Bayes Classifier

Supplement

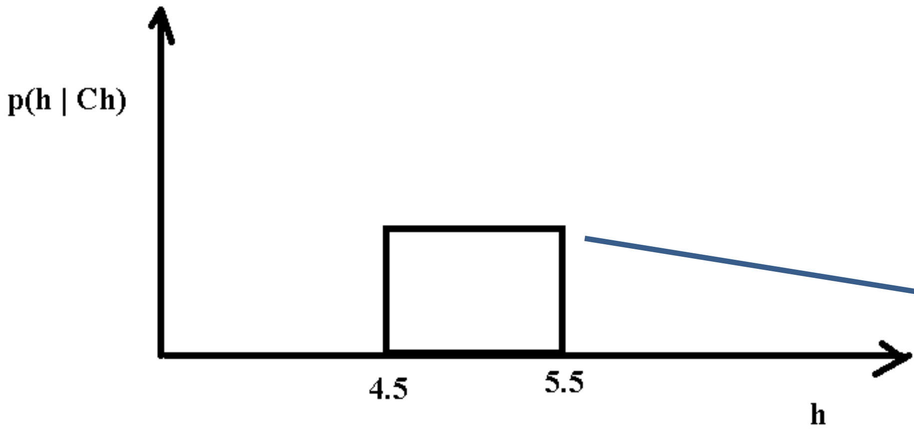
Probability Theory Review

Data is coming from where?

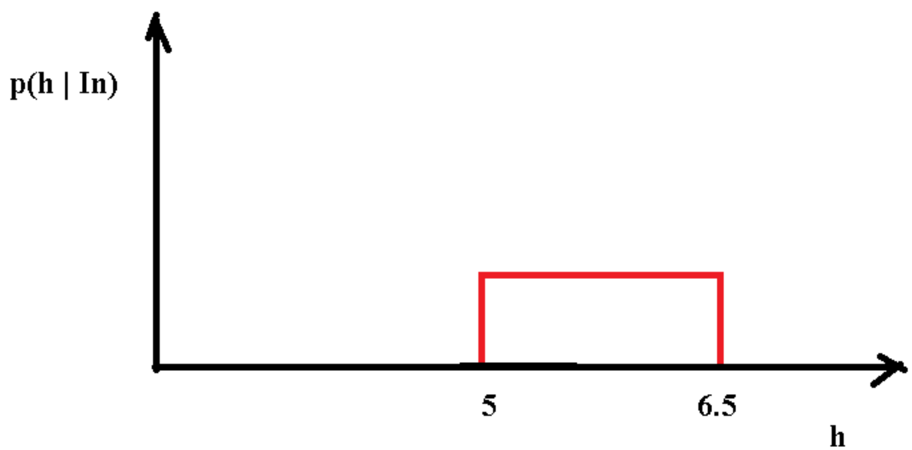
- From a distribution !
- We may not know this distribution from which the data is generated.
- But this is how the nature works !!
- We, often assume simple distributions (like Gaussian, Uniform, etc.)

Example

- Let us take height of people as our data.
- We assume these heights of Chinese are coming from a uniform distribution [4.5, 5.5]
- Similarly assume that the heights of Indians are from [5, 6.5].



What is the height of this rectangle?



- Now in Sri City 30% of people are chinese; rest are Indians.

- Now how to draw heights of people living in Sri City?

