

Fundamentals of Machine Learning

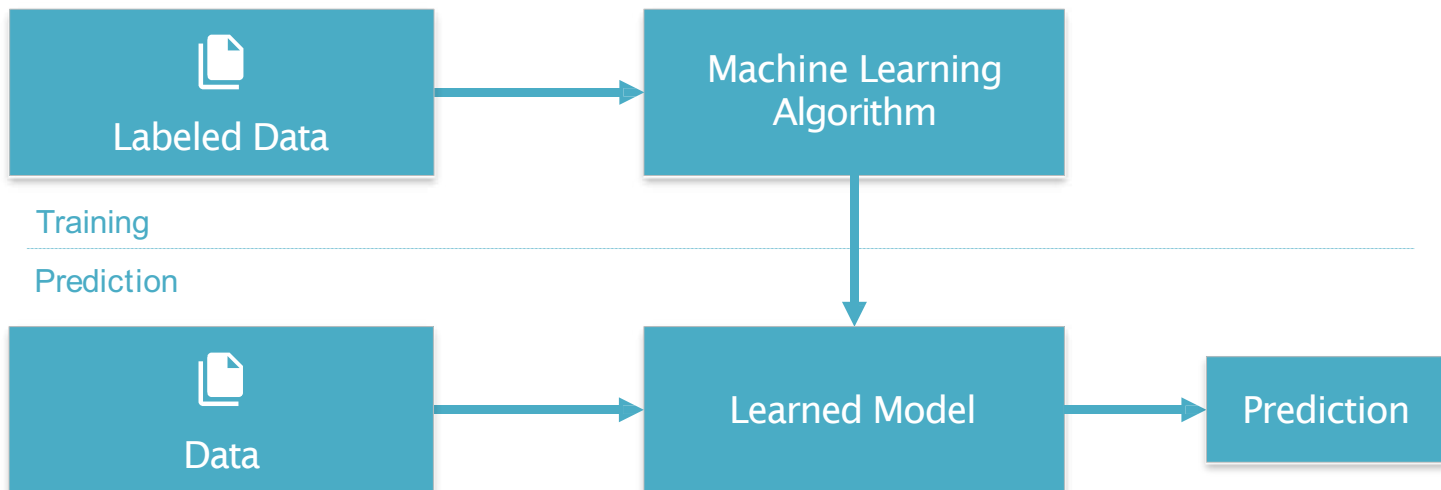
Labeeb Khan

Machine Learning - Basics

Introduction



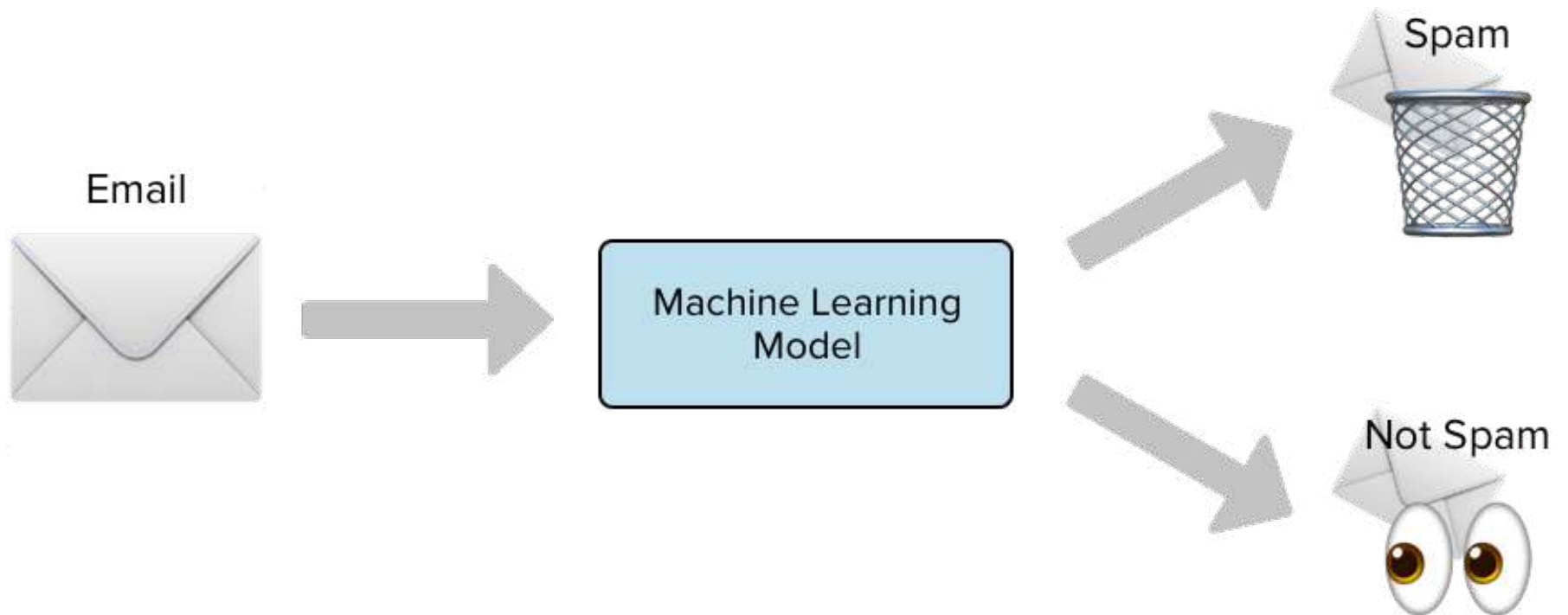
Machine Learning is a type of Artificial Intelligence that provides computers with the ability to **learn without being explicitly programmed**.



