ECE 0402 - Pattern Recognition

LECTURE 1

Today (1/10): Introduction and Learning Challenge–feasibility of learning.

Wednesday(1/12): Why it is important to do inference from data using statistical perspective?

Example: The Netflix prize: Predict how a use will rate a movie

10% improvement = $1 million prize

This problem captures the essence of ML:

- a pattern exists
- cannot formulate it analytically
- have data

what we look for in a problem

Features

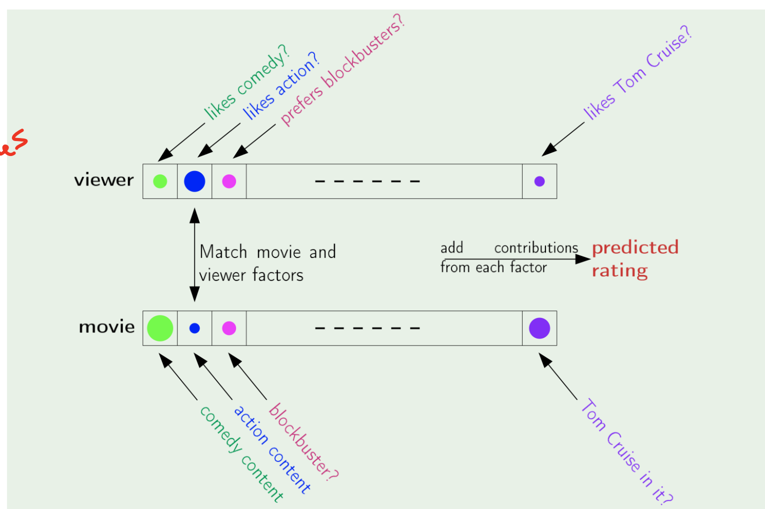


Figure 1: * A solution

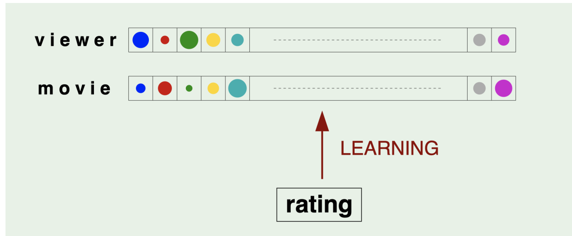
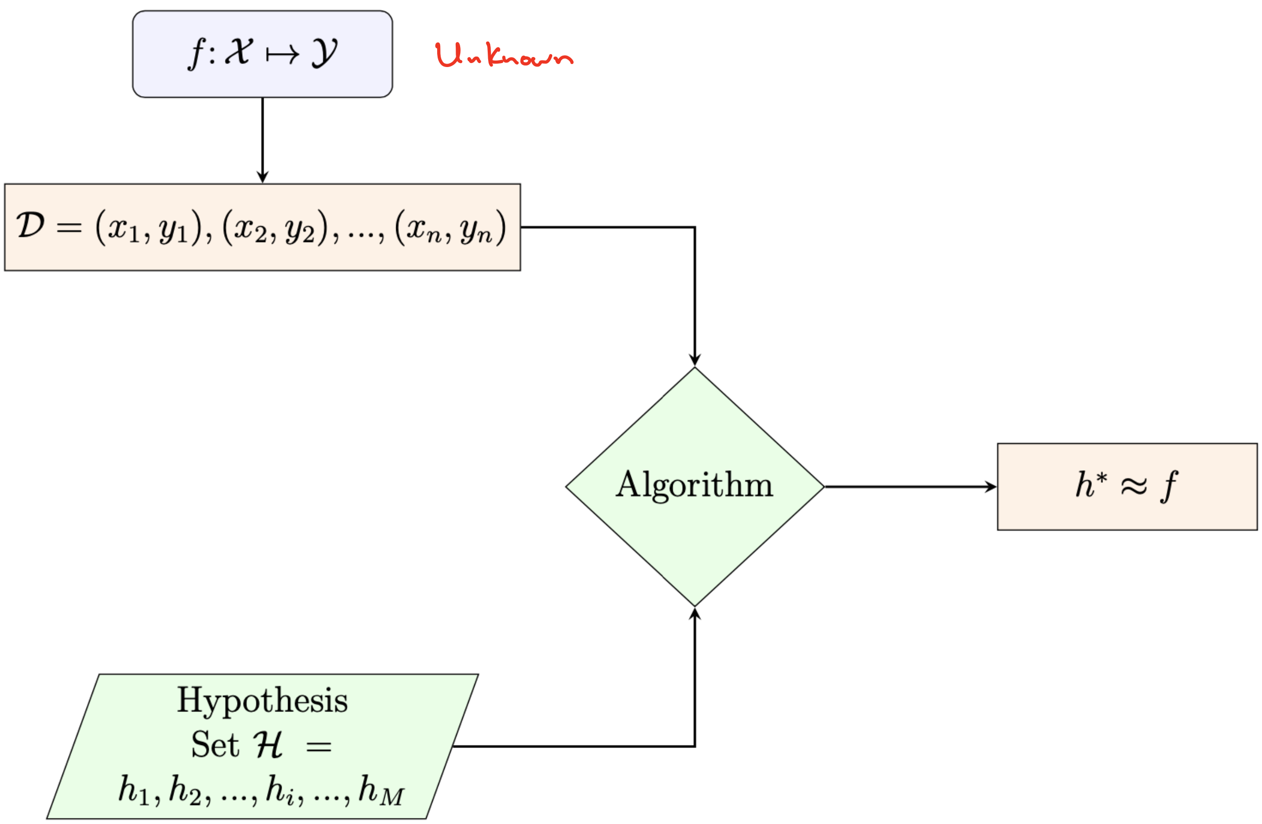


