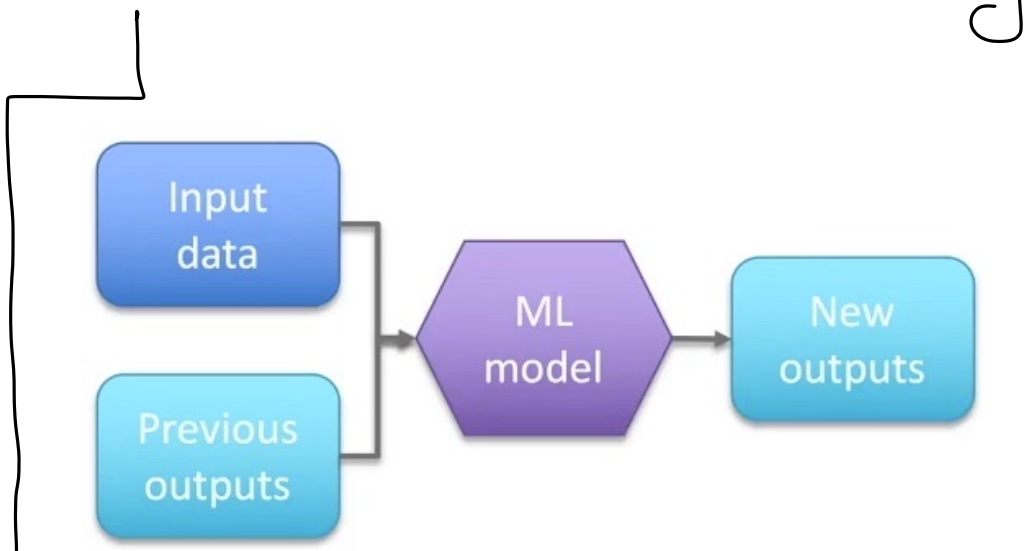
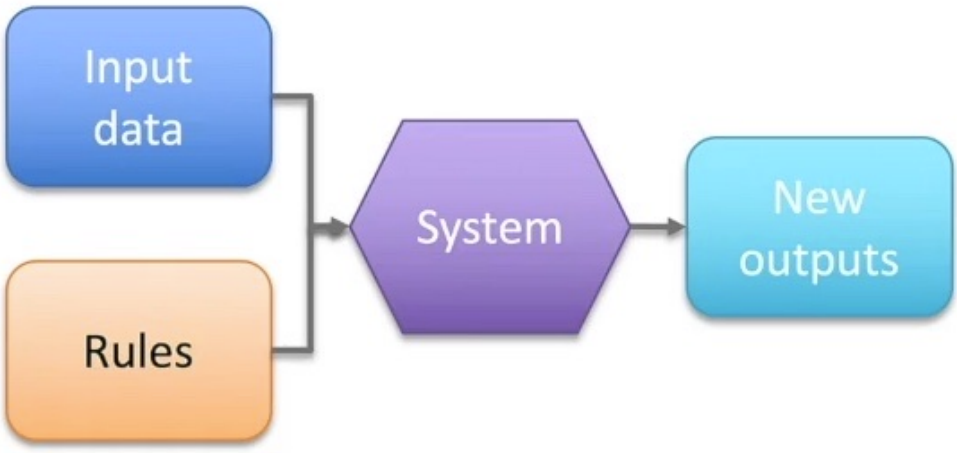
Machine learning Foundations for Product Managers