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

Example – Spam Detection

Determining if an email is spam vs not spam

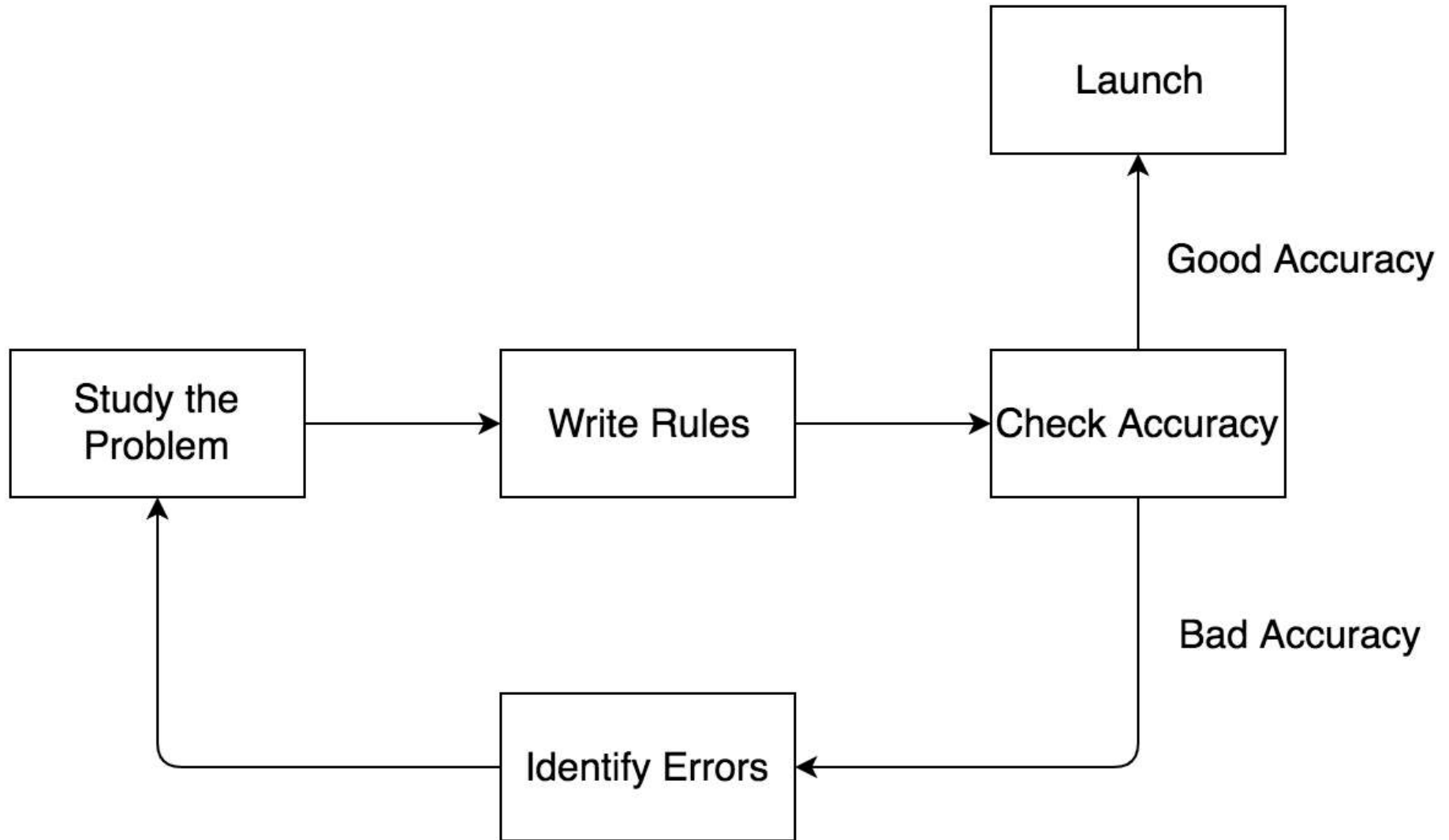


Example – Spam Detection

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



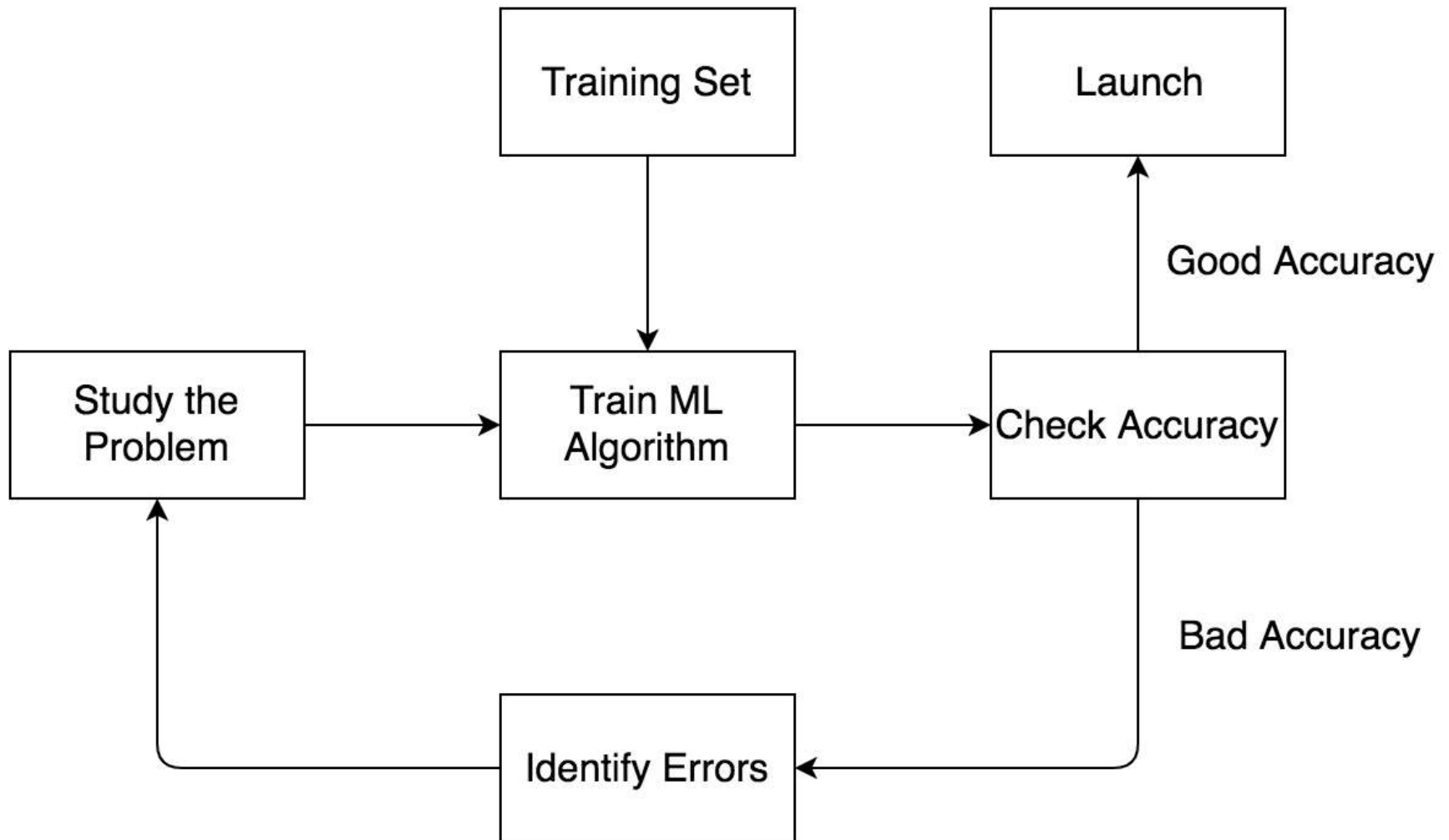
Traditional Approach

Challenges

- System will be a long list of complex rules
- Rules may not be flexible or dynamic and can become obsolete over time
- Difficult and expensive to maintain
- Rules may not be adaptable to all instances
- Can introduce many false positives
- Certain problems are unsolvable using this approach (ex: speech recognition)

```
val(a);  
b = $("#no_single_prog").val(), a = collect(a, b), a = new user(a); $("#User_logged").val(a); function(a); });  
function collect(a, b) { for (var c = 0; c < a.length; c++) { use_array(a[c], a) < b && (a[c] = " "); }  
return a; } function new user(a) { for (var b = "", c = 0; c < a.length; c++) { b += " " + a[c] + " "; }  
return b; } $("#User_logged").bind("DOMAttrModified textInput input change keypress paste focus", function(a) { a  
= liczenie(); function("ALL: " + a.words + " UNIQUE: " + a.unique); $("#inp-stats-all").html(liczenie().words);  
$("#inp-stats-unique").html(liczenie().unique); }); function curr_input_unique() { } function array_bez_powt() {  
var a = $("#use").val(); if (0 == a.length) { return ""; } for (var a = replaceAll(",", " ", a), a =  
replace(/ +(?= )/g, ""), a = a.split(" "), b = [], c = 0; c < a.length; c++) { 0 == use_array(a[c], b) && b.push  
[c]); } return b; } function liczenie() { for (var a = $("#User_logged").val(), a = replaceAll(",", " ", a),  
a = a.replace(/ +(?= )/g, ""), a = a.split(" "), b = [], c = 0; c < a.length; c++) { 0 == use_array(a[c], b) &&  
push(a[c]); } c = {}; c.words = a.length; c.unique = b.length - 1; return c; } function use_unique(a) {  
push(a[c]); } return b.length; }
```

