Day 8, Oct-10, 2024

Quick Summary of Machine Learning, Terms, Concepts and Ideas

→ Machine Learning - type of learning algorithm

Data + Output ⇒ Machine Learning Algorithm

→ Subset of AI

→ Used where Human Expertise is not enough or does not exist (navigating on Mars)

→ Medical field, Astronomy & Images, business.

→ Computational Biology, Space Exploration, Information extraction, Social Network are some ML applications.

Basic Paradigm

- Training Data: Example dataset for the observation & training
- Testing Data: unseen set of example for Prediction, use to evaluate the model's fitness

- Types of ML Algorithms:
  - Supervised – input and output given, allow & train model to predict the labels associated with a previously

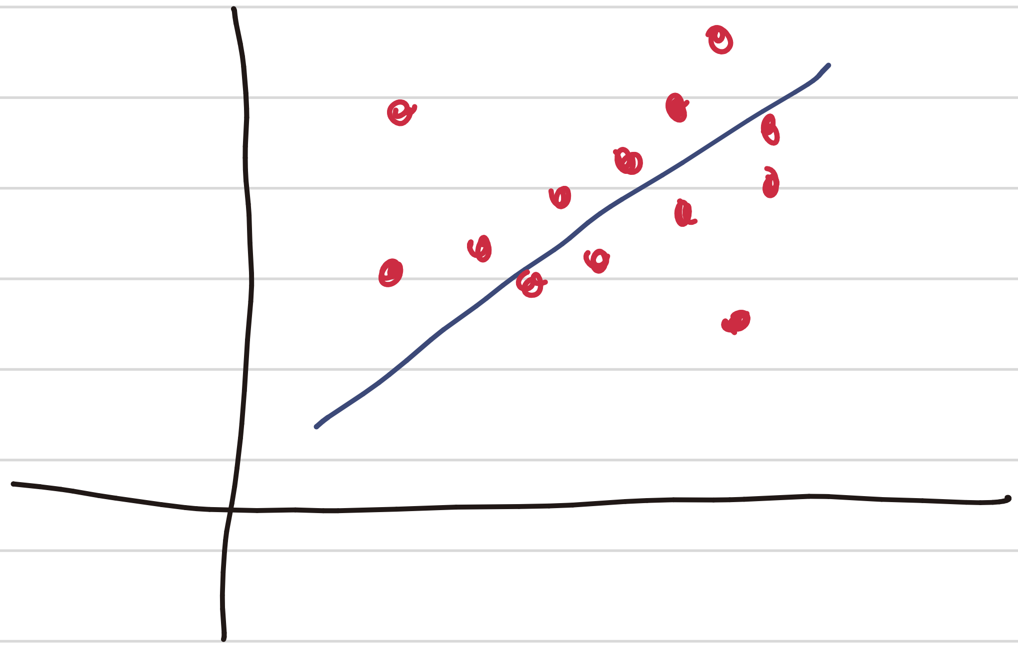
- Unsupervised - Only input given, o/p is not given to the model, model itself group or predict the class of given dataset or cluster the given set of features vector

- Reinforcement - This type of ML algorithm are actually resembles to the true AI definition because these AI agents are allow and left out in the environment for each mistakes punishment is given and for each correctness reward is given.

Three Canonical Learning Problems

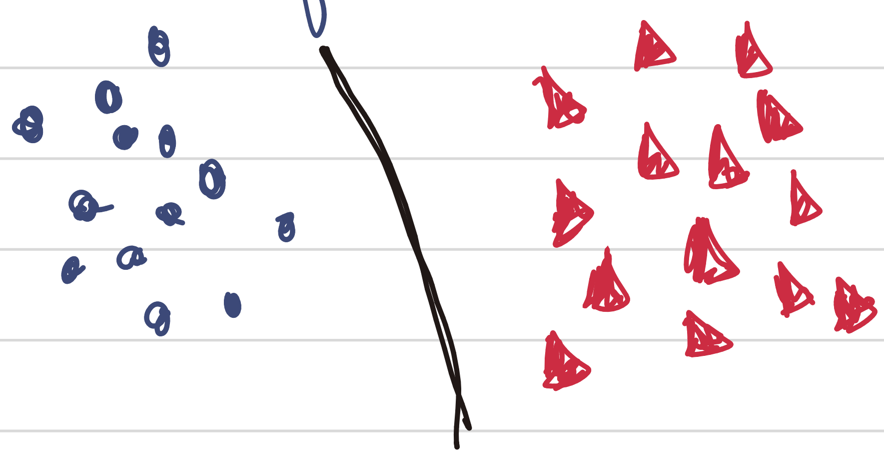
→ Regression → Supervised → Classification - Supervised

