Module 1: Machine Learning – An overview

Lecture 2 - Part 1

Theme: Hypothesis Space

Credit: ARTQU

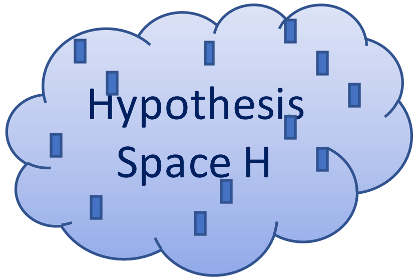
Photo by Pawel Nolbert on Unsplash

Topic 1: Supervised Learning –
A Search through Hypothesis Space

- Supervised learning constructs a Hypothesis Function

Hypothesis
h: X Y

- Hypothesis Space **H**: Set of all possible hypotheses



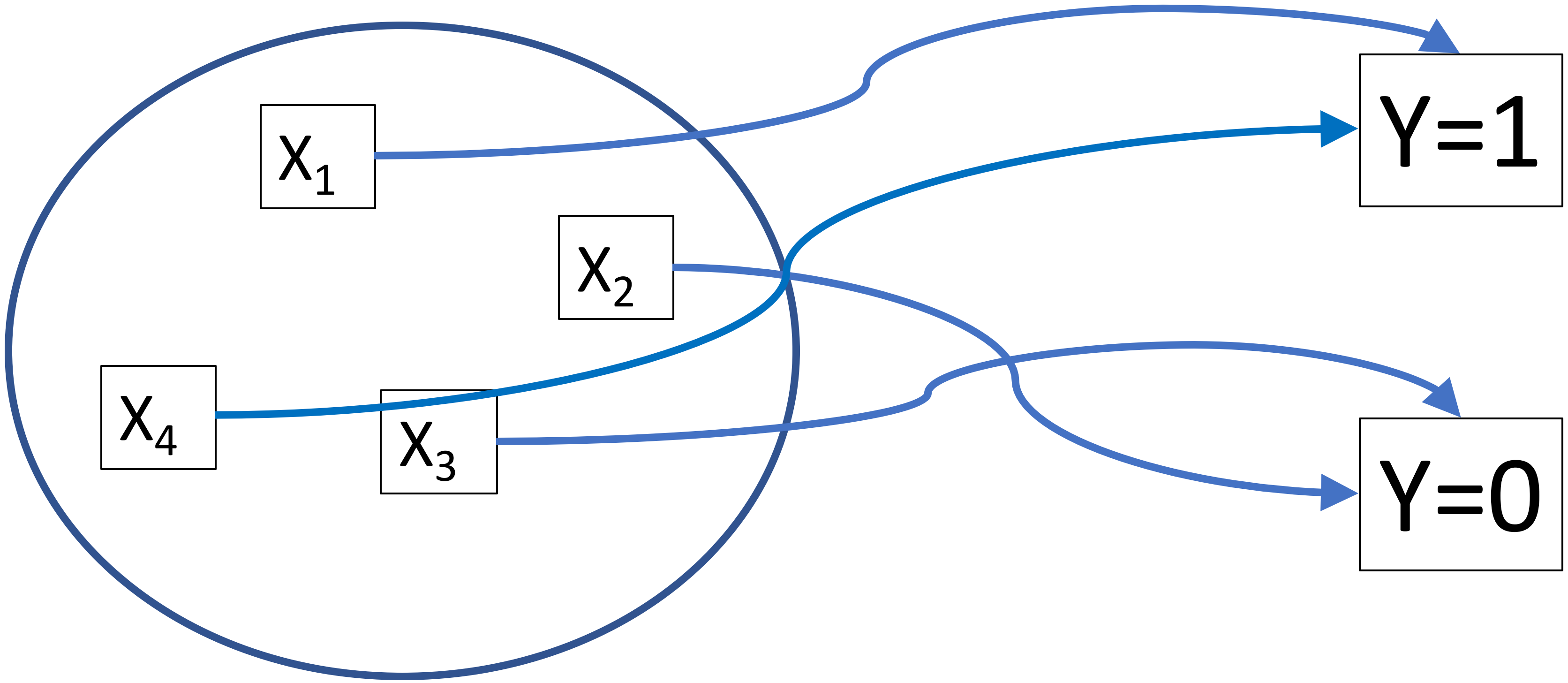
Boolean expression: Maximum Size of the hypothesis space