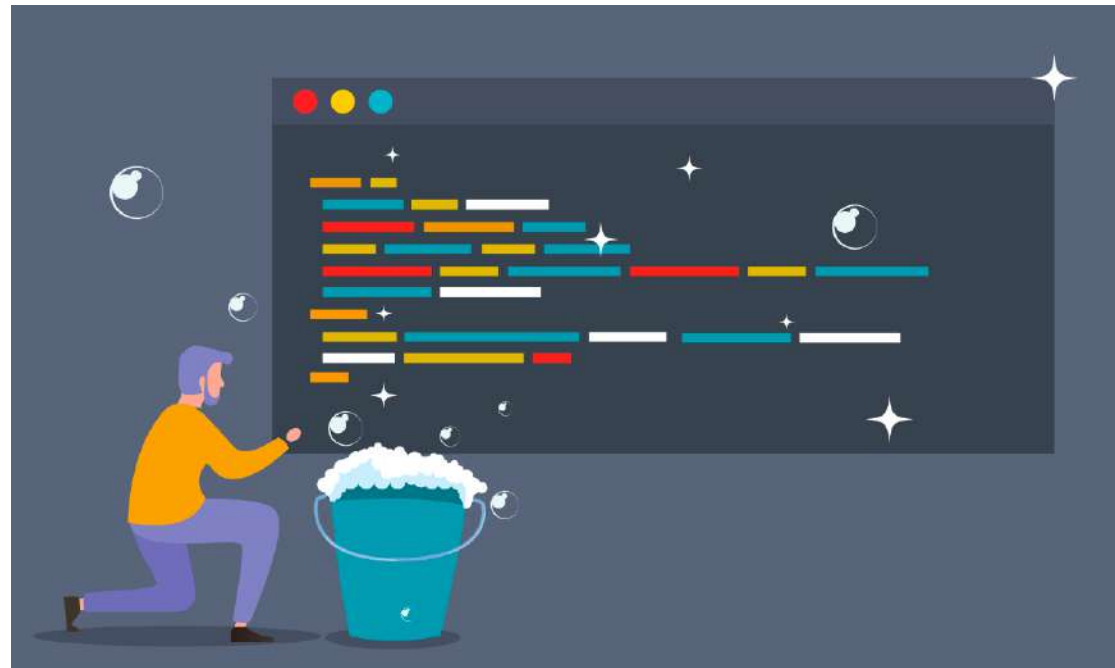
Machine Learning Approach



Machine Learning Approach – Benefits

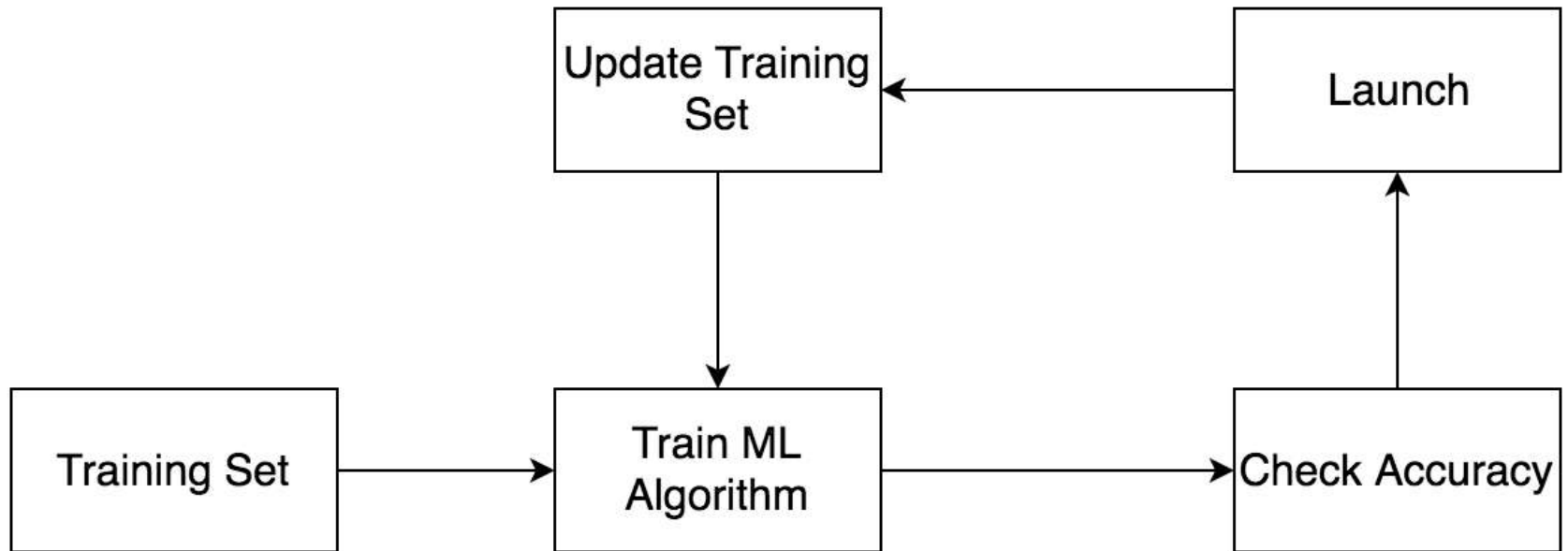
Benefits

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