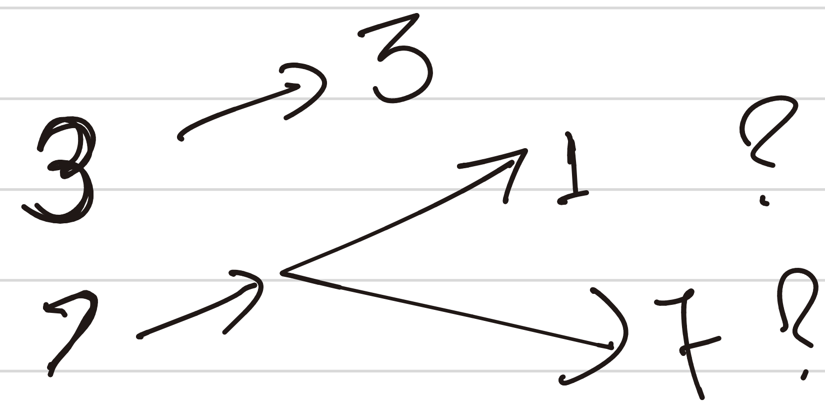
Regression - Supervised estimate parameters, eg: weight vs height



Classification - Supervised estimate class ; eg: handwritten digit

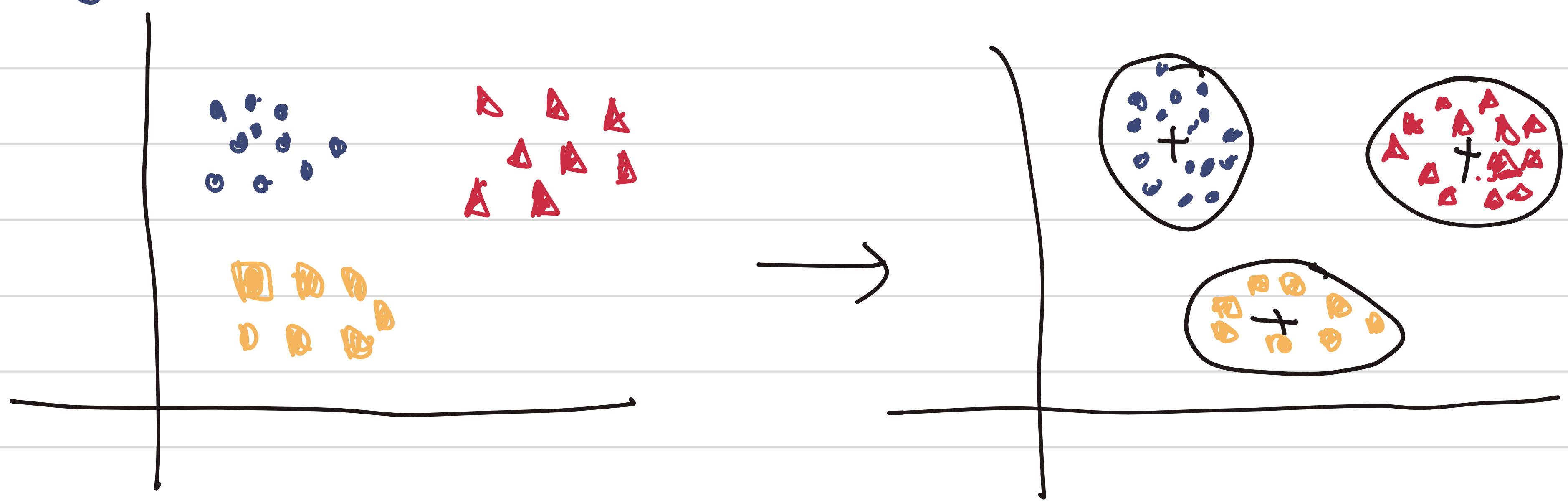
Classification.