Figure 2: * Learning approach

Notation:

Supervised learning
Unsupervised learning
Reinforcement learning
Harder

$$\text{Input}: x$$
$$\text{Output}: y$$
$$\text{classification}: y \in \mathbb{Z}$$ integer
$$\text{regression}: y \in \mathbb{R}$$ real number
$$\text{Unknown}: f\colon \mathcal{X} \mapsto \mathcal{Y}$$ target function
$$\text{Data}: \mathcal{D} = (x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$$
$$\text{Hypothesis}: h^* : \mathcal{X} \mapsto \mathcal{Y}$$



Linear model

n: number of samples
d: number of features

$x \in \mathbb{R}^d$ or $\mathbb{Z}^d$

Example: Perceptron Model Let $x = (x_1, ..., x_d)$ be 'attributes of a customer'

$$\text{approve credit if } \sum_{i=1}^{d} w_i x_i > \text{threshold}$$

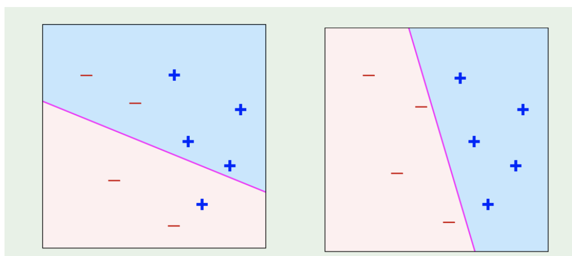
$$\text{deny } \sum_{i=1}^{d} w_i x_i < \text{threshold}$$

* Introduce (artificially) $x_0 = 1$, and combine

$w_0 = b = \text{threshold}$

$x_0 = 1$

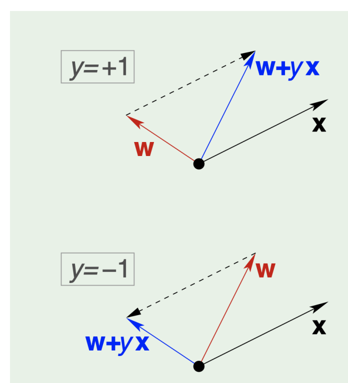
$$h(x) = sign\left(\sum_{i=0}^{d} w_i x_i\right)$$



'linearly separable' data

Learning algorithm:

- takes the training set $(x_1, y_1), ...(x_2, y_N)$
- picks a **misclassified point**, i.e., $sign(w^T x_n) \neq y_n$
- updates the weight vector: $w \leftarrow w + y_n x_n$



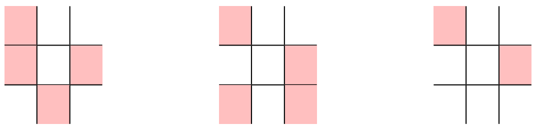
Interesting fact: it works as long as the data is linearly separable.

Types of Learning

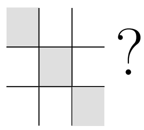
Proposition: "using a set of observations to uncover an underlying process"
Big promise - large territory, many types...An instance – supervised learning

Learning Challenge

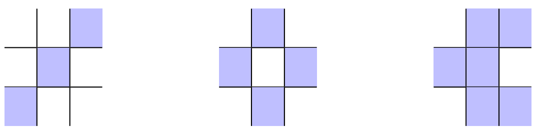
Example 1*:



All have shade in upper left square



→ Fits both criteria



All are symmetric

Example 2*:

x	y	h^*	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8
0 0 0	0	0	0	0	0	0	0	0	0	0
0 0 1	1	1	1	1	1	1	1	1	1	1
0 1 0	1	1	1	1	1	1	1	1	1	1
0 1 1	0	0	0	0	0	0	0	0	0	0
1 0 0	1	1	1	1	1	1	1	1	1	1
1 0 1		?	0	0	0	0	1	1	1	1
1 1 0		?	0	0	1	1	0	0	1	1
1 1 1		?	0	1	0	1	0	1	0	1

How well do you generalize to unseen examples

If we remain true to the notion of unknown target function, we cannot exclude any $f_1, ..., f_8$ from being the true f. It is easy to show that any 3 bits that replace the question marks are as good as any other 3 bits.
Some thoughts:

- Learning vs. Memorizing
  - The purpose of learning f is to able to predict the value of f on points that we haven't seen before.
  - The quality of the learning is going to be determined by how close our prediction is to the true value.

I References

1. Chapter 1, Elements of Statistical Learning (by Hastie, Tibshirani, and Friedman)

2. Chapter 1, Learning From Data, by Abu-Mostafa, Magdon-Ismail, Lin. * Example 1&2, and this lecture's figures are taken from this textbook.

https://web.stanford.edu/~hastie/ElemStatLearn/printings/ESLII_print12_toc.pdf