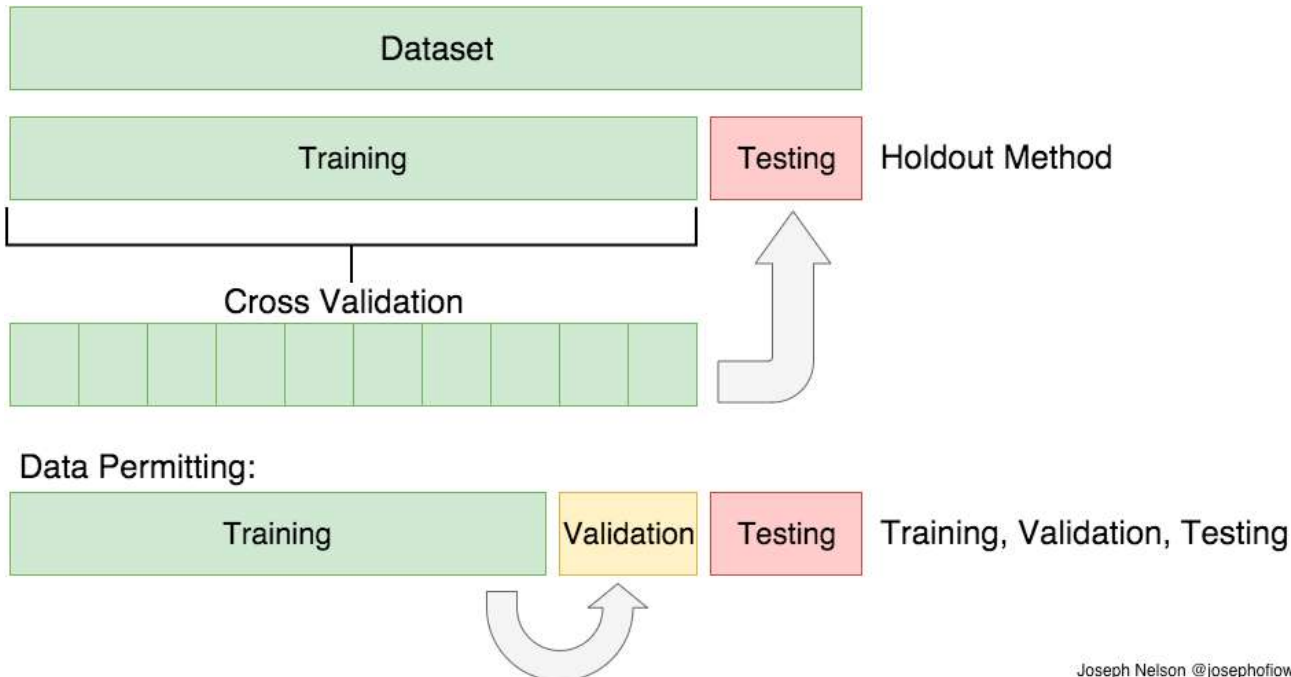
- Fix issues in the data (ex: missing values)
- Merge / aggregate the data
- Do feature engineering
- Organize it for an algorithm-specific structure
- Create a data pipeline