- |H| = Number of valid label assignments:

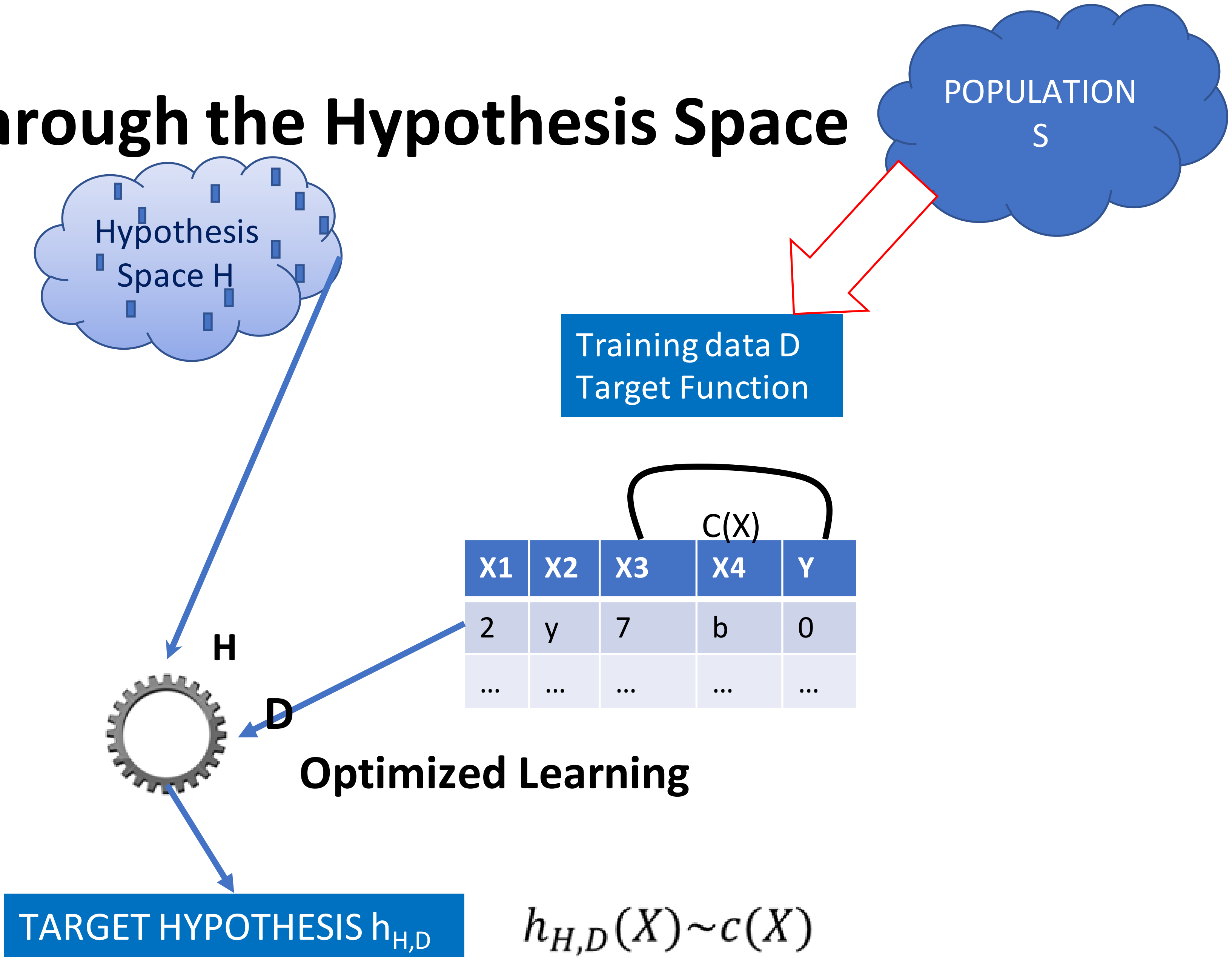
  $$\text{No of values Y can take}^{\text{Number of combinations of feature values}}$$