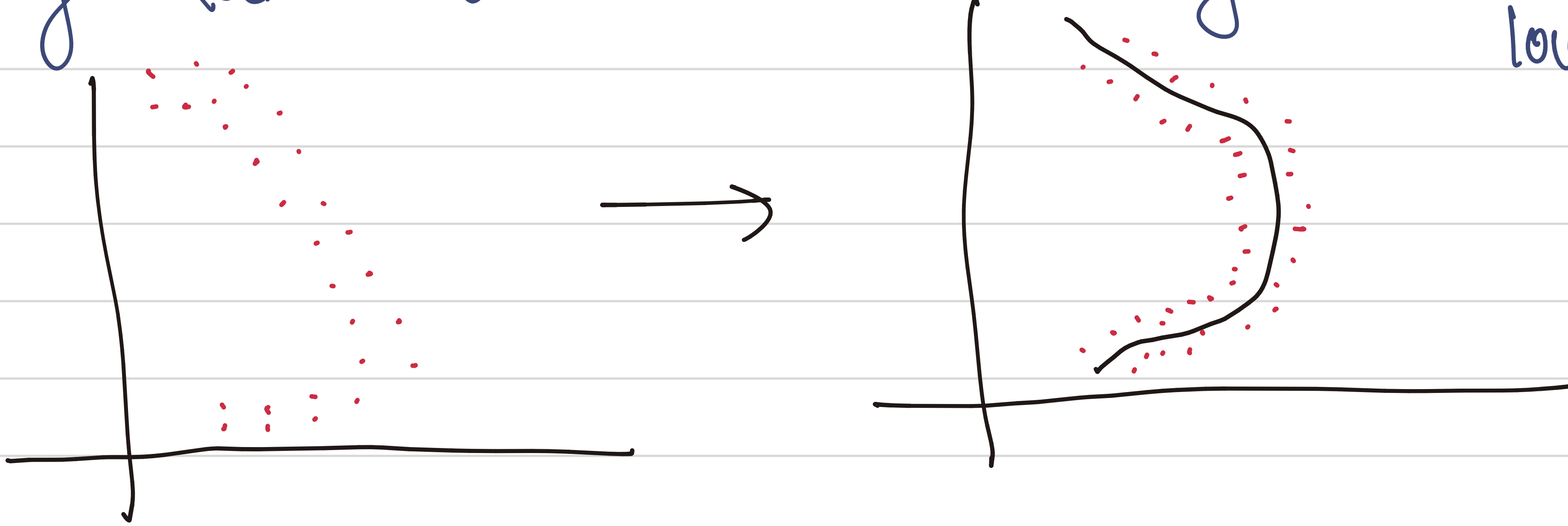


③ Unsupervised learning model the data

- Clustering



- dimensionality Reduction → reduce the high-dimension into lower-dimension



Three Components of ML

1) Representation

> Regression model

> Neural network

> Decision Tree

> Support Vector Machine

2) Evaluation

> Accuracy

> Squared Error

> Precision and Recall

> posterior Probability

> Entropy

3) optimization

> Combinatorial optimization eg. Greedy Search

> Convex optimization eg Gradient Descent

> Constrained optimization eg: Linear Programming

Cost function - Average error of entire dataset

Loss function - Single data error

Supervised Learning : Motivation / Examples

1) predict weight from gender, height, age
2) predict Google stock price today from Google, Yahoo, Microsoft prices yesterday
3) predict each pixel intensity in Robots current camera image, from previous image and previous action
4) classifying the resume

Formal Motivation :

Given a set of measurements (x_i, y_i) for $i = 1, \ldots N$ subjects known as the training data, we want to construct a

Prediction Rule for an arbitrary input x_0.

- X's are often called predictors or Covariates
- Y is the outcome variable; can be continuous, categorical or ordinal
- y is used to represent a realization of the random variable Y.

Glimpse into Linear Regression

Model continuous outcome as a linear function of predictors.

$$Y = \sum_{K=1}^{P} X^{K} \beta_{K} + \epsilon \Rightarrow X\beta + \epsilon$$

where ϵ is the noise

Y : A continuous outcome

$X \Rightarrow [X_1 \dots X_p]$: matrix of p predictors or features.

β : a vector of coefficients

ϵ : is sometimes called noise or error term.

How to predict Y at an arbitary X_0 ?

① Training: Define objective function to estimate β using training data.

② Prediction: Predict outcome of X_0 as

$$E(Y \mid x_0) = \hat{Y} \mid x_0$$

$$\Rightarrow \hat{\beta} x_0$$

Likelihood Approach.

$$\hat{\beta} \Rightarrow \operatorname{argmax}_\beta [\text{likelihood}(\beta)]$$

$$\Rightarrow \operatorname{argmin}_\beta [-\text{likelihood}(\beta)]$$

In Linear Regression, fitting the best line, finding the best parameters (slope, intercept) that maximize the likelihood of data under this model.

