ECE 0402 - Pattern Recognition

LECTURE 1

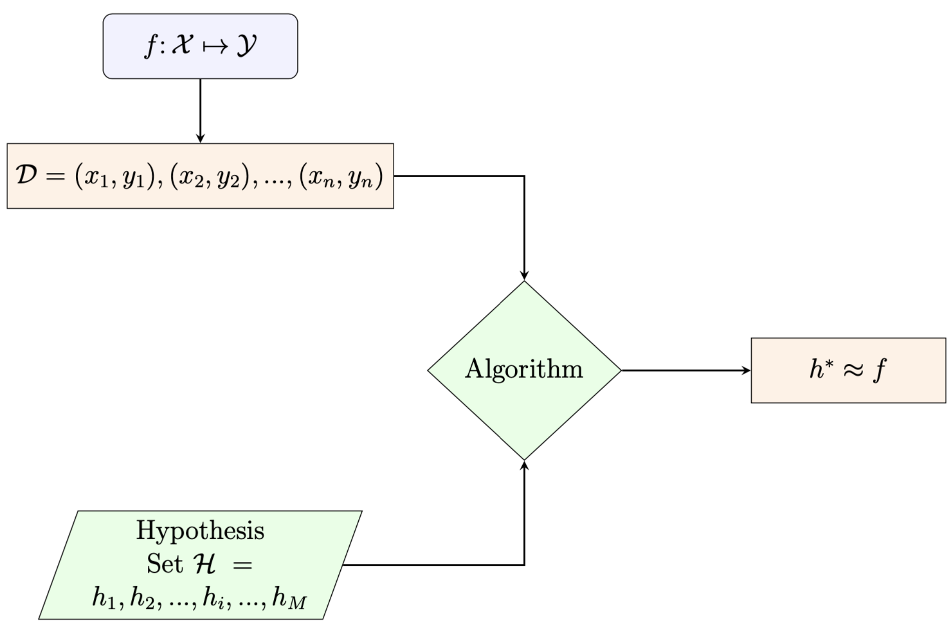
Today: Introduction and Learning Challenge. Why it is important to do inference from data using statistical perspective?

I Problem and Proposition

Proposition: "from seen uncover the unseen"
Problem: Is it possible?

Big promise - large territory.
An instance – supervised setting:

$$classification: y \in \mathbb{Z}$$
$$regression: y \in \mathbb{R}$$

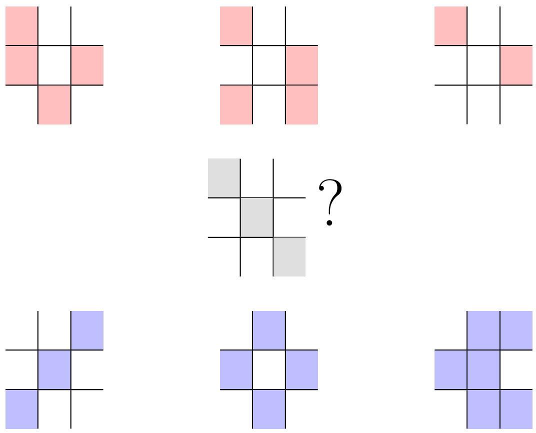


Goal: **generalize** the input-output relationship to unseen input x.

Learning Challenge

Learning : the acquisition of knowledge or skills through experience, study, or by being taught.

Example 1*:



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

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.

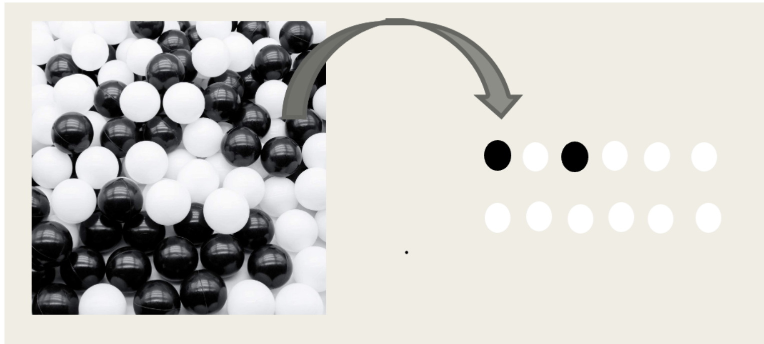


Figure 1: Possible? Probable?

- Can probability help the dilemma?
  - It is feasible to infer something outside $\mathcal{D}$, using only $\mathcal{D}$–but in a probabilistic way.

Let's consider the simplest case of picking a sample from a bag of white and black balls. Does the sample we pick tell anything about the white black ratio in the ball? Maybe.
Assume the ration of black and white balls in the bag such that if we pick a ball at random, the probability that will be black is μ. This is a good candidate for an unknown. Let's pick N balls out of this bag which has possible infinitely many of them. And let's count the black and white balls and say the ratio is ν. How much ν can tell us about μ? Well, it depends. There is this unlikely but possible scenario of getting all white (or all black) balls even though (maybe) our bag is mostly filled black balls. I hear you say, that is just not **Probable**! You can't deny the fact that regardless of the colors of the balls you picked, you still don't/can't know the color of any marble that you didn't pick.
Perhaps it is not that dramatic. We see polls all the time (in fact the biggest one is on the news). A random sample from a population to agree with the views of the population at large (I mean not in US politics).
ν is just a RV, μ is not. μ is more interesting because we don't know, but it is just a number–parameter is a better word. Parameter of the unknown distribution of the balls inside the bag. By LLN (or CLT), we know that if sample size is big, ν tends to be close (proximity) of μ. Actually Dr. Hoeffding studied more on this and proposed one of the most famous concentration inequality in the field of ML, i.e. Hoeffding's Inequality. It states that for any sample size N,

$$\mathbb{P}[\ \mid \nu - \mu \mid \leq 2e^{-2\epsilon^2 N}]$$

Here $\epsilon > 0$. In English, as our sample size grows, it becomes unlikely that ν will deviate from μ by more than a tolerance we picked, ϵ. The bad event probability drops exponentially with N. Intuitively no surprise here, almost sounds like a restatement of LLN. The nice thing here is the right side of the inequality does not depend on μ, the unknown. We are able the probability precisely. Only N effects the probability (and upper bound), not the size of the bin. The bin can be large or small, finite or infinite, and we still get the same bound for fixed N. The price you pay is obvious too, if you want a tight approximation of

μ (we can do that by picking ϵ small), you have to increase N so that the right side of the inequality small. Knowing that we can be within $\pm\epsilon$ of μ is a good promise compared to not knowing about it at all.
The key to the magic is that samples are drawn randomly from the bag. If the sample was not randomly selected, we would lose the benefit of probabilistic analysis, and be in the dark outside of sample again.
Well there is quite a **heap** in connecting intuition about a singe unknown μ to a whole blown function $f\colon \mathcal{X} \mapsto \mathcal{Y}$. Maybe illustration is a better way to explain here, switch to board.

II 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 and 2 are taken from this textbook.

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