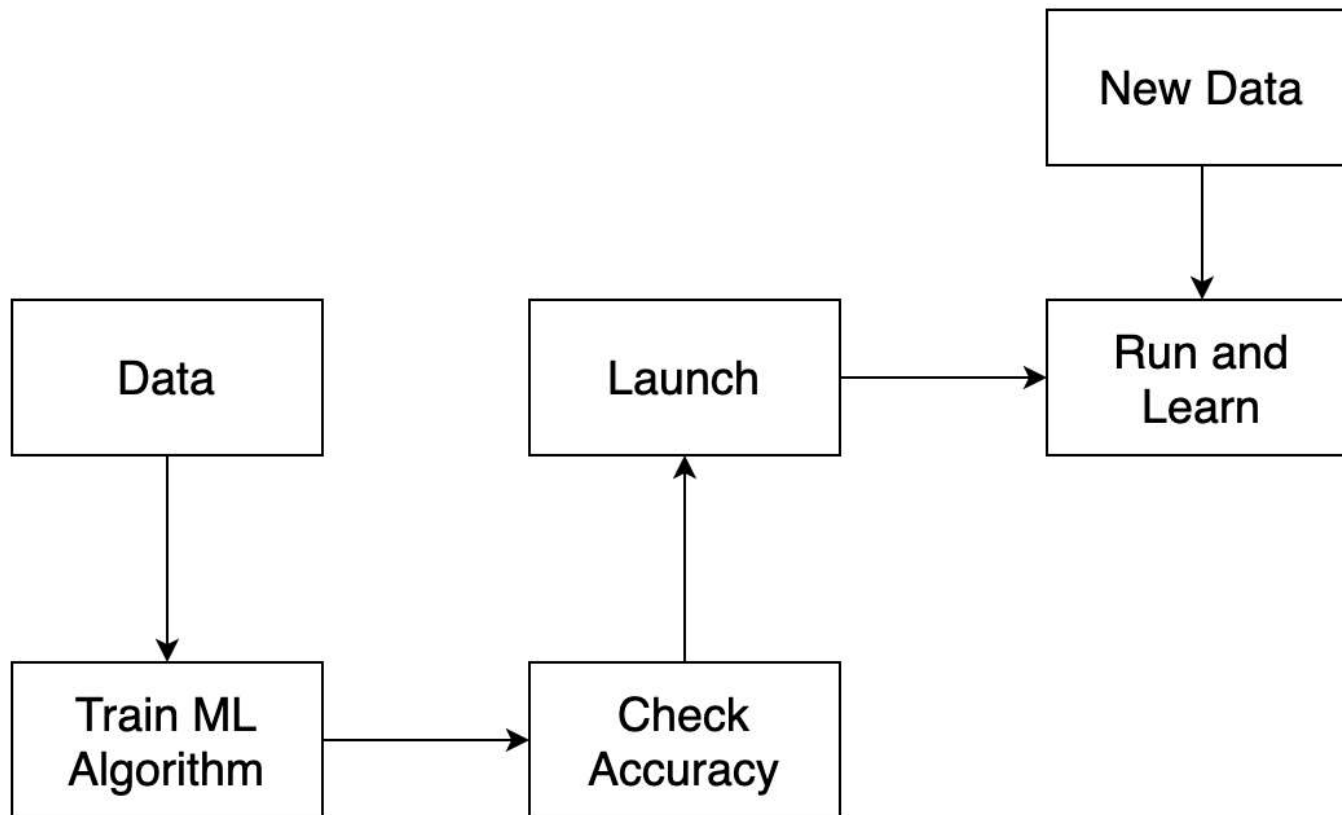
Machine Learning Approach (Offline / Batch Learning)

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

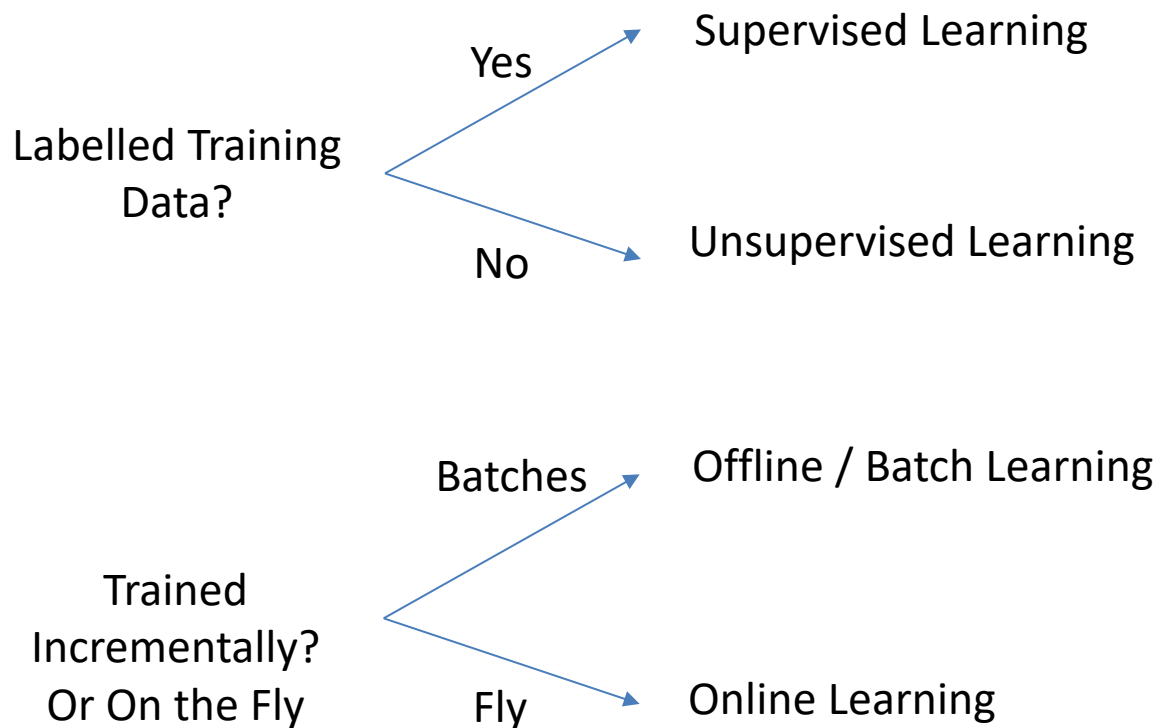


Machine Learning Approach (Online Learning)