⚠ what is Machine Learning

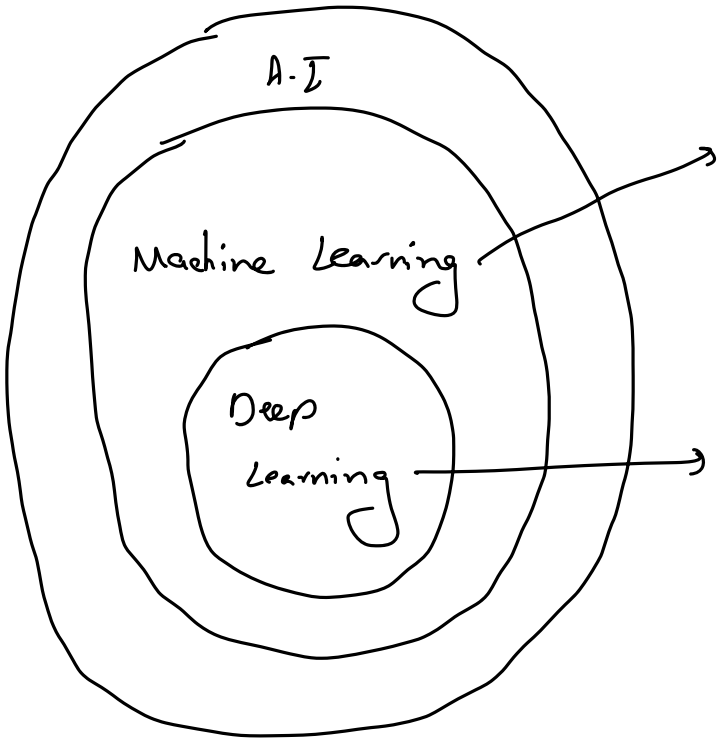
↳ field of study that gives computers the ability to learn without being explicitly programmed

⚠ Traditional software vs. Machine learning





⚠ AI vs. Machine Learning vs. Deep Learning



set of methods & tools which help realize the goal of the field of A.I

the use of neural networks containing many layers

⚠ Structured data vs. unstructed data

Structured Data	Unstructured Data
set structure based on pre-defined fields for each record	does not follow a defined format of fields
often stored in relational databases	Many types - images, videos sounds, text
easy to enter, search and analyze	requires specialized tools to work with
works well with command tools	

⚠ Continuous Data, Categorical Data, Discrete Data

- Continuous Data → temperature, height, time
- Categorical Data → gender, color
- Discrete Data → # of apples

⚠ A model is an approximation of the relationship between variables

⚠ To create a model, we define four things

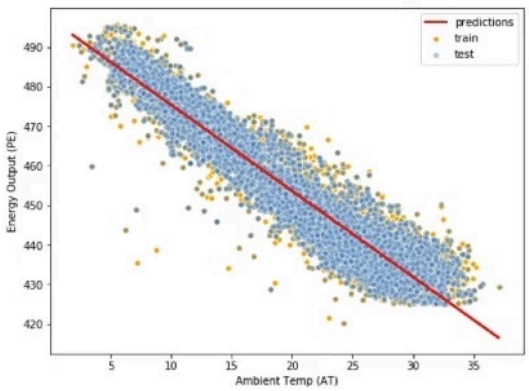
- Features ~ to use
- Algorithm ~ acts as a form / template for model
- Hyperparameter ~ values for algorithm
- Loss function ~ to optimize

⚠ Types of Machine Learning

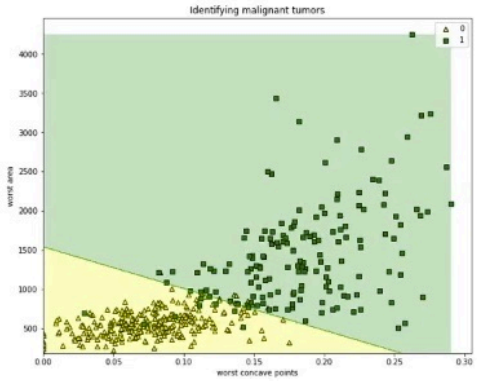
	Supervised Learning	Unsupervised Learning	Reinforcement Learning
Objective	Prediction of a target variable	Organize data by inherent structure	Learn strategies via interaction
Learning Task(s)	Classification Regression	Clustering Anomaly detection	Achieve a goal
Target Data Required?	Yes	No	Yes, but delayed
Examples	• Identifying pneumonia from xray images • Predicting real estate prices	• Market segmentation • Identifying fraudulent activity	• AlphaZero • Autonomous vehicles

⚠ Regression vs. Classification

- Predict one or more **numerical** target variables
- E.g. home price, number of power outages, product demand



- Predicts a **class / category** – either binary or out of a set
- E.g. lung disease detection, identifying types of plants, sentiment analysis, detecting spam



⚠ What ML can not do well

- understand context
- determine causation
- explain "why" things happen
- determine the impact of interventions / find solutions