Let for each i, $\epsilon_i \sim \text{Normal}(0, \sigma^2)$

$$\Rightarrow Y_i \mid X_i, \beta \sim \text{Normal}(X_i \beta, \sigma^2)$$

$$\text{likelihood}(\beta) \Rightarrow P(Y \mid X, \beta)$$

$$\Rightarrow \prod_{i=1}^{n} P(Y_i \mid X_i, \beta)$$

MLE (Maximum-Likelihood Estimation) Linear Regression:

- $Y_1, Y_2 \ldots\ldots Y_n$ are i outcomes
- for each i, $\epsilon_i \sim \text{Normal}(0, \sigma^2)$

$$\Rightarrow Y_i \mid X_i, \beta \sim \text{Normal}(X_i\beta, \sigma^2)$$

$$\text{Likelihood}(\beta) = p(Y \mid X, \beta)$$

$$= \prod_{i=1}^{n} p(Y_i \mid X_i, \beta)$$

$$\text{Log-Likelihood}(\beta) \Rightarrow \log p(Y \mid X, \beta)$$

$$\Rightarrow \sum_{i=1}^{n} \log p(Y_i \mid X_i, \beta)$$

$$\bullet \log p(Y_i \mid X_i; \beta, \sigma^2) \Rightarrow \log \frac{1}{\sqrt{2\pi\sigma}} + \log \exp\left[-\frac{(Y_i - X_i\beta)^2}{2\sigma^2} \right]$$

Mean Square Error Loss

$$\hat{\beta} \Rightarrow \operatorname{argmin}_{\beta} \| Y - Y_{\beta} \|_2^2$$

$$= \operatorname{argmin}_{\beta} \sum_{i=1}^{n} (Y_i - X_i\beta)^2$$

$$\hat{\beta} \Rightarrow (X^T X)^{-1} X^T Y$$

In logistic Regression (M-likelihood E) is finding the best parameters to make Observed data more Probable. It maximize the (log- likelihood) and Correctly Classifying the outcomes.

- When $\epsilon \sim Normal(\mu, \sigma^2)$ minimizing squared error loss is equivalent to maximizing likelihood.

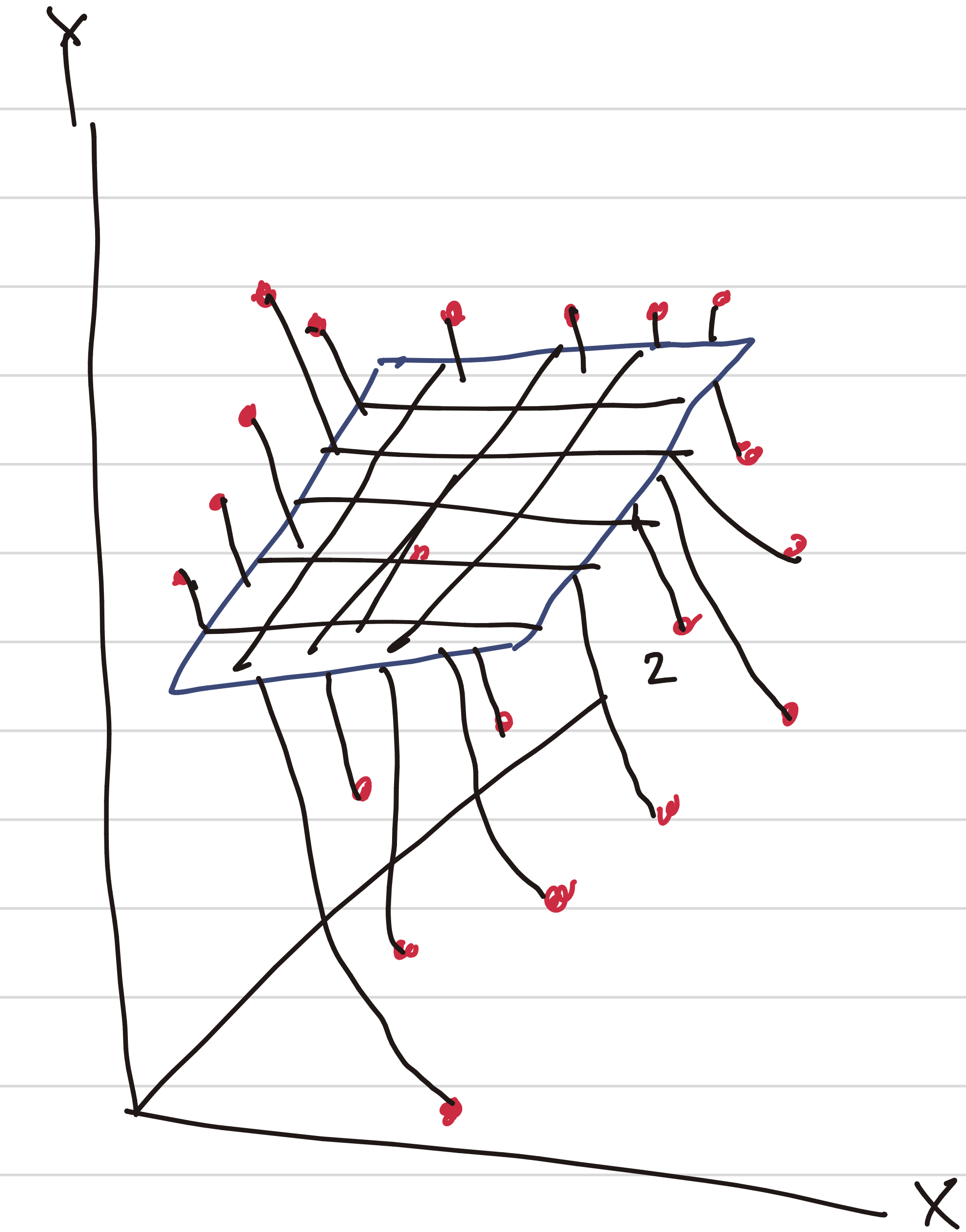
Pros:

