Theoretical Concepts and Important Terminologies in ML - II

Dr. Muhammad Wasim

Hypothesis Space (H)

- The learner has to apply some hypothesis, that introduces a search bias to reduce the size of the concept space
- This reduced concept space becomes the hypothesis space.

  For example, the most common bias is one that uses the AND relationship between the attributes.
- In other words, the hypothesis space uses the conjunctions (AND) of the attributes T and BP

  i.e. h = <T,BP>

Hypothesis Space

- H denotes the hypothesis space
- Here it is the conjunction of attributes T and BP

  If written in English, it would mean:

  H = < t, bp >: IF “temperature”= t AND “Blood Pressure” = bp

  Then

  H = 1 Otherwise H = 0
- In other words, the function gives a 1 output for all conjunctions of T and BP, e.g., H and H, H and L, H and N, etc.

Hypothesis Space

- h = < H,H > : < temp,bp >

BP				
H	0	0	1	
N	0	0	0	
L	0	0	0	
	L	N	H	**T**

Hypothesis Space

- h = < L,L >

BP				
H	0	0	0	
N	0	0	0	
L	1	0	0	
	L	N	H	**T**

- Notice that this is C2 that we discussed earlier in the concept space slide

0	0	0
0	0	0
1	0	0

Hypothesis Space

- H = < T,BP >
  - Where T and BP can take on five values
    - H, N, L (High, Normal, Low)
    - Also ? and ϕ
- ? means that for all values of the input H = 1 (don't care)
- ϕ means that there will be no value for which H will be 1

Hypothesis Space

- For example, h1 = <?, ?>: [For any value of T and BP, the person is sick
- The person is always sick

BP				
H	1	1	1	
N	1	1	1	
L	1	1	1	
	L	N	H	**T**

Hypothesis Space

- For example, h2 = <?, H>: [For any value of T AND for BP = High, the person is sick]
- Irrespective of temperature, if BP is High, the person is sick

BP				
H	1	1	1	
N	0	0	0	
L	0	0	0	
	L	N	H	**T**

Hypothesis Space

- For example, h3 = < ϕ,ϕ >: [For no value of T or BP, the person is sick]
- The person is never sick