- In reality, it depends on ***how the hypothesis is expressed and restricted.***

Training Data has **Target function: c(X)→ Y**

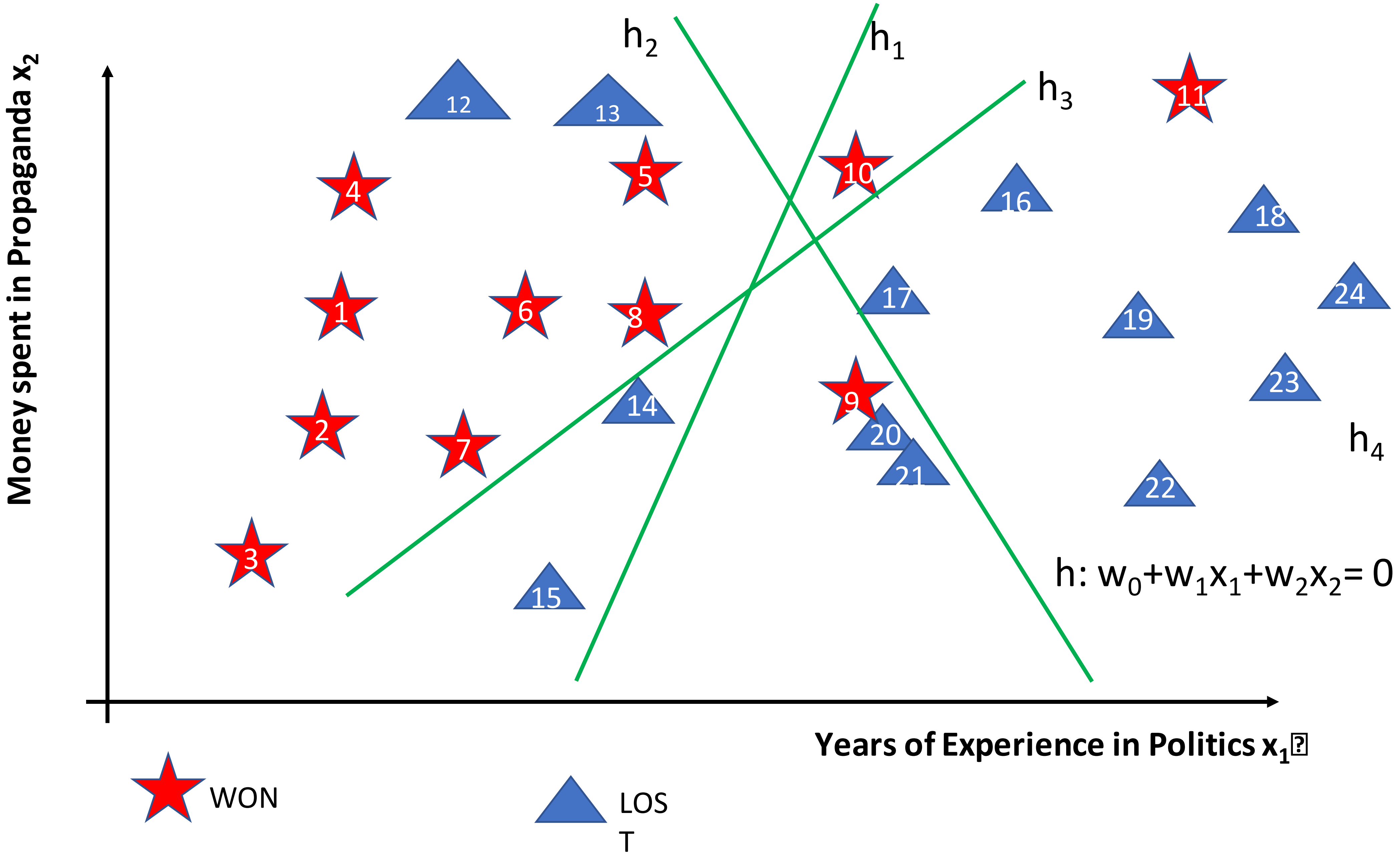


Searching through the Hypothesis Space



X1	X2	X3	X4	Y
2	y	7	b	0
…	…	…	…	…

$$h_{H,D}(X) \sim c(X)$$