→ Simplicity, Interpretability

→ Solution exists (in most cases)

→ Numerous software packages to fit the model

→ Favorable large-sample statistical properties



∴ Lines are supposed to be straight ☺

Cons of LR:

Real-world is filled with complexity!

→ Too restrictive: Assumes linear Relationship between features and outcomes

→ Big Data: Not suitable when feature space is large ($p >> n$)

Supervised Learning: Classification Models:

Outcome is discrete:

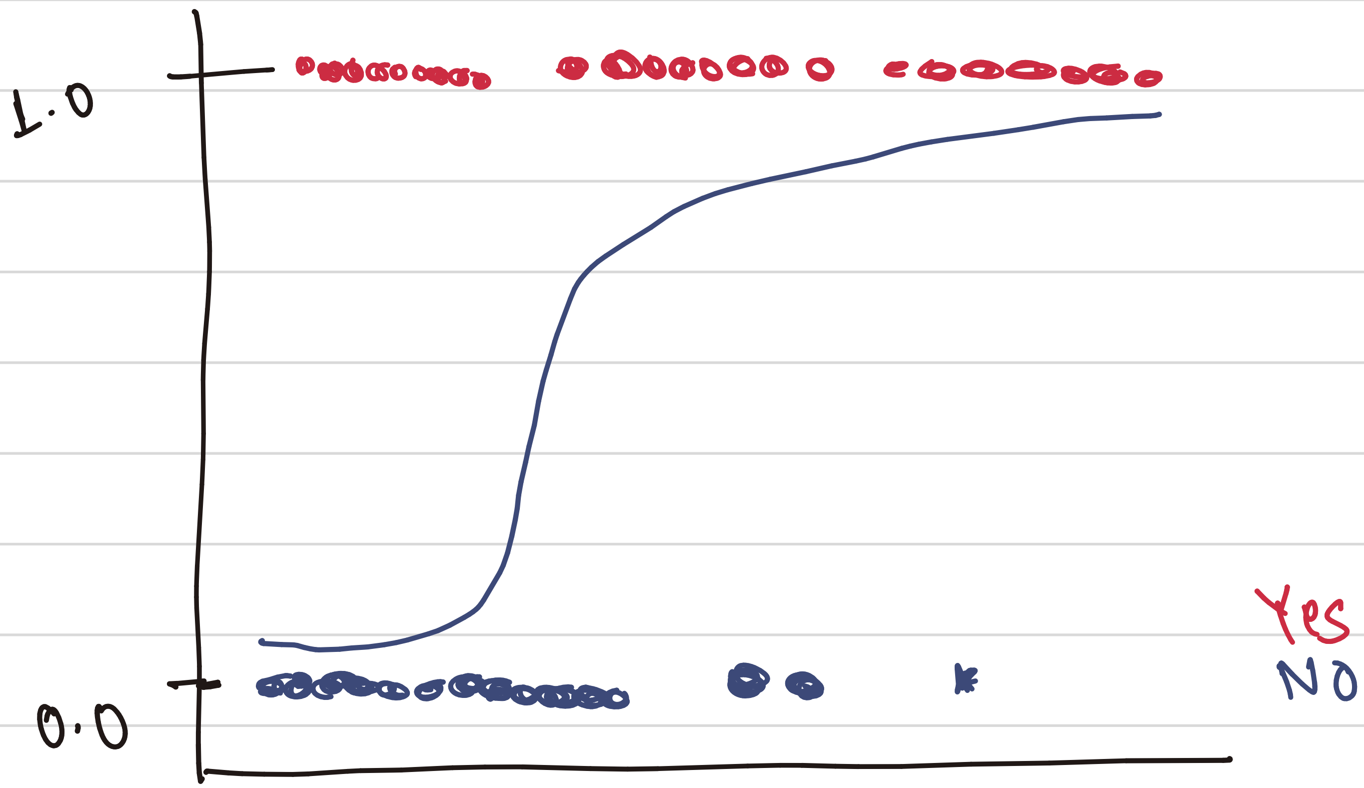
→ predict weather, rainy versus sunny, based on past weather