Stages of a Machine Learning Project

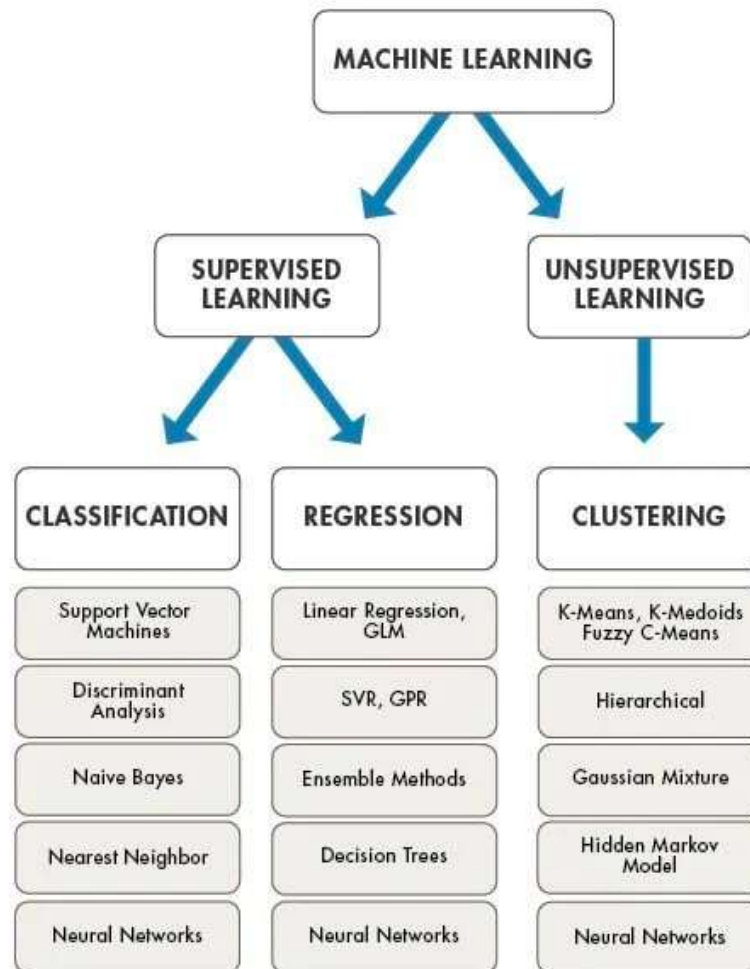
Modelling

- Train a machine learning model
- Evaluate the performance of the model
- Make predictions on a holdout/test set