⚠ CRISP-DM Process

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modeling
5. Evaluation
6. Deployment

⚠ Modeling Process

① Select Features ② choose algorithm

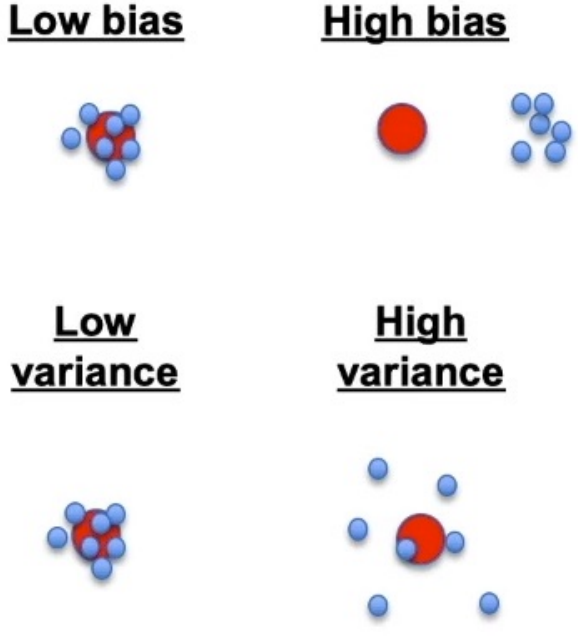
③ set hyperparameters ④ train model

⑤ Graduate Model

⚠ Features ⟷ characteristics of data

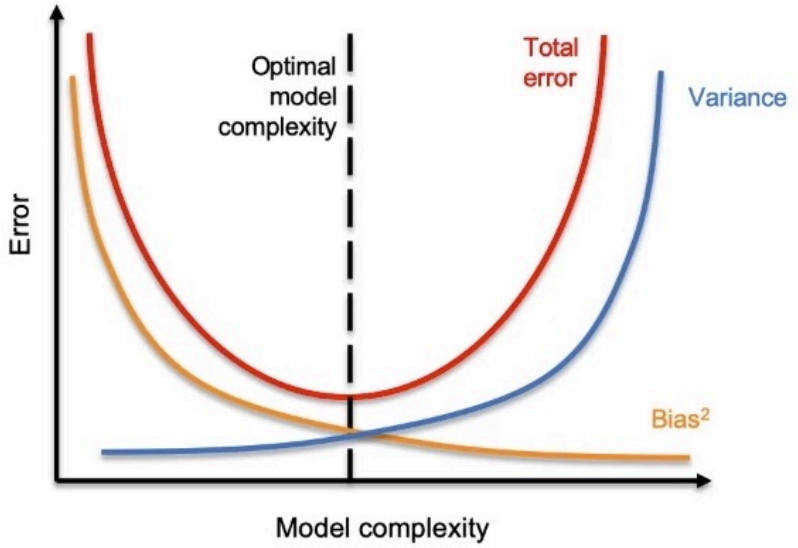
⚠ Bias is error introduced by modeling a real life problem using a simpler model that is unable to fully capture the underlying patterns in data

⚠ Variance refers to the sensitivity of model to small fluctuations in the data because it models fine patterns which may just be noise