→ predict whether a patient will survive after a major surgery using vitals prior to the surgery

→ Classifying patients into high or low risk of Hypertension based on their age

- Goal: find a Role to correctly classify subjects into classes by learing from the existing data.

Figure:



$$\log \frac{Pr(Y=1)}{Pr(Y=0)} = X\beta$$

Prediction:

$$\hat{p} = \hat{P}r\left(Y=1\right) | X_0$$

$$\Rightarrow \frac{1}{1+Exp\left(-X_0\hat{\beta}\right)}$$

$$\hat{Y} \Rightarrow I\left(\hat{p} > 0.5\right)$$

Decision Boundary: Linear classifier

- A convenient property of this form is that it leads to a simple linear expression for classification.

- To classify any given Y we generally want to assign the value K that maximizes $Pr\left(Y=K | X\right)$, for $K = \{0, 1\}$.

- Classify $Y=1$ if

$$Pr(Y=1|X) > Pr(Y=0|X)$$

$$\Rightarrow 1 > Exp(\beta X) \equiv 0 > \beta X$$

More classification Rule is linear in x.

Supervised Learning: Regression Models

$$E(Y|X) = f(X)$$

- Many other functional forms on f is possible
- f is often refered to as a link function in statistical learning

- Parametric versus non-parametric models: Restrict f to rely on few parameters versus using f in a bigger class of family with no restrictions.
- How to decide which f to use?

General Estimation: Minimization of Expected Risk or loss function

$$\hat{f} = \arg\min_{f} E(\text{prediction loss}(f))$$

parametric Models of f (use when you have prior knowledge, data points (few))

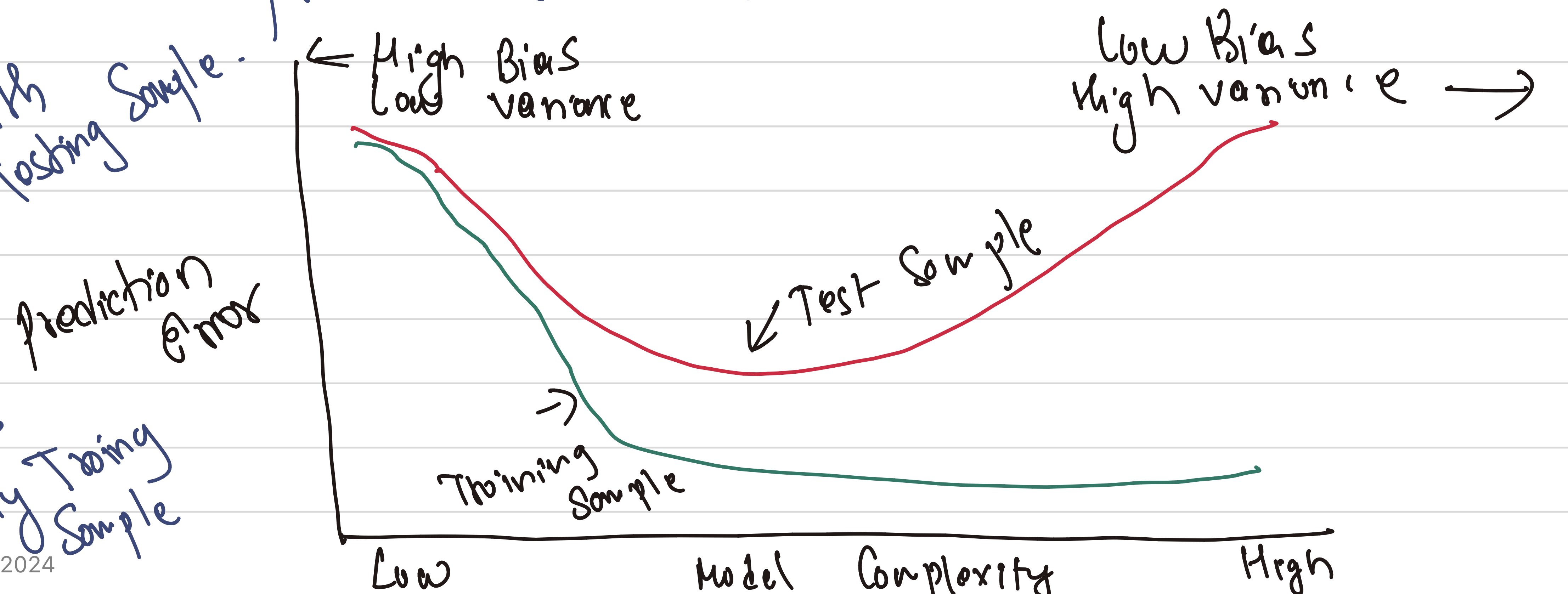
Non-Parametric Models of f (use when large, complex datasets, no prior knowledge)