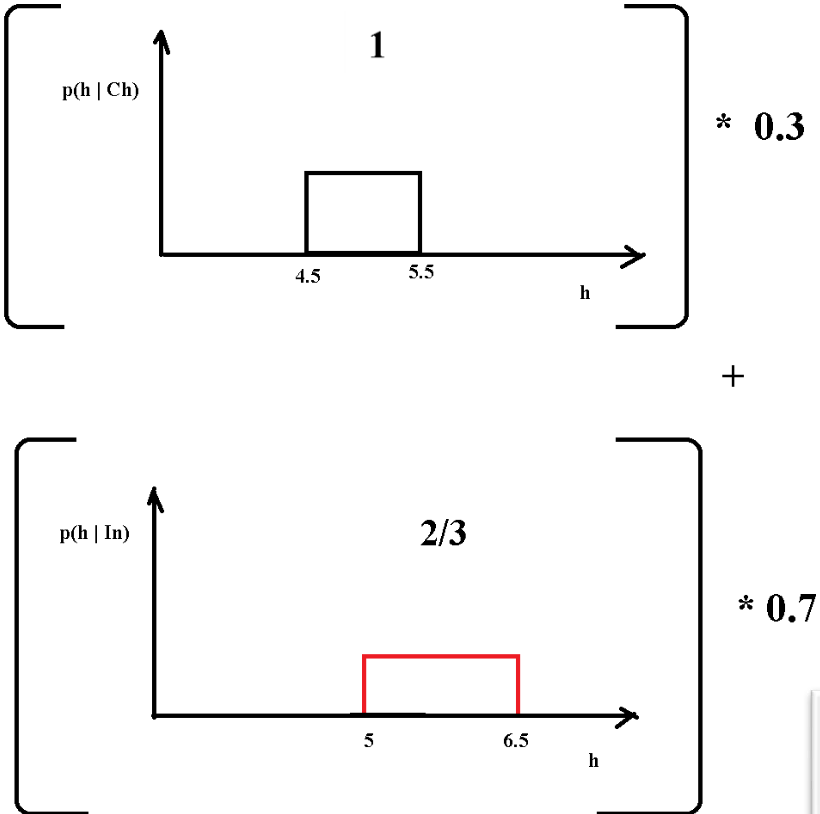
One way is,

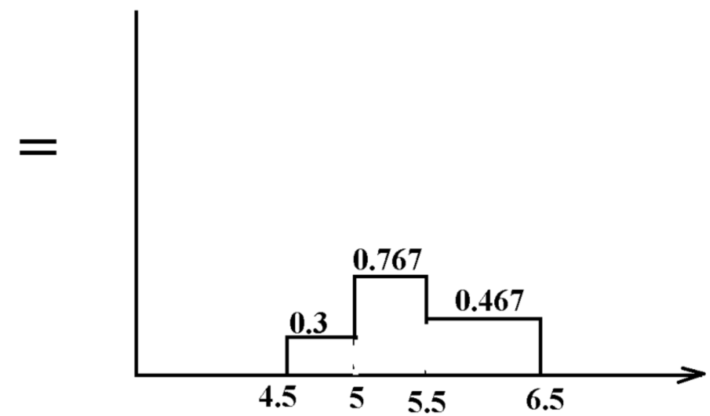
- Toss a biassed coin whose faces are marked “China” and “India”.
- The P(China) = 0.3 and P(India) = 0.7.

- Now toss this coin, if we get China then draw a height from the Chinese distribution;
- Else draw a height from the Indians distribution.

- We are generation a set of numbers which is heights.
- This can be seen to be generated from a mixture … may be you can understand this as…

Mixture





$$p(h) = p(h|Ch)P(Ch) + p(h|In)P(In)$$

$$= \begin{cases} 0.3, & \text{for } h \text{ in } [4.5, 5] \\ 1 * 0.3 + \dfrac{2}{3} * 0.7 = 0.767, & \text{for } h \text{ in } [5, 5.5] \\ \dfrac{2}{3} * 0.7 = 0.467, & \text{for } h \text{ in } [5.5, 6.5] \end{cases}$$

