Statistical Learning Method

- Statistical Learning is a set of tools for understanding data.

- These tools broadly come under two classes: supervised learning & unsupervised learning.

- Generally, supervised learning refers to predicting or estimating an output based on one or more inputs.

- Unsupervised learning, on the other hand, provides a relationship or finds a pattern within the given data without a supervised output.
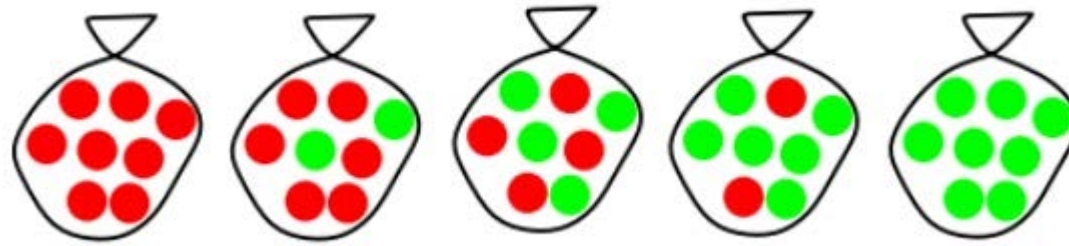
Cont...

- View learning as Bayesian updating of a probability distribution over the hypothesis space

- H is the hypothesis variable, values h1, h2, . . ., prior P(H) jth observation dj gives the outcome of random variable Dj training data d = d1, . . . , dN

- Given the data so far, each hypothesis has a posterior probability:

- $P(h_i|d) = \alpha P(d|h_i)P(h_i)$

- where P(d|hi) is called the likelihood

cont...

- Example

- Suppose there are five kinds of bags of candies: 10% are h1: 100% cherry candies

- 20% are h2: 75% cherry candies + 25% lime candies 40% are h3: 50% cherry candies + 50% lime candies 20% are h4: 25% cherry candies + 75% lime candies 10% are h5: 100% lime candies

Cont…

Cont…

1.The true hypothesis eventually dominates the Bayesian prediction given that the true hypothesis is in the prior

2.The Bayesian prediction is optimal, whether the data set be small or large