General Estimation: Maximization of Expected Risk or loss function.

- optimization methods are used to do the maximization
  - Challenging if the loss function is not convex.
- Desired to build a most parsimonious model that maximizes the Prediction error.

Bias: Properly does with Training & testing Sample.

Variance: Properly does with only Training Sample



Performance Metrics

Truth Table (Confusion Matrix)

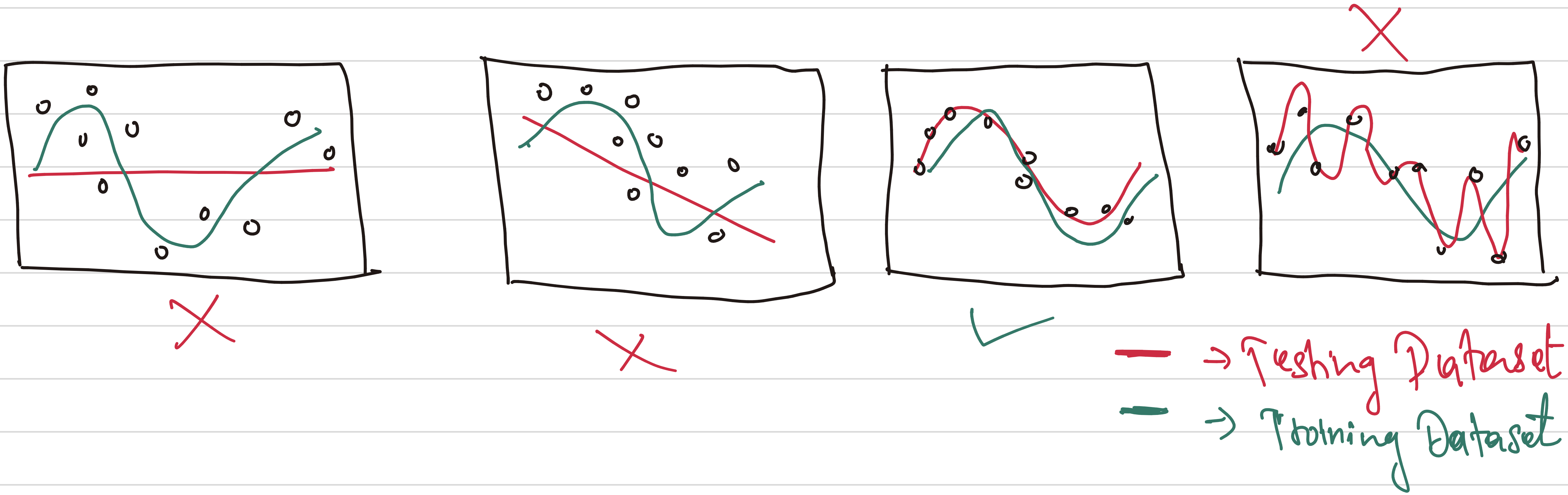
	P	N
P	True (TP)	False (FN)
N	False (FP)	True (TN)

- Accuracy $\Rightarrow \dfrac{TP + TN}{P + N}$

- Sensitivity $\Rightarrow \dfrac{TP}{\text{Actual Positives}}$

- Specificity $\Rightarrow \dfrac{TN}{\text{Actual Negatives}}$

Some fits to the data: which is the best?



From Textbook.