Stages of a Machine Learning Project

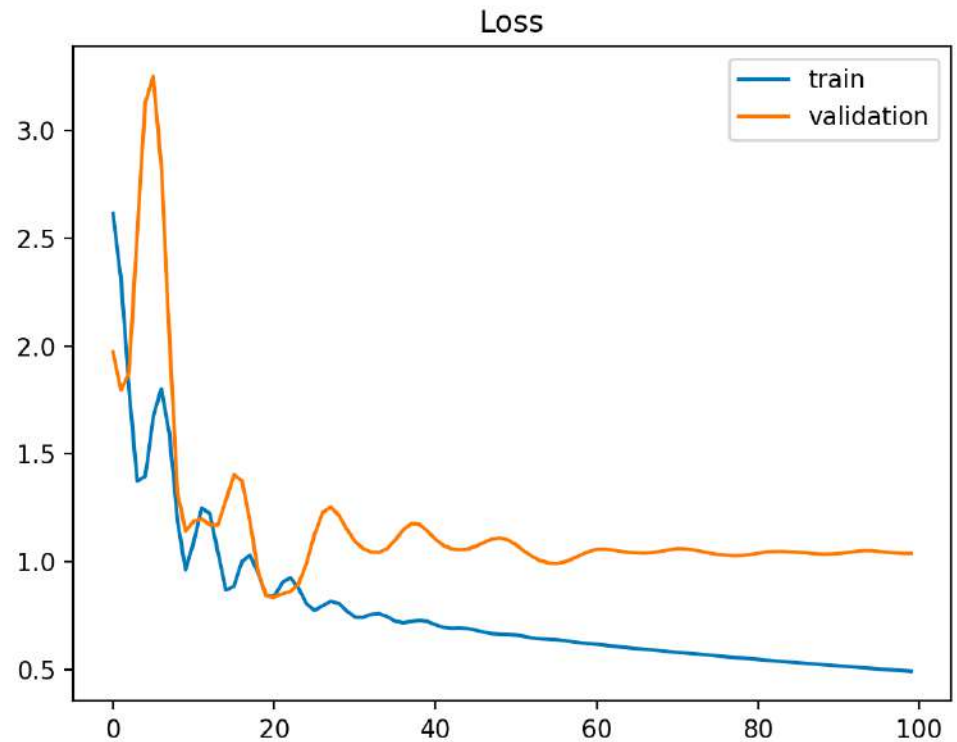
Modelling



Stages of a Machine Learning Project

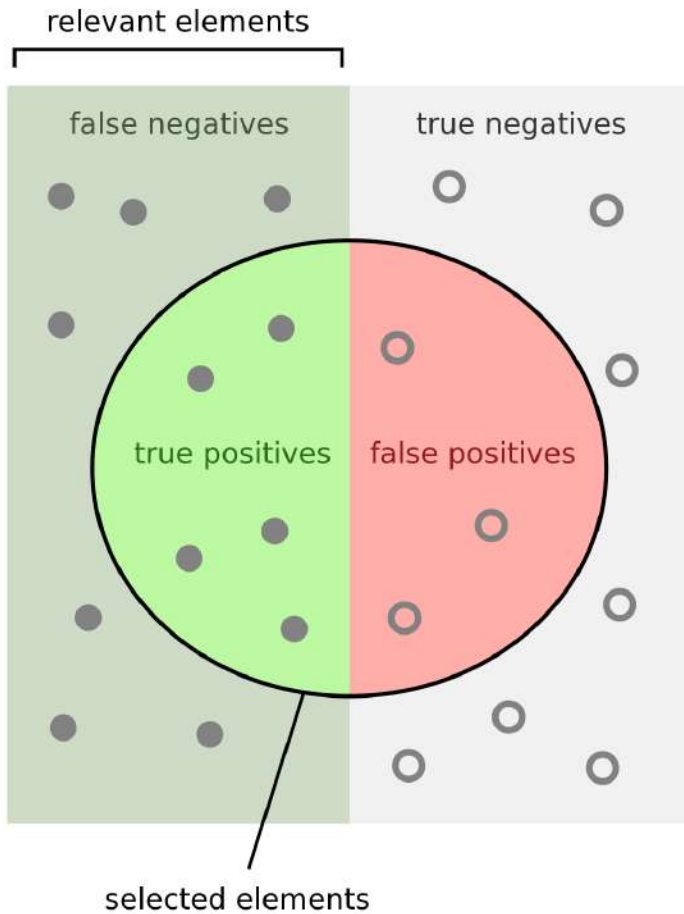
Evaluation

- Select a performance measure
- Analyze the best models
- Fine tune the model



Stages of a Machine Learning Project

Evaluation



How many selected items are relevant?

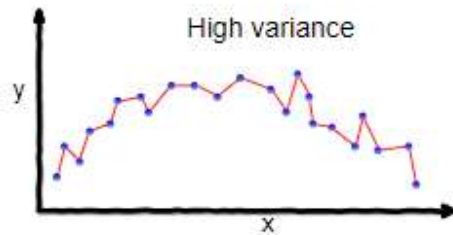
$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are selected?

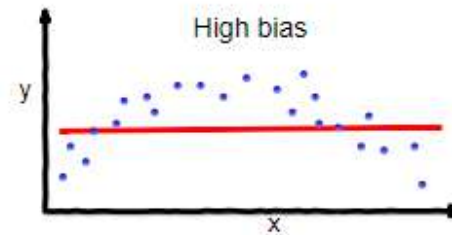
$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