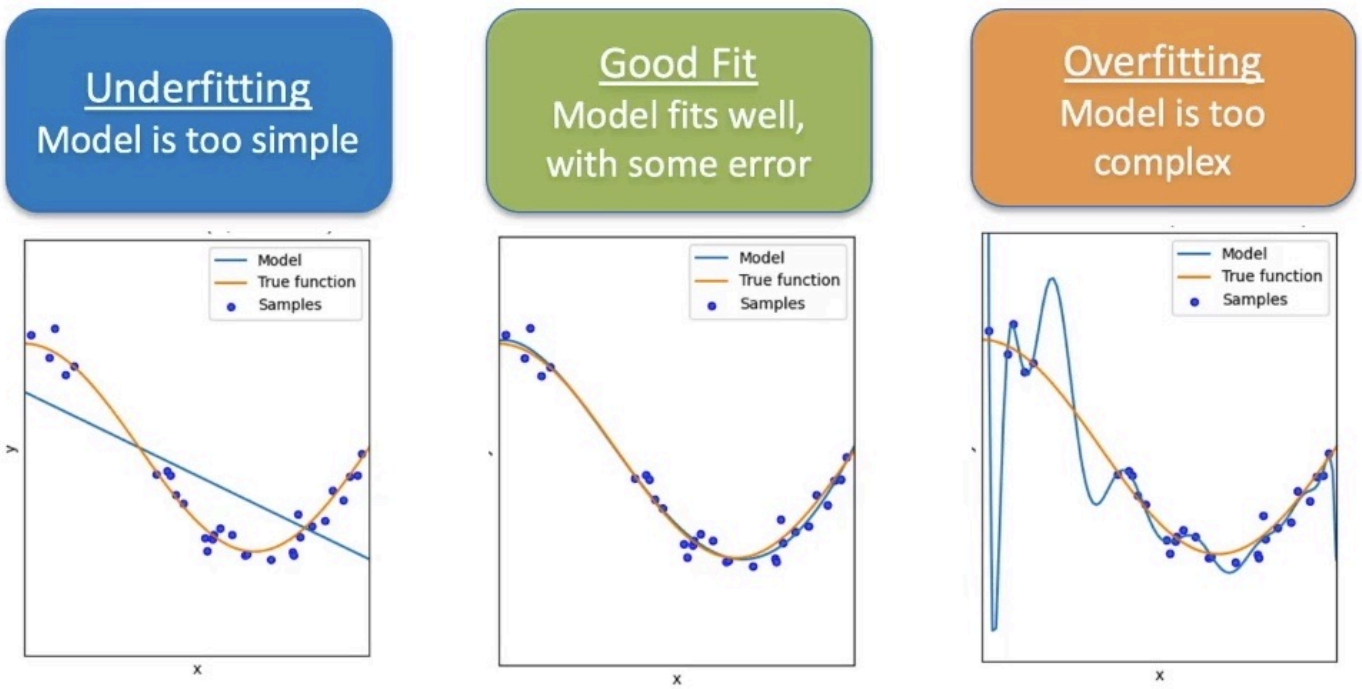


⚠ Bias - Variance Tradeoff

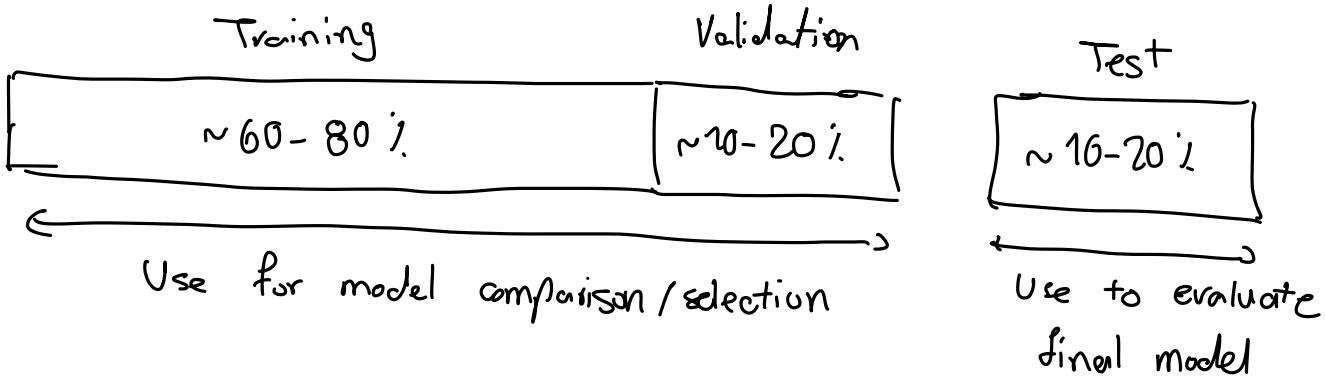
↳ Total Error = $Bias^2 + Var + \sigma_e^2$