① Knowledge Representation in AI

> Knowledge and Importance of Knowledge
> Issues in Knowledge Representation

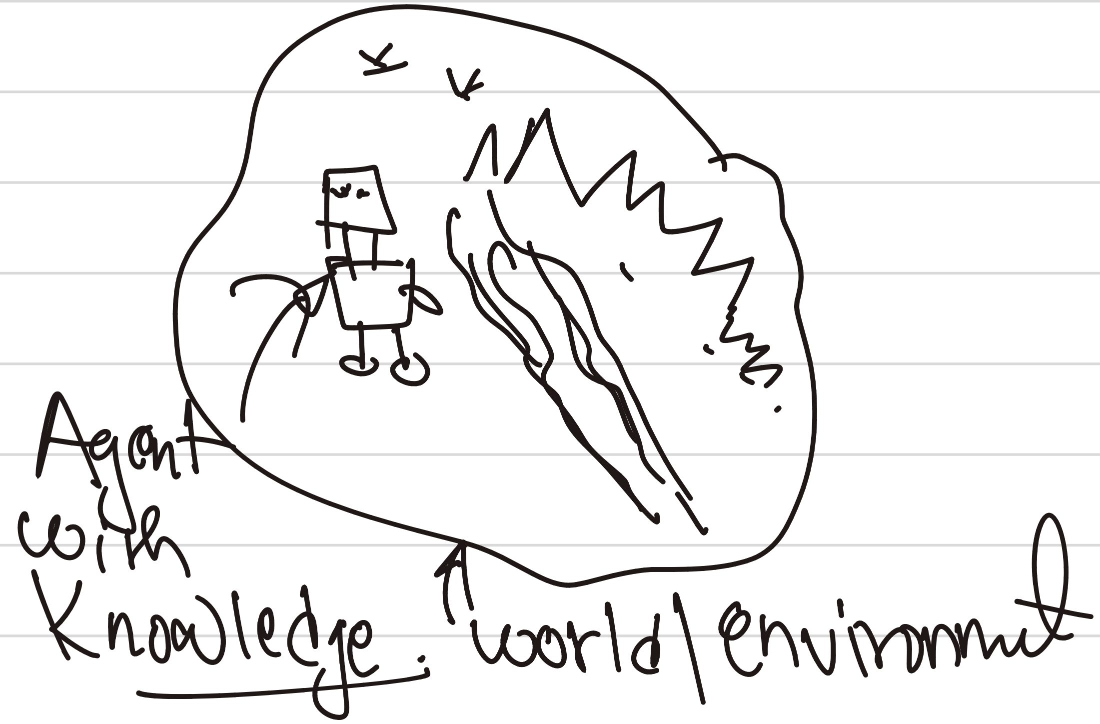
> Knowledge Representation Systems in AI

What is Knowledge ?

→ Only information is not knowledge but insights + Information + skill + experience ⇒ Knowledge that help an intelligent system to make decision, solve problems understand the world.

→ Without proper knowledge an AI agent cannot function, take proper decision, interact and mimic like human. Knowledge helps to derive, reason, deduce and support decision making.

Issues in Knowledge Representation



① Complexity of Representation:
How to efficiently store and manipulate Knowledge so that it can be efficiently Computable and Manageable.

② Ambiguity: Unclear Knowledge, Ambigious and uncertainties.

③ Inference & Reasoning: making sure correct inference & logical reasoning

Scaling: At what scale? volume of Knowledge as data representation without degarding the Performance?

Context Sensitive: Knowledge must be adoptable, context & Semantical to represent and store

Knowledge Representation System:

1) Rule-Based System
2) Semantic-Networks
3) Frames
4) Ontologies
5) Logical Representation

Summary

1. Learning Algorithms + Statistical methods + Data + Output ⇒ Machine Learning

2. Supervised, Unsupervised Learning are classification, Regression and clustering, K-map, Dimensionality Reduction.

3. Linear Regression and Logistic Regression Mathematical notation and figure with the loss function.
Where Linear Regression → Regression (Continuous Variable Predict)
Whereas Logistic Reg. → outcome (0 or 1 of Probability)

④ Components of Machine Learning Testing, training, Loss function, Parametric & Non-Parametric Model, Gradient Descent.

⑤ Low Bias & High Variance and High Bias & Low Varian.

High B + low Variance ⇒ Model overfit

High V + low B ⇒ Model adaptable

⑥ AI Agent must have Knowledge & represent them within itself and to the world to make decision, solve, understand, navigate the real or environment or virtual world with facing Challenges.

⑦ The Knowledge can be represented via Semantics, Ontologies, Rule-base, etc.