BP				
H	0	0	0	
N	0	0	0	
L	0	0	0	
	L	N	H	**T**

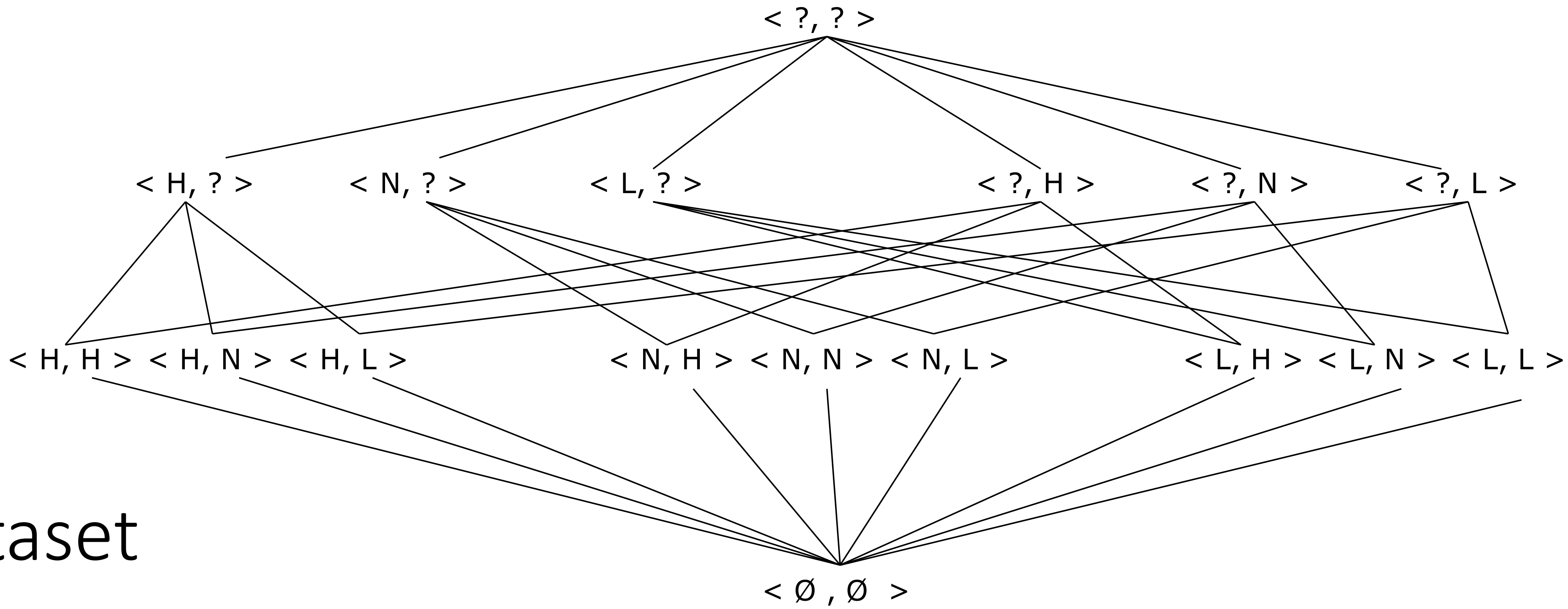
Hypothesis Space

- Having said all this, how does this still reduce the hypothesis space to 17?
- Well, its simple, now each attribute Temp and BP can take 5 values each: L, N, H, ? and ϕ
- So, there are 5x5=25 total number of possible hypothesis.
- Now, this is a tremendous reduction from 2^9 or 512 to 25
- This number can be reduced further
- There are redundancies within these 25 hypothesis caused by ϕ

Hypothesis Space

- These redundancies are caused by ϕ
- Whenever there is a ϕ in any of the inputs and we are considering conjunctions (min) the output will always be 0
- If there is ϕ in T or BP or both, we'll have the same hypothesis as the outcome is always, all zeros
- For a ?: we will either get a full column of 1's, or a full row of 1's in the concept matrix representation.
- For both ?: all 1's