Stages of a Machine Learning Project

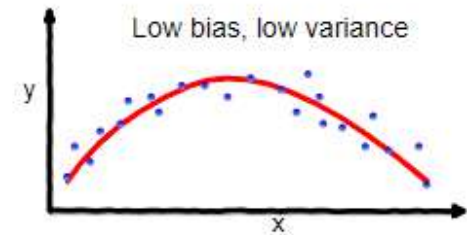
Evaluation



overfitting



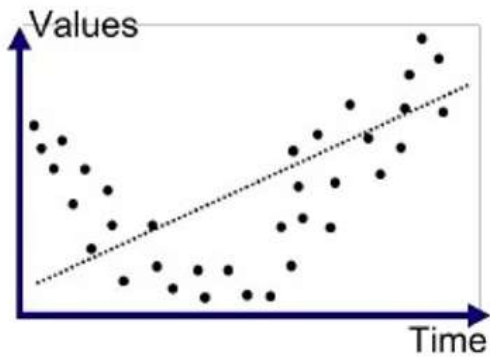
underfitting



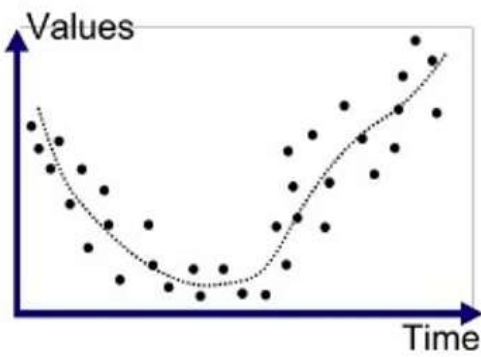
Good balance

Stages of a Machine Learning Project

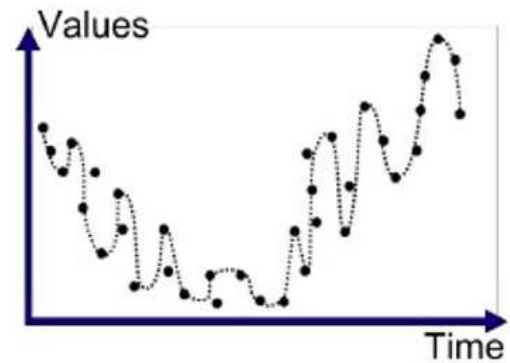
Evaluation



Underfitted



Good Fit/Robust

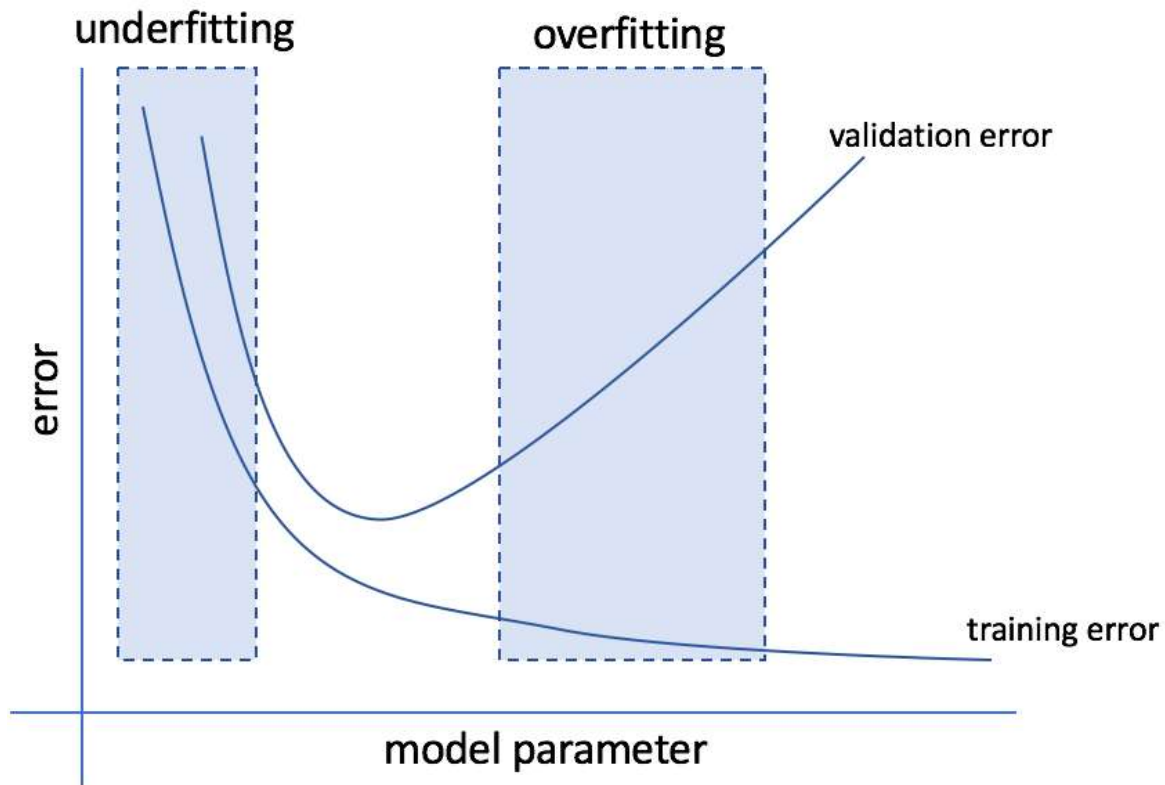


Overfitted