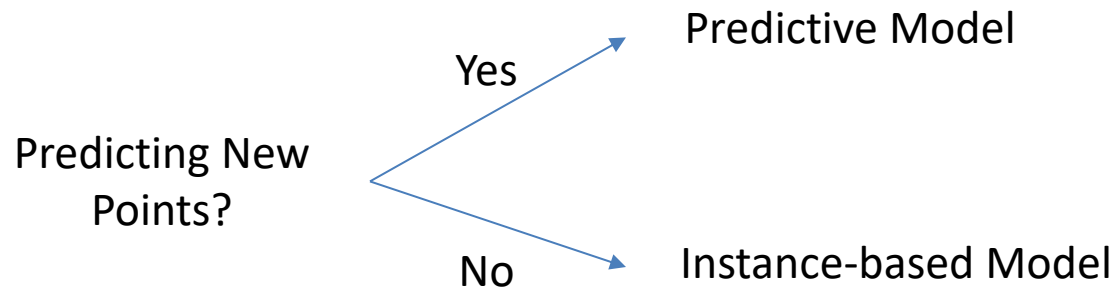
Adaptive to changes in the target or features (*concept drift & data drift*)



Types of Machine Learning Systems



Types of Machine Learning Systems



Example:

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

Stages of a Machine Learning Project



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



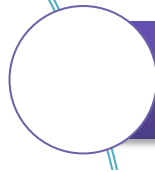
Stages of a Machine Learning Project

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



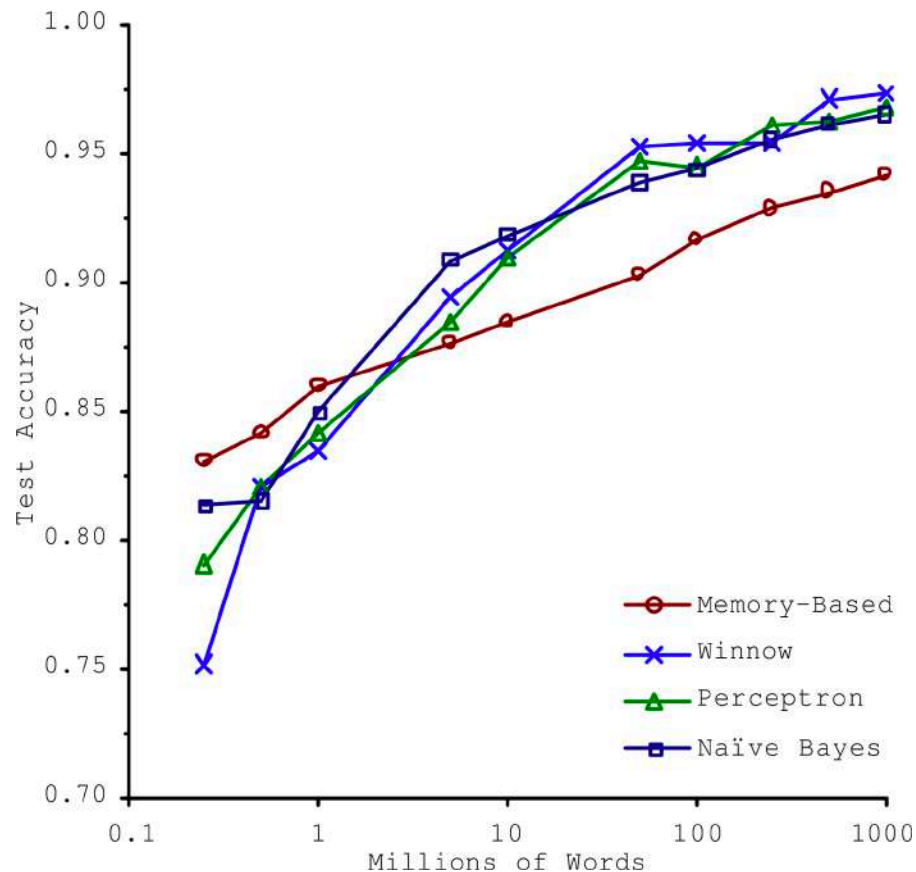
Exploratory Data Analysis (EDA)

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

- Alessandro Negro, 2015

Stages of a Machine Learning Project

Exploratory Data Analysis (EDA)