⚠ Overfitting vs. underfitting

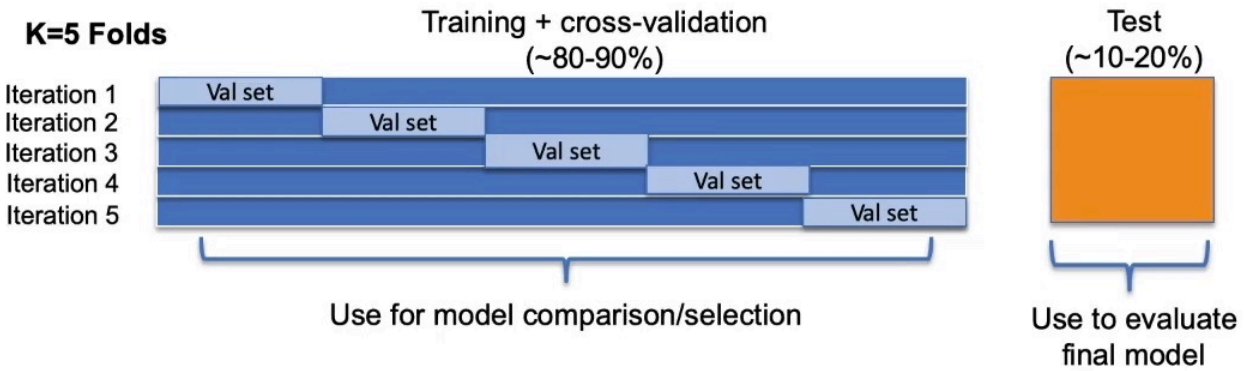


⚠ Training Data vs. Validation Data vs. Test Data



⚠ k-Folds Cross Validation

↳ rather than using a fixed validation set, we train & run the models multiple times, each time using a different subset as the validation set



↳ 5-Folds Cross Validation

⚠ Outcomes vs Outputs

	A tool to predict turbulence for airlines	A power demand forecasting tool for a utility
Outcome	• Low # of safety incidents per year, or lower $ of safety-related claims	• Lower cost per MWh of power produced • Lower emissions rate per MWh
Output	• Classification error metric (binary or 1-5 scale)	• Regression error metric

⚠ Mean Squared Error (MSE)

$$\text{MSE} = \frac{1}{n}\sum_i (y_i - \hat{y}_i)^2$$

y_i → actual target value

$\hat{y}_i$ → actual predicted value

⚠ Mean Absolute Error (MAE)

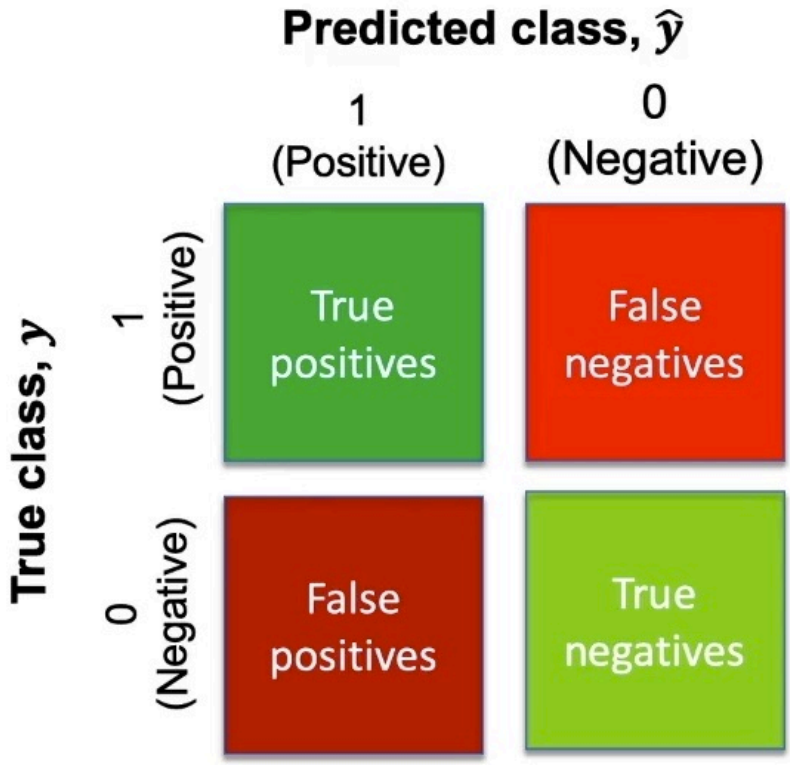
$$\text{MAE} = \frac{1}{n}\sum_i |y_i - \hat{y}_i|$$