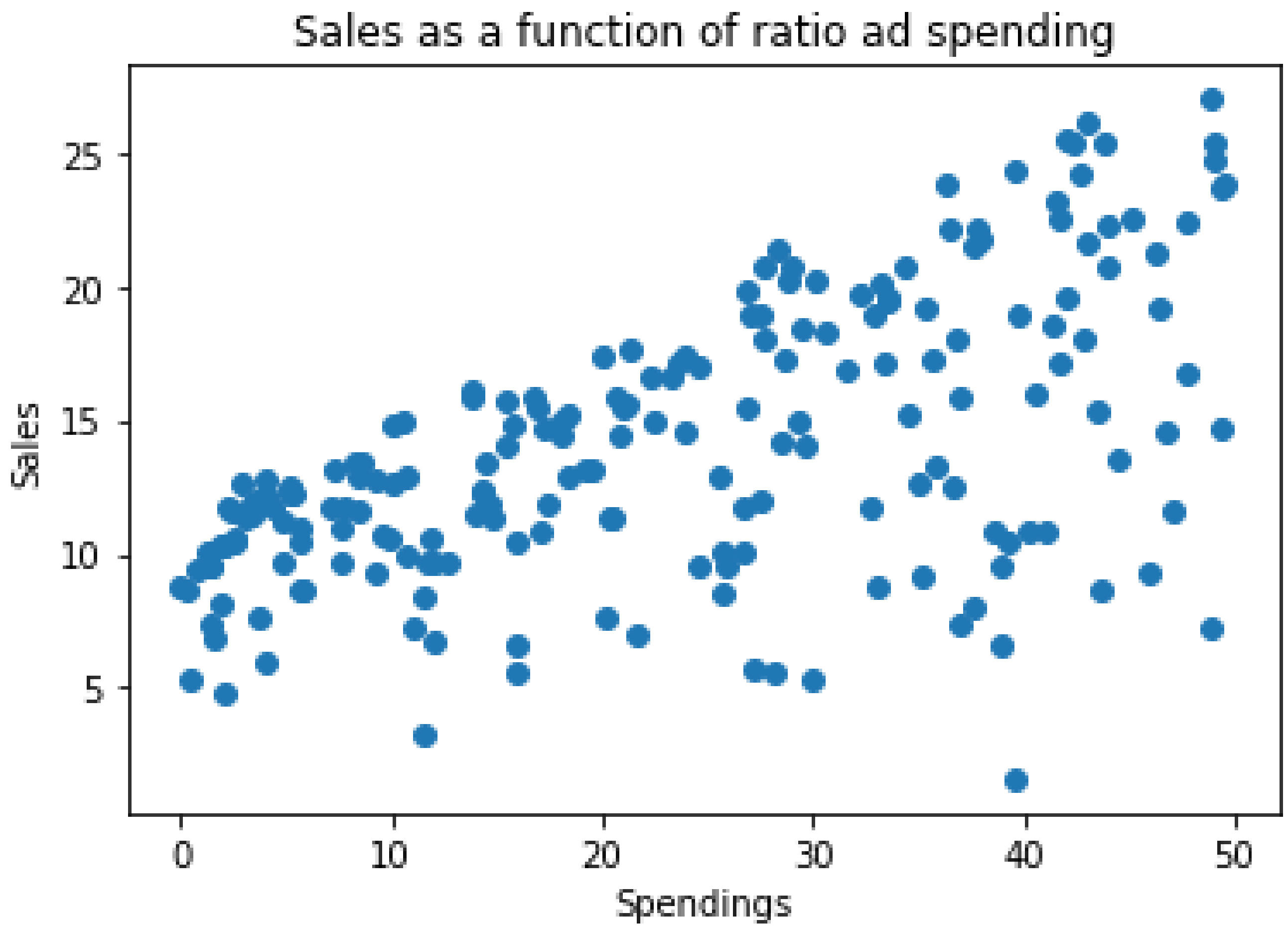
Hypothesis Space



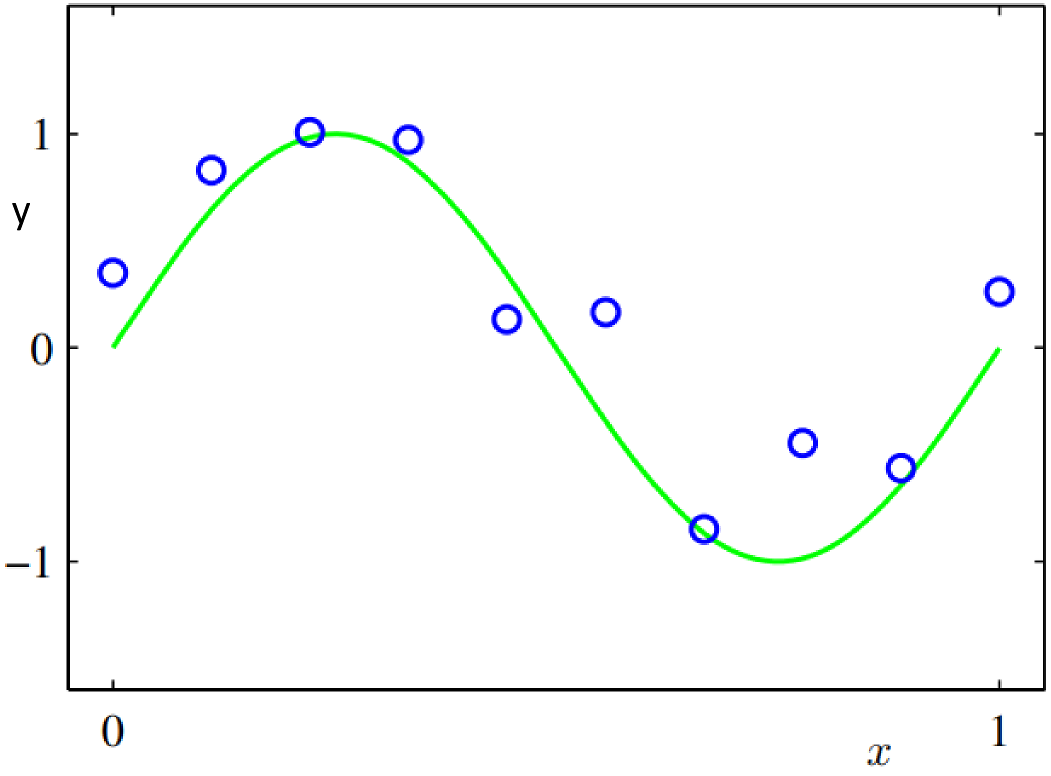
Dataset

Example	T	BP	SICK (SK)
1	H	H	1
2	L	L	0
3	N	H	1

Example of a Regression Problem (Line Fitting)



Assume we know the target function!