LINEAR SEPARATOR HYPOTHESES FOR ELECTION RESULTS

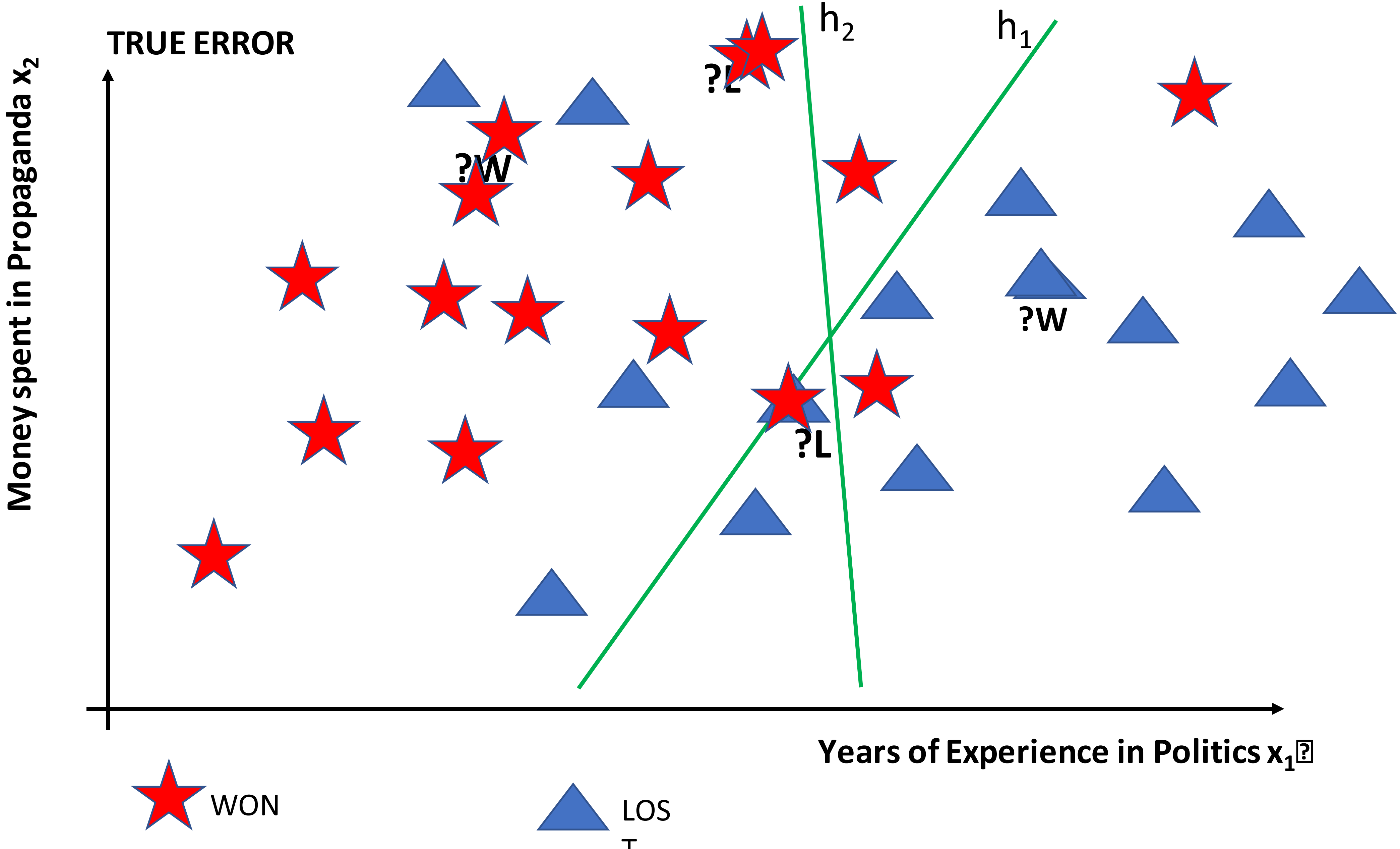
Training Error

- $err_{train}(h) = Pr_{x \in D}[h(x) \neq c(x)] = \frac{count\ of\ mismatches}{|D|}$
- $err_{train}(h_1) = \frac{7}{24} = 29\%$ $err_{train}(h_2) = \frac{66}{24} = 25\%$