Mixture of two normal distributions

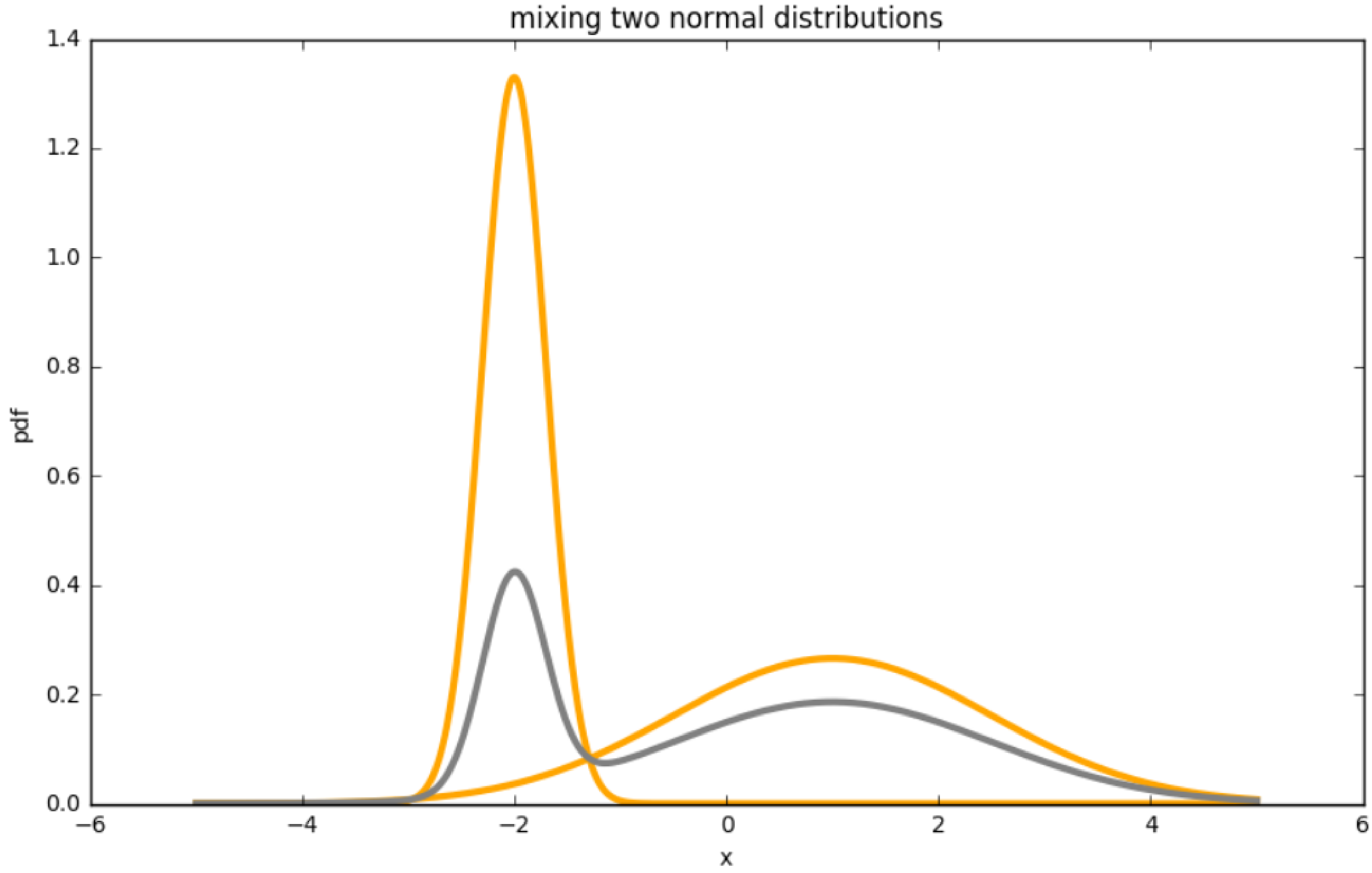
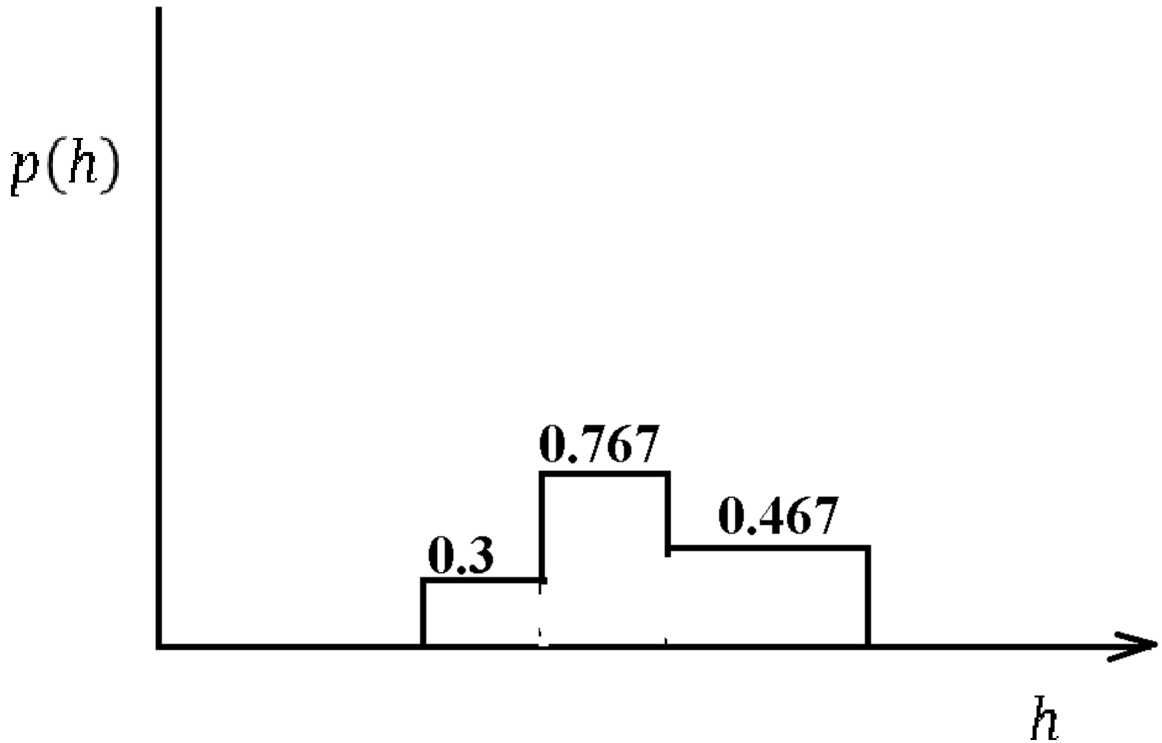