⚠ Mean Absolute Percent Error (MAPE)

$$\text{MAPE} = \frac{1}{n}\sum_i \frac{|y_i - \hat{y}_i|}{y_i}$$

⚠ Confusion Matrix

(Binary classification)



⚠ $\text{Recall} = \frac{TP}{TP + FN}$ $\quad \text{Precision} = \frac{TP}{TP + FP}$

⚠ ROC Curves

↳ plots TP rate and FP rate for different threshold values

⚠ Linear Regression

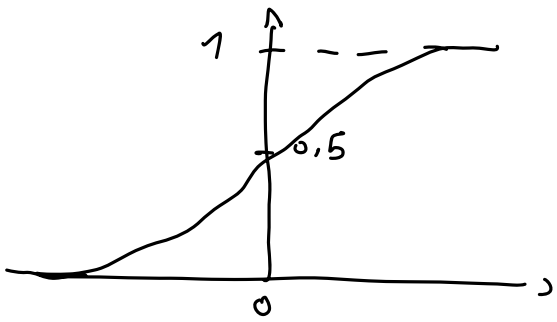
↳ Model which assumes linear relationships between features and targets, defined by a set of coefficients

↳ $y = w_1 x + w_0$ ← bias

↳ weight / coefficient

⚠ Sigmoid function

↳ $\sigma(z) = \frac{1}{1+e^{-z}}$



⚠ Logistic regression function gave us the probability of the positive class

⚠ Instead of the sigmoid function, we use the softmax function to give us the probability of belonging to each class

⚠ Decision Trees

↳ Ask a series of questions to narrow in on the label

⚠ Depth of a tree (max # of splits) is a hyperparameter

↳ shallow trees ⇝ underfitting

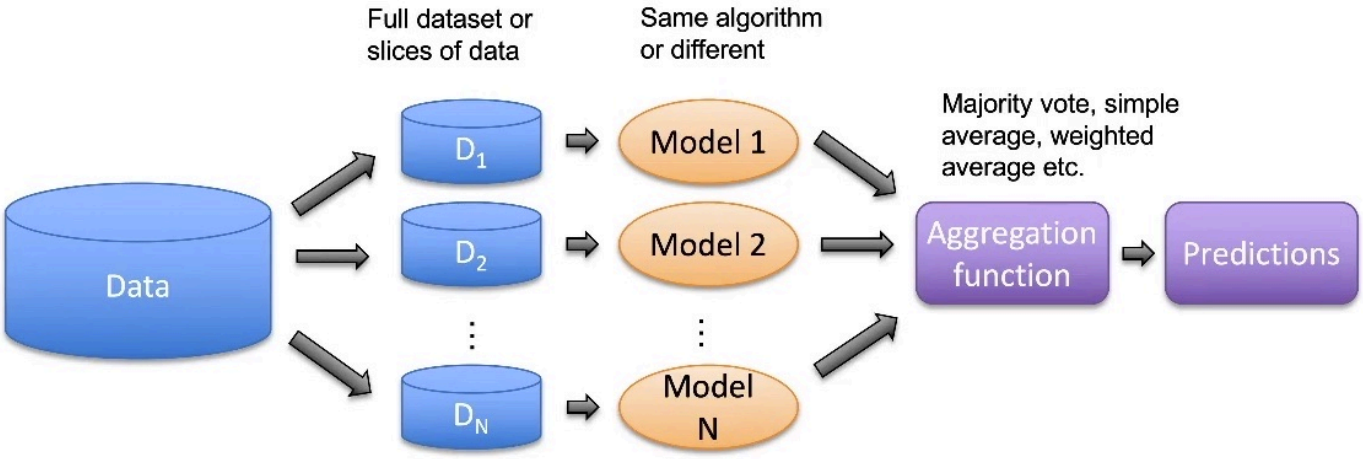
↳ Deep trees ⇝ overfitting

⚠ Regression Trees

↳ rather than taking the majority vote of samples in leaf, we calculate the mean target value of the samples

⚠ Ensemble Models

↳ Goal is to combine multiple models together into a meta-model that has better generalization performance