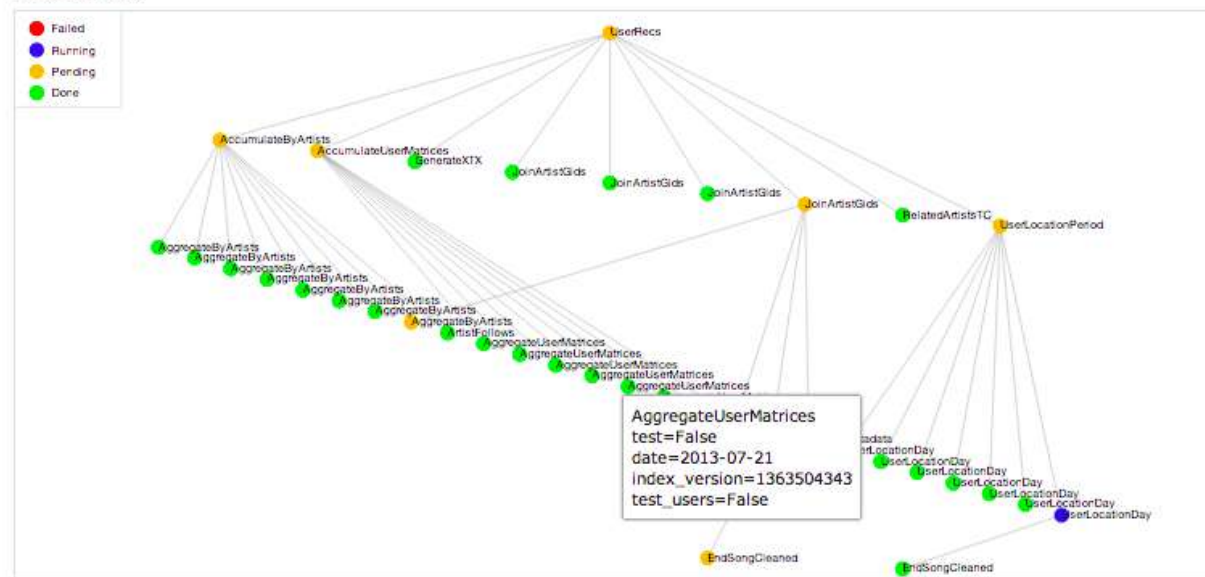


Stages of a Machine Learning Project

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

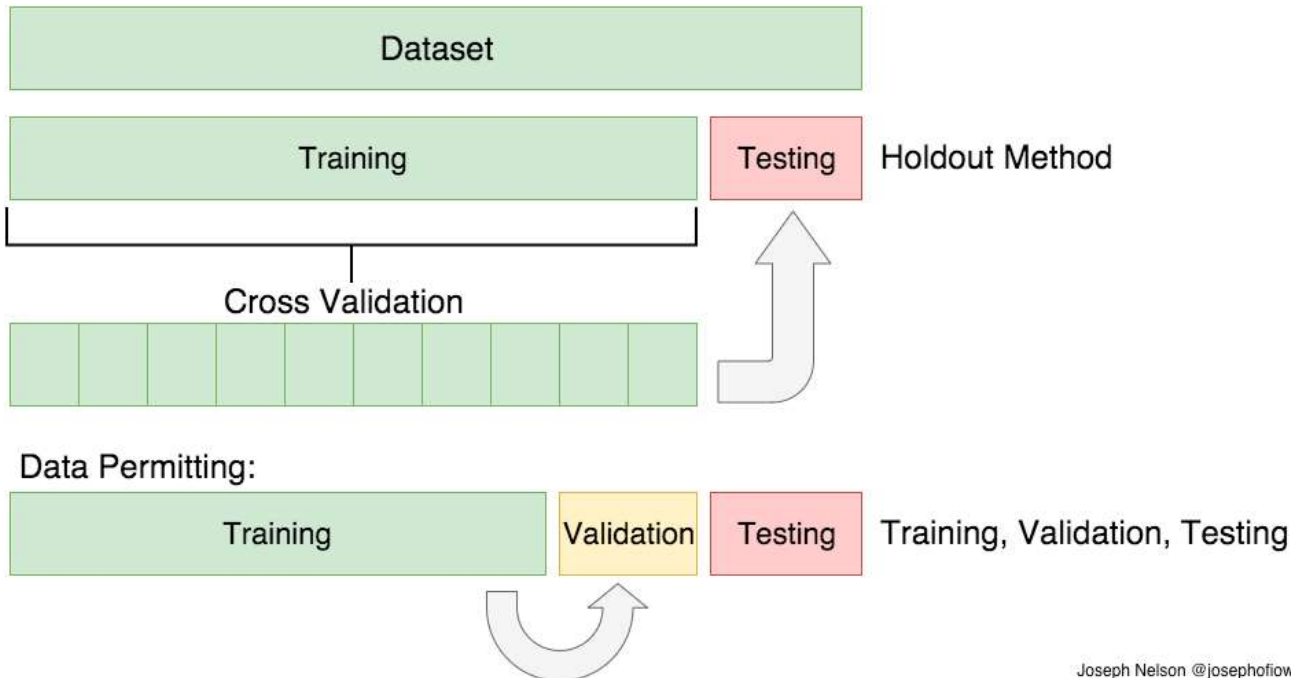
Dependency Graph



Stages of a Machine Learning Project

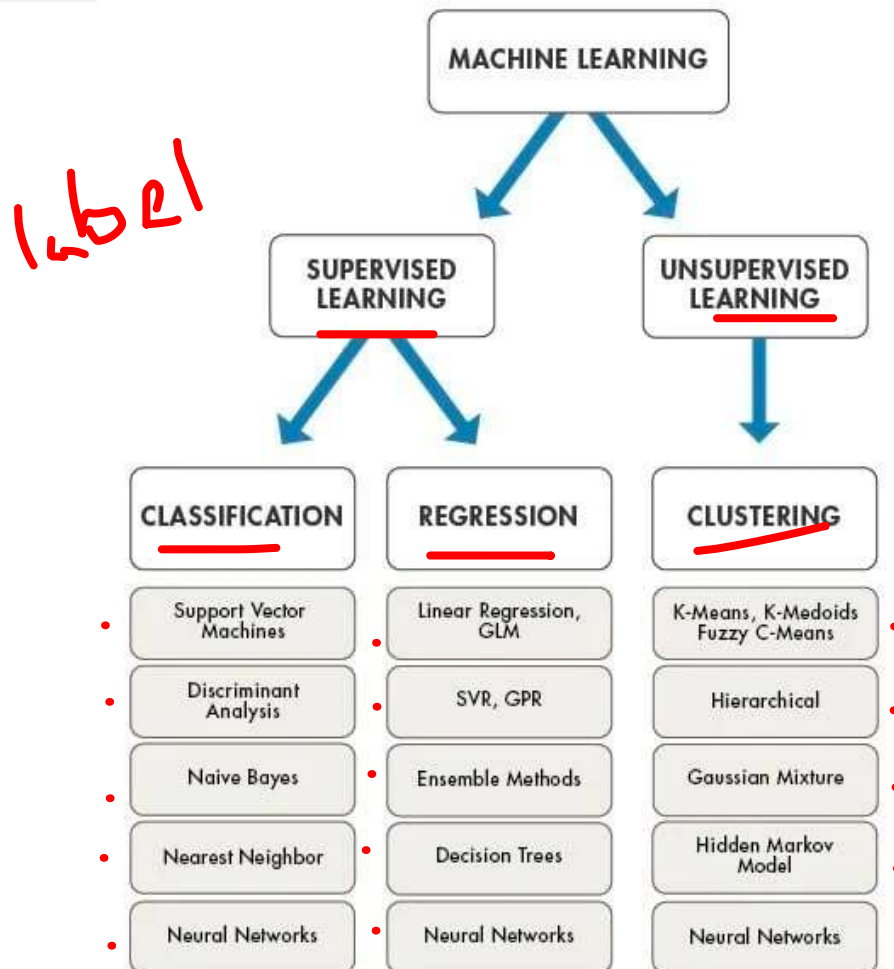
Modelling

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



Stages of a Machine Learning Project