$$err_{train}(h_3) = \frac{4}{24} = 17\%$$

- $h_{target} = \underset{h \in H}{\operatorname{argmin}}\{err_{train}(h)\} = h_3$

TRUE ERROR



LINEAR SEPARATOR HYPOTHESES FOR ELECTION RESULTS

Learning Errors

- Training error:

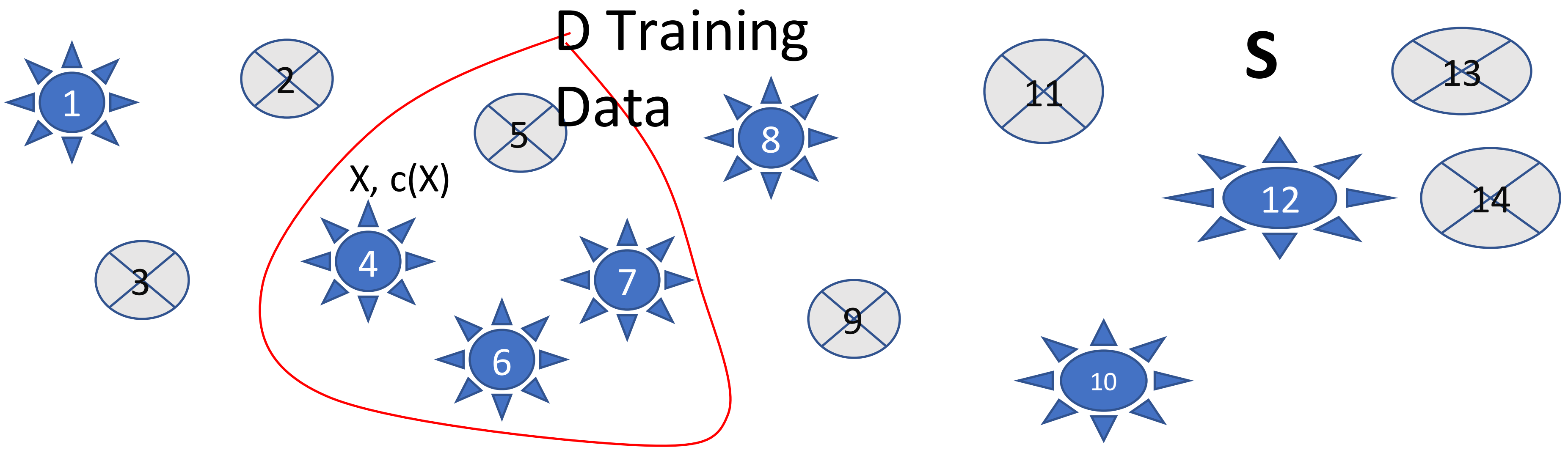
$$err_{train}(h) = Pr_{x \in D}[h(x) \neq c(x)] = \frac{count\ of\ mismatches}{|D|}$$

- True error:

$$err_{true}(h) = Pr_{x \in S}[h(x) \neq c(x)]$$

True and Training Errors

- True error is mistakes committed when any random case from S is presented to a candidate hypothesis $h_{H,D}$:



Topic 2:

Learning –

An optimization Problem

- Learning $\Rightarrow$ Continuous search through the hypothesis space for finding the *best possible target hypothesis*
- This hypothesis must minimize both Training error as well as True error to generalize
- However, 100% accuracy is impractical Near best solution
- The learning process is thus seen as an Optimization problem.