Figure 1. Mixing two normal distributions (orange) to get a bimodal distribution (gray).

Drawing h values from Sri City

- Means drawing from this distribution.



Error

- Now, given $h = 5.6$, somebody classified the person as Indian.
- What is the error of this decision?

- The classifier is correct with probability:

  $$P(Indian \mid h = 5.6)$$

- So the classifier went wrong with the probability: $1 - P(Indian \mid h = 5.6)$

Error rate of a classifer

- For a given pattern X, the classifier outputs the label which is $f(X)$.
- The probability by which this decision is correct is: $P(f(X) \mid X)$
- The probability by which this decision is wrong is: $1 - P(f(X) \mid X)$

Average error

- The probability by which this decision is wrong is: $1 - P(\, f(X) \mid X)$
- Averaging over all patterns means taking the expectation, so
- $P(error) \; = \int [1 - P(\, f(X) \mid X)] \; p(X) \, dX$

What the Bayes Classifier is saying?

- For any given X, see that the probability by which this decision is wrong is, i.e., $1 - P(\, f(X) \mid X)$ is as small as possible.
- So choose the class such that $P(\, f(X) \mid X)$ is maximum.
- So if class labels are $\omega_1, \omega_2, \ldots, \omega_c$ Choose the class ω_k such that

  $P(\omega_k|X) \geq P(\omega_i|X)$ for all i.

- So one has to find the posterior probabilities and act accordingly.