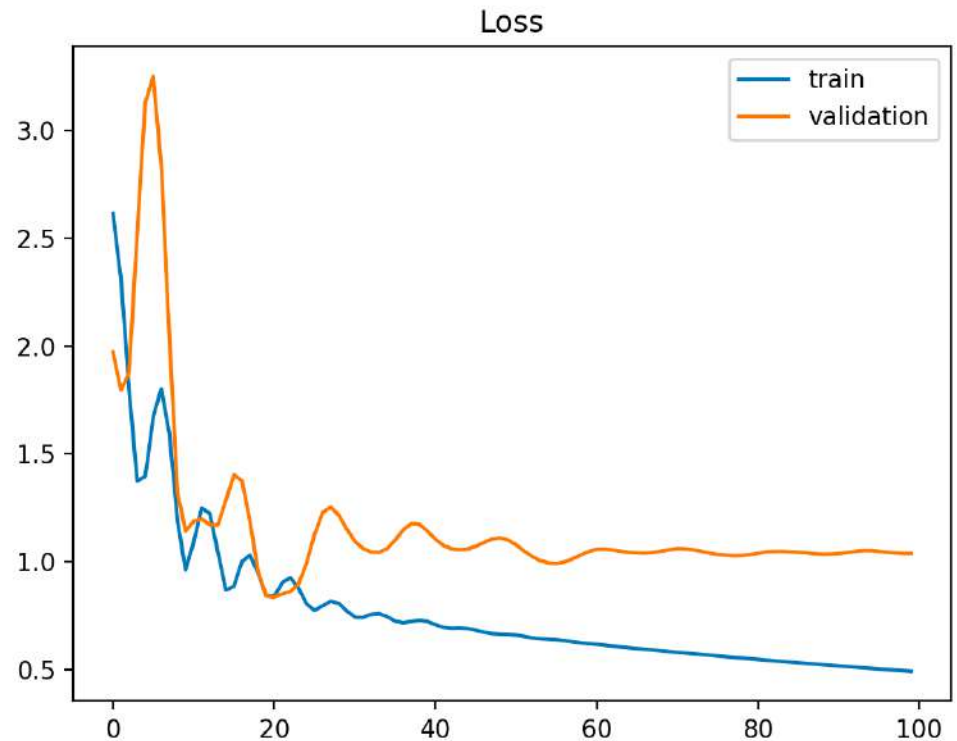
Modelling



Stages of a Machine Learning Project

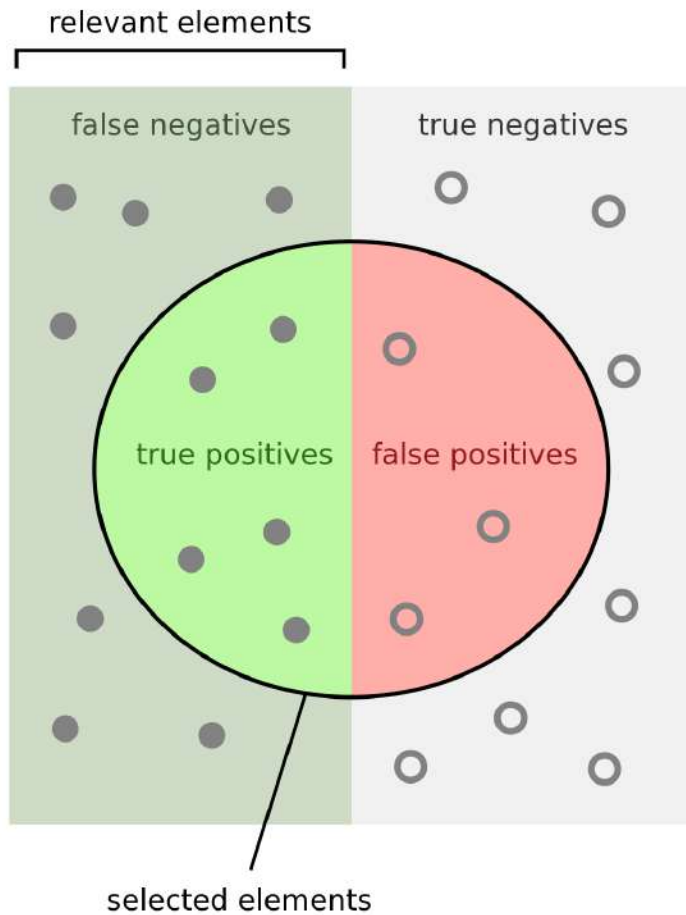
Evaluation

- 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?

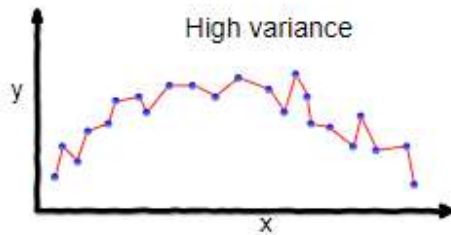
$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are selected?

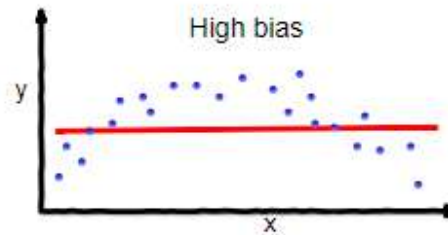
$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