2.1 How Hypotheses is expressed?

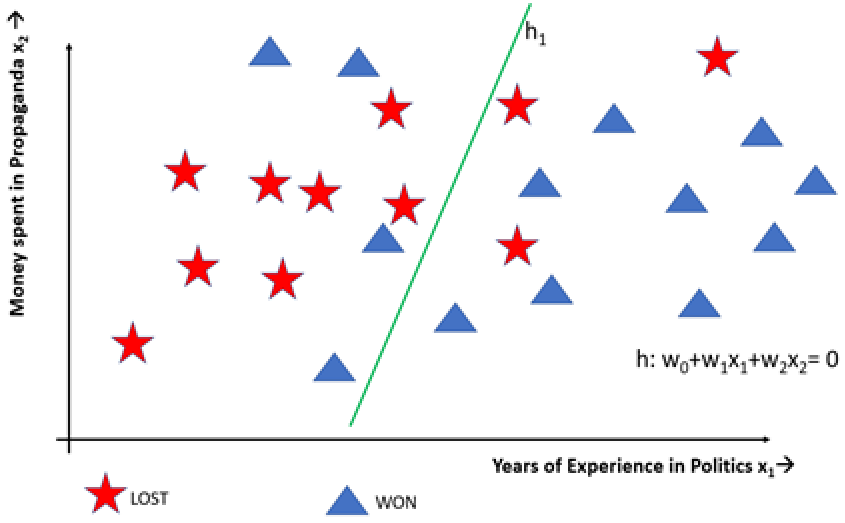
- Boolean Expression $\mathbf{y=(x_1 \wedge x_2) \vee x_3}$