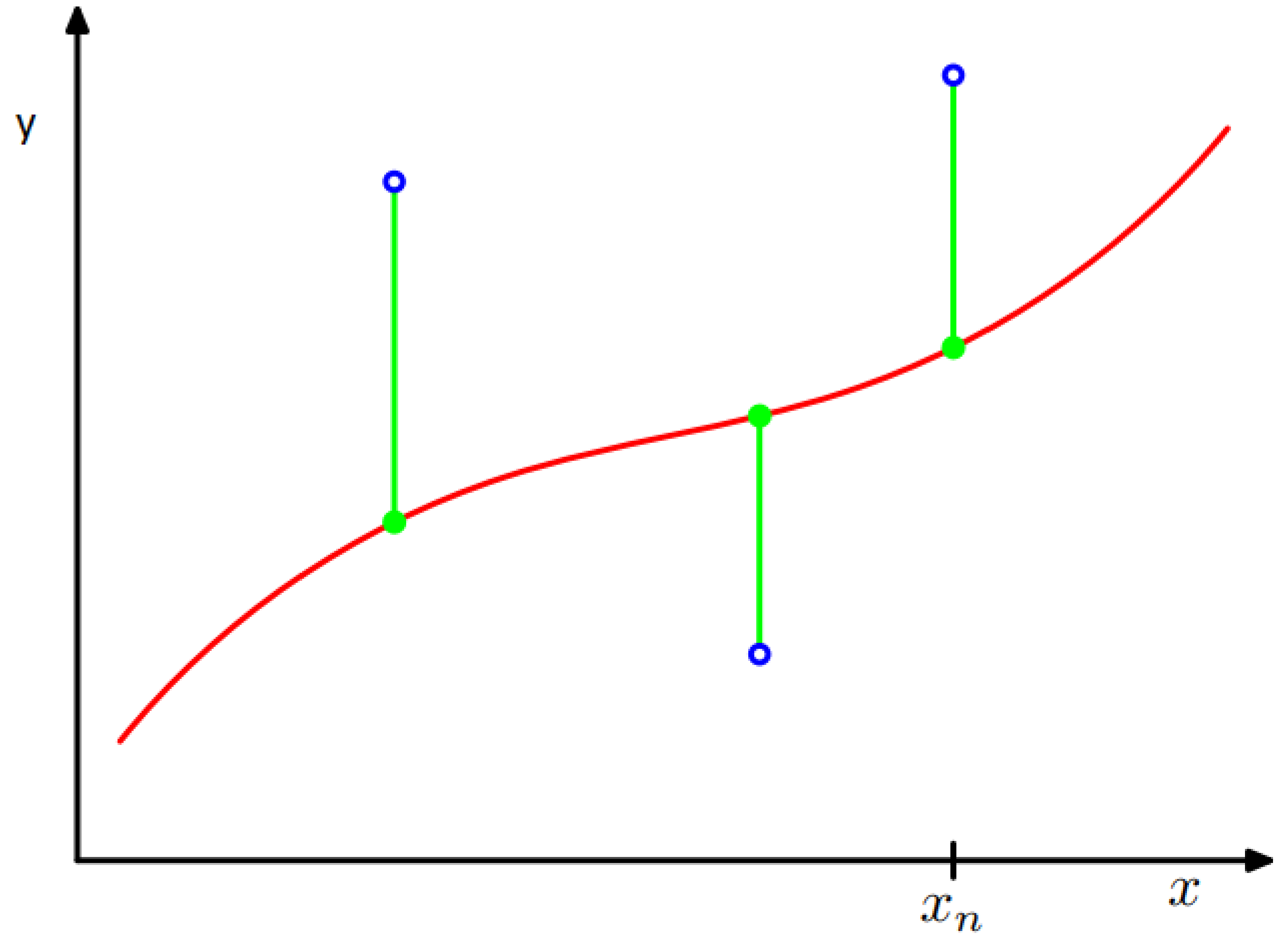


$y = \sin(2\pi x)$ ← Target Function

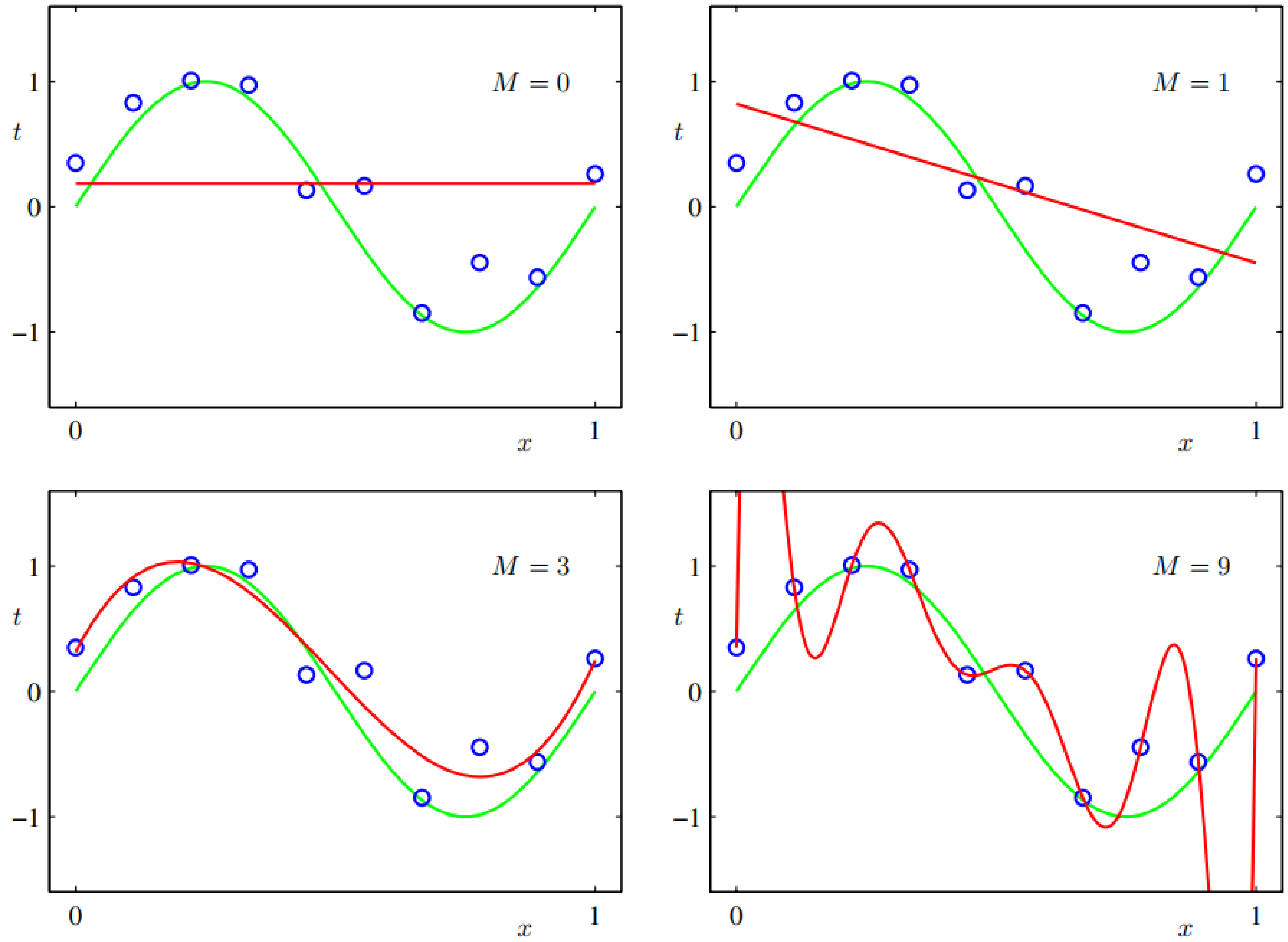
$\hat{y} = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$ ← Hypothesis

$E(w) = \frac{1}{2} \sum_{n=1}^{N} (y - \hat{y})^2$ ← Error Function / Loss

Visualizing the Error Function



Polynomials having various Orders of M



Note: y is also known as t (target)

- The objective is generalization not memorization.
- We always want our model to generalize well to unseen data. So, we split data in train/test split before training our model.
- Although, the higher order polynomial fits the data well, it may not perform well on unseen data.
- This phenomena is known as **overfitting**.