Stages of a Machine Learning Project

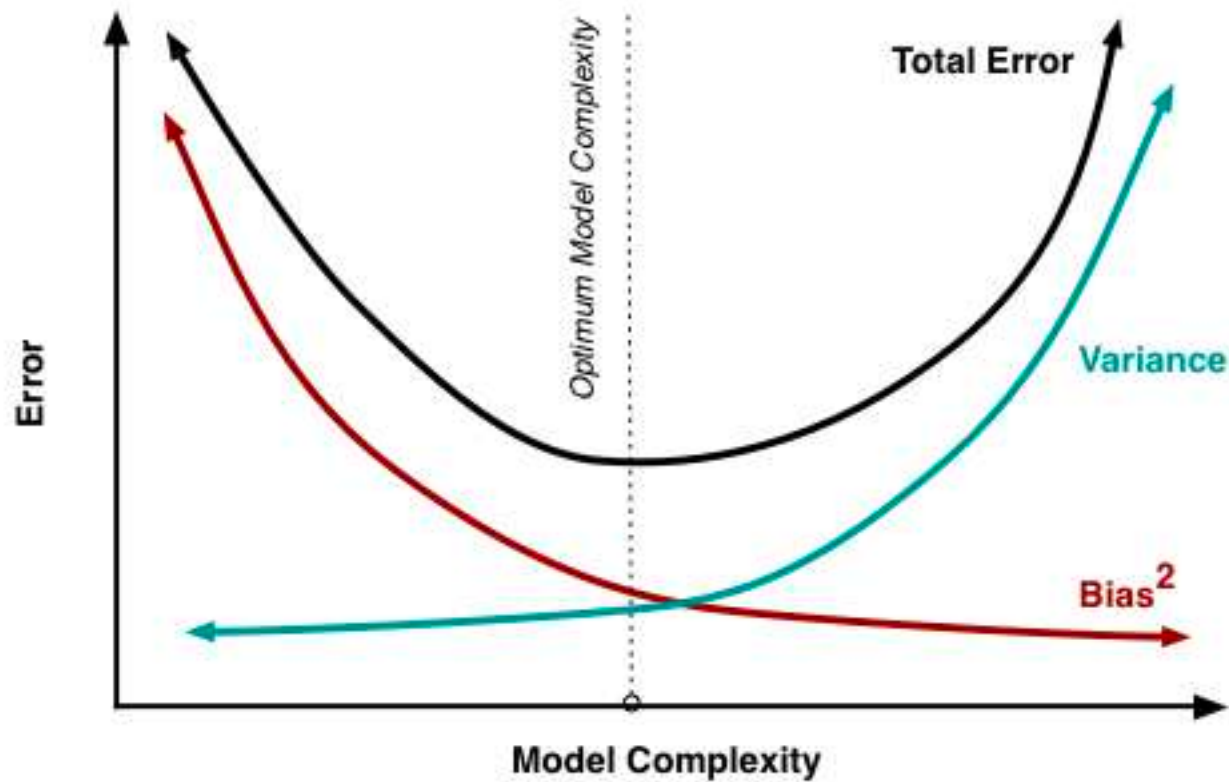


Evaluation



Stages of a Machine Learning Project

Evaluation



Stages of a Machine Learning Project



Deployment

- Define the architecture
- Deploy the project in a production environment
- Monitor the performance of the model
- Communicate findings to key stakeholders (using plots and interactive visualizations)