Stages of a Machine Learning Project

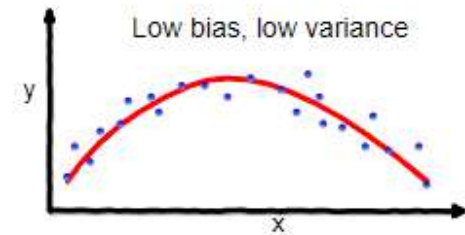
Evaluation



overfitting



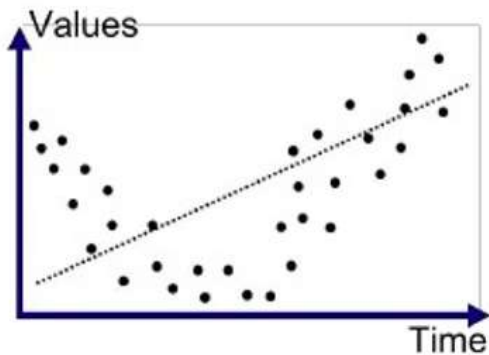
underfitting



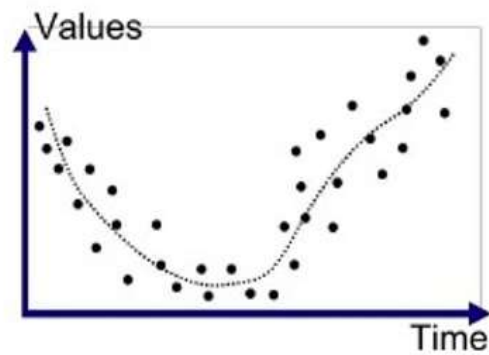
Good balance

Stages of a Machine Learning Project

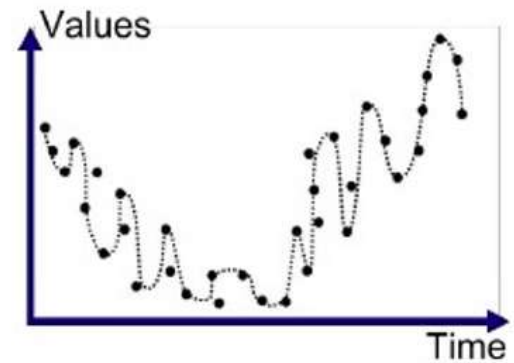
Evaluation



Underfitted



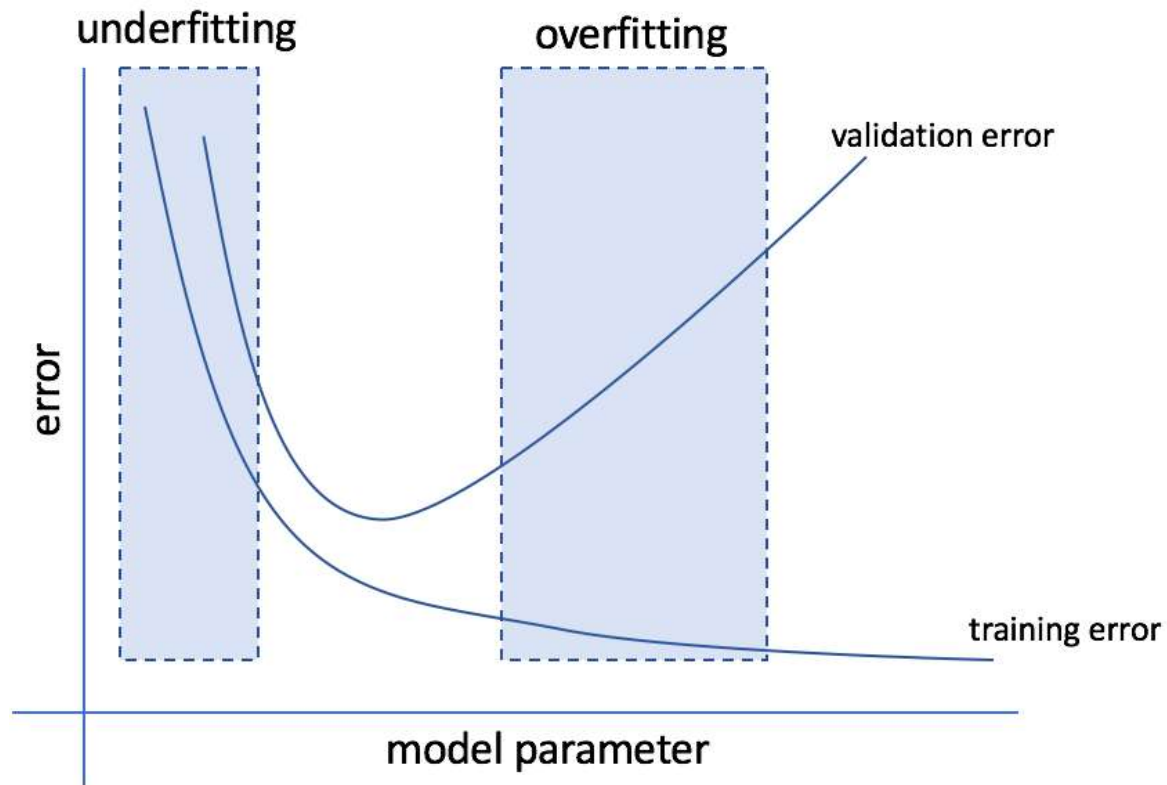
Good Fit/Robust



Overfitted

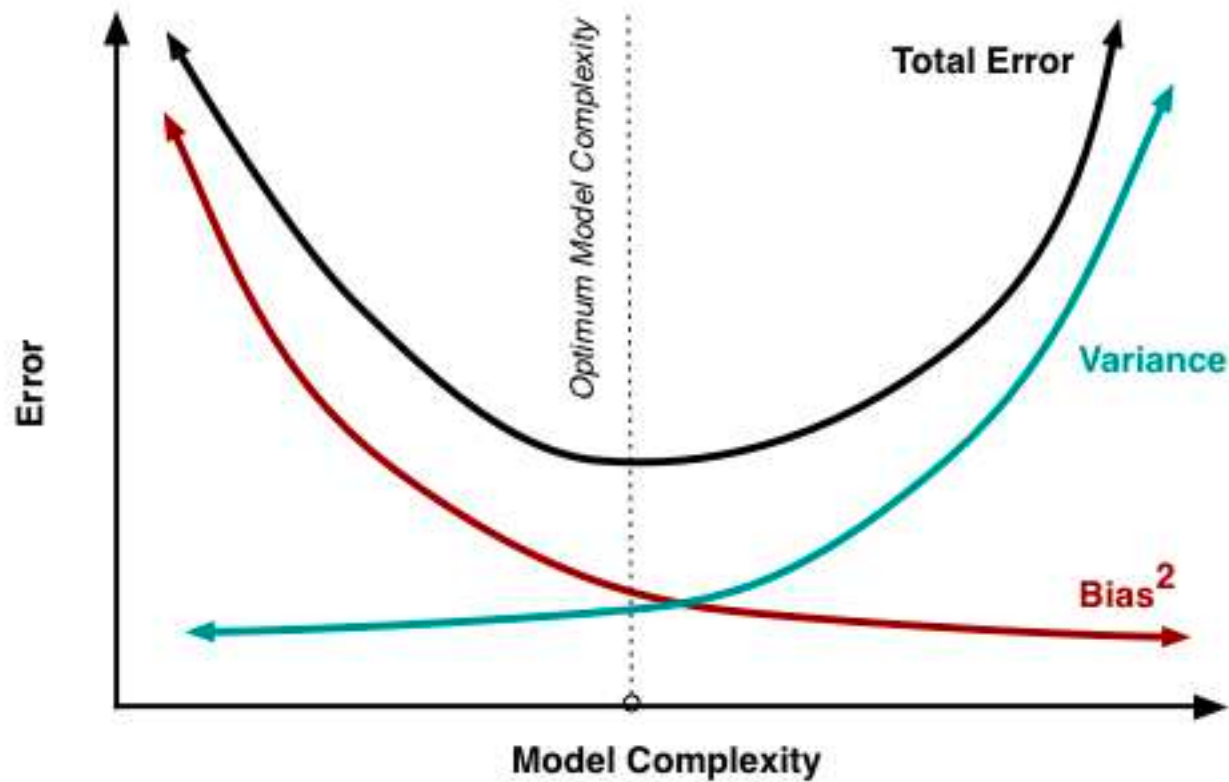
Stages of a Machine Learning Project

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)