  Finite Hypothesis Space

- Linear Separator Line /plane for Classification

  $\boldsymbol{h: w_0 + w_1 x_1 + w_2 x_2 .....}$

  Infinite Hypothesis Space



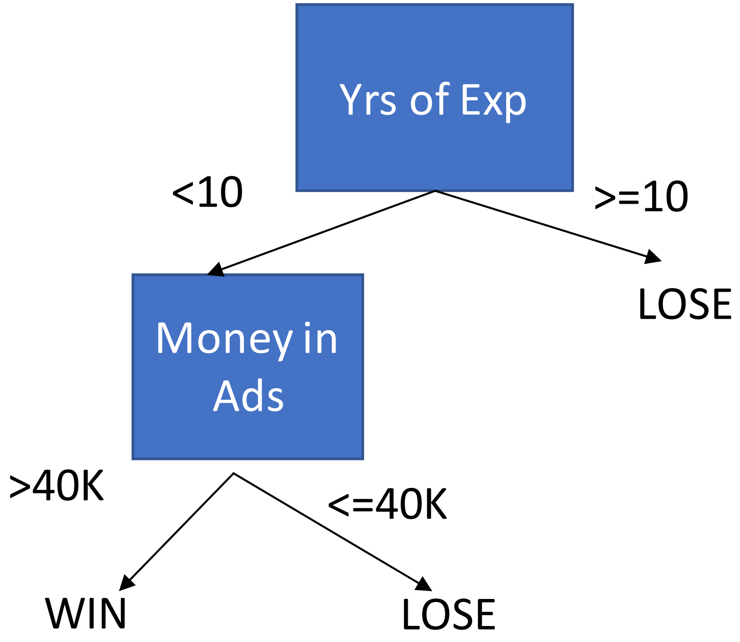
HYPOTHESES SPACE FOR ELECTION RESULTS

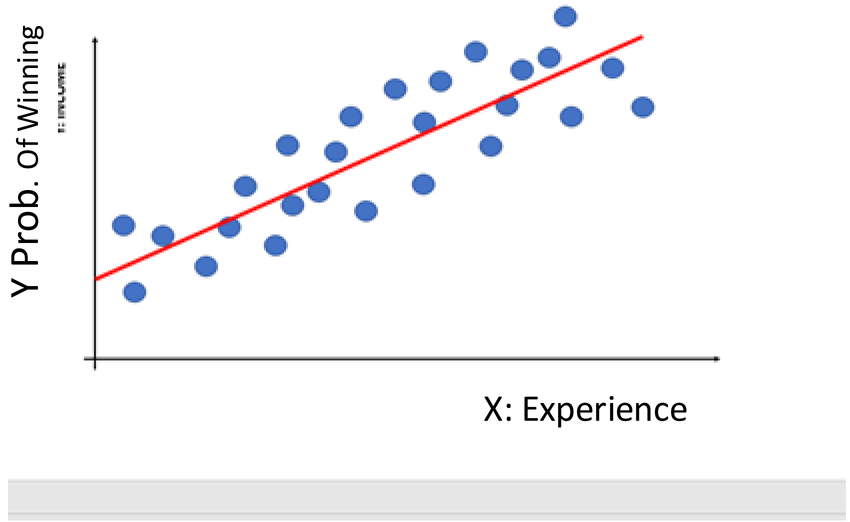
Hypotheses expression

- A Line / Plane that fits continuous data for Regression

  $y = w_0 + w_1 x$

- A Decision Tree:





Hypothesis expression…

- A set of logical rules:

  If Years of Experience >25 years AND Money spent in Ads < 20K THEN LOSE

- Probabilities:

  If probability that experience > 25 years is 30% and Probability of spending 20K in Propaganda >70%, then Probability of winning is 60%.

A.2 Objective Function

- **Objective Function (OF) of learning:** a “goodness measure” that evaluates how well the learning search process is shaping up.
- OF is either maximized or minimized. Example Regression: *Minimize Mean of Summation of Squared Errors MSE*:

$$Minimize_{data}\{\frac{1}{N}\sum_{i=1}^{N}(predicted\ response_i - actual\ response_i)^2\}$$

$$\text{Classification: Maximize}\{Accuracy = \frac{TP+TN}{TP+TN+FP+FN}\}$$

A.3 Constraints

- Constraints guide the process of hypothesis space search by limiting certain aspects.
- Avoid over-training, Simplifies the hypothesis and reduces true error

Topic 3:

The Inductive Learning Hypothesis

The Inductive Learning Hypothesis

- ***Any hypothesis (model) h(x) that approximates the target function c(x) well over a sufficiently large number of training examples D, will also approximate the target function well over new unseen examples T.***
- This implies that both training error as well as true error is minimized
- The Inductive Learning Hypothesis embodies **Generalization**

Assumptions of Inductive Learning Hypothesis

Assumptions on training data:

- The training examples D represent the sample population S ⍰ ***same distribution***
- The input features express enough variety to ***allow discrimination***

Unbiased Learning is useless!

- An Unbiased learner searches the complete hypothesis space:
  - Has very Expressive hypothesis space
  - Complex learning models/ Target hypothesis
  - Performs rote learning
  - Low training error but high true error
  - Learner not able to generalize to new examples
- We need to apply some biased learning to enable generalization

The Inductive Bias of Target Hypothesis

- **There is Inductive bias:** A set of assumptions about the *Target Hypothesis* that give an idea about its generalization ability.
- Each learning algorithm has its own inductive bias

Learning Points

- **Hypothesis is a function that maps a set of independent attributes X to the dependent response h: X → y. A collection of *possible hypotheses* forms the hypothesis space**
- **Different learning algorithms express the hypothesis function in different ways such as curves, graphs, and rules**
- **Learning scans hypothesis space H, given training data D, to find the best target hypothesis that matches the target function c(X) in D**
- **Training error is over training data D only, while True error is over entire feature space S with new unseen examples**

Learning Points

- Learning is an optimization process that minimizes/maximizes an Objective Function (OF) such as *MIN*{Mean Squared Error}/ *MAX*{Accuracy}, by tuning its parameters / hyperparameters
- The basis of learning is Inductive Learning Hypothesis (ILH) □ can induce the learnt hypothesis to new examples.
- ILH assumes that the training data has variety and truly represents sample space
- ILH also assumes inductive bias – a set of assumptions about the target hypothesis to show that it generalizes.

Self Assessment Link in Description Box

IF I CEASE SEARCHING, THEN WOE IS ME, I AM LOST!

VINCENT VAN GOGH