⚠ Clustering

↳ A technique used to organize data points into logical groups without using explicit group labels

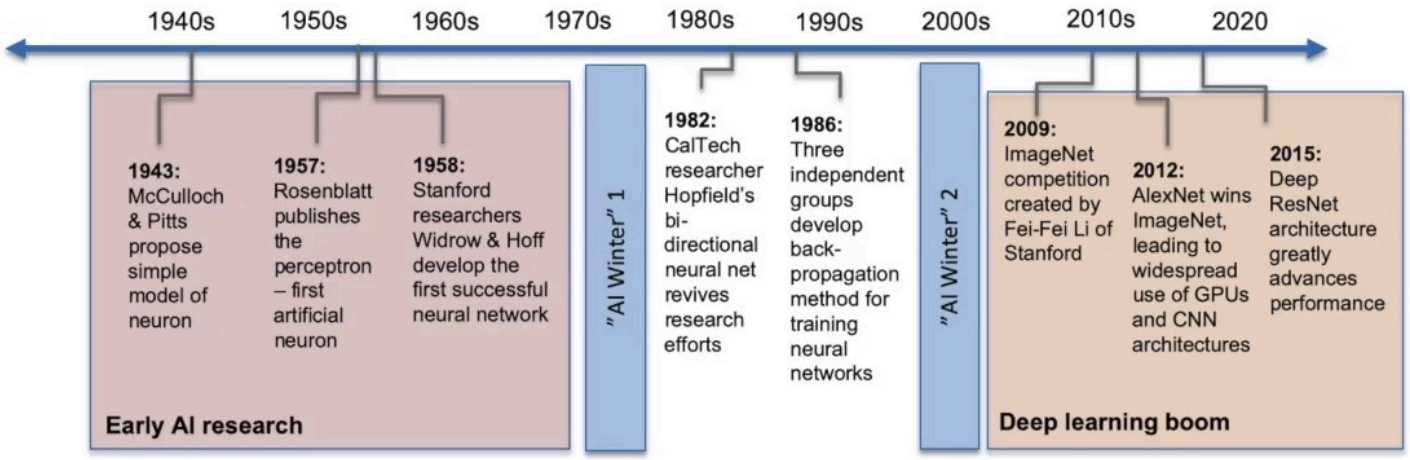
⚠ K-means Clustering

↳ groups points into clusters based on distance from the nearest cluster center

↳ objective: minimize the sum of the distances from each point to its assigned cluster center

⚠ A neural network with many layers is called a deep neural network

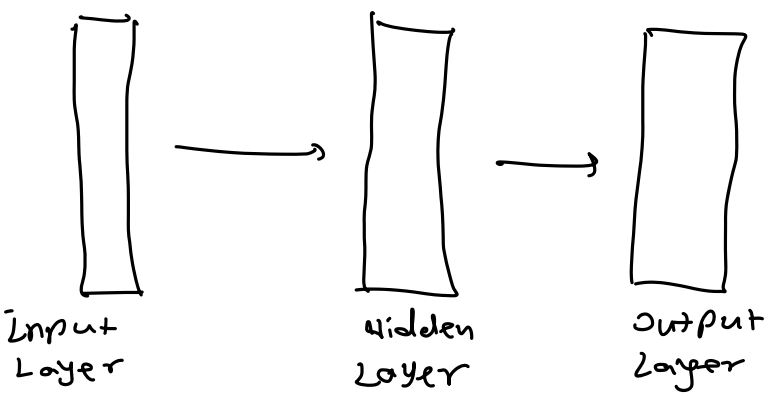
⚠ History of Neural Networks



⚠ Deep Learning excels in applications which have

- Vast amounts of training data
- Very large number of features
- Complex relationships between features & the target
- Low concern for explainability

⚠ Neural Network architecture



⚠ Computer Vision tasks

- Image classification
- object detection
- Semantic segmentation
- Image generation

⚠ Convolutional neural networks (CNNs)

- utilize two aditional types of layers
  - Convolutional → acts as a filter to learn patterns in data
  - pooling → reduce dimensions of data

⚠ Common Natural language Processing (NLP) tasks

- Text classification
- Sentiment analysis
- Search
- Machine translation
- text generation