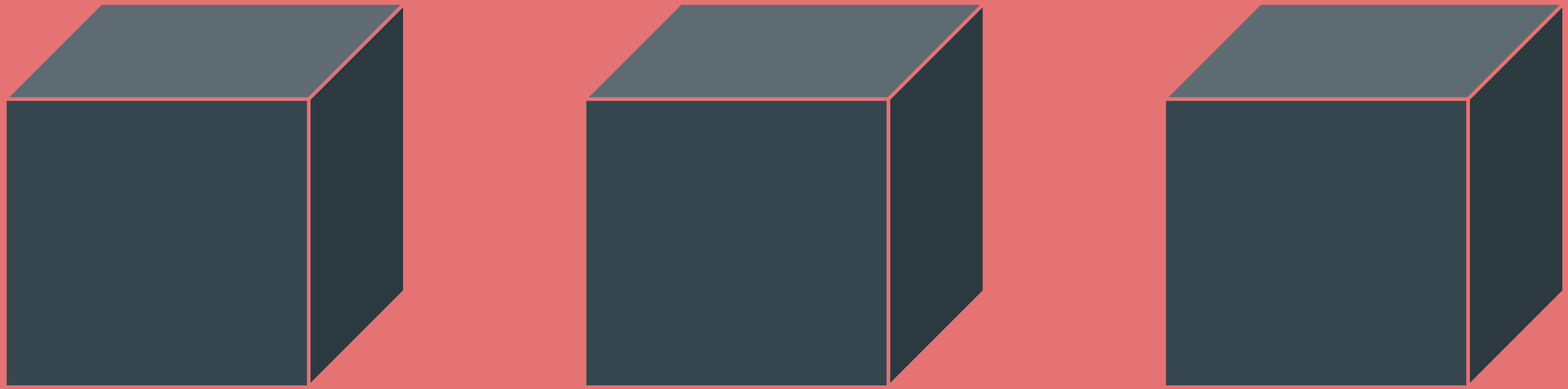
Naïve Bayes Classifier



R=100 red balls

B=100 blue balls

M=50 red balls and 50 blue balls

Suppose that you are to choose a box.

What is the probability that the box you chose is box R? and what is the probability that the box you chose is B? and what about the probability that the box is M?

$$P(H_R) = 0.333$$
$$P(H_B) = 0.333$$
$$P(H_M) = 0.333$$

Suppose that on the box you chose you pick a ball. And the ball you picked turned out to be red.

How does this evidence affect the earlier probabilities?

$$P(H_R|E_R) > P(H_M|E_R) > P(H_B|E_R) = 0$$

Bayes Theorem

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

where:
$P(H|E)$ is the **posterior probability** or the probability that hypothesis H is true given evidence E.
$P(E|H)$ is the **likelihood** or the probability that E is true given H
$P(H)$ is the **prior** probability or the probability that H is true
$P(E)$ is the **marginal likelihood** or the probability that E is true

Applying Bayes Theorem

$$P(H_R|E_R) = \frac{P(E_R|H_R)P(H_R)}{P(E_R)} = \frac{1 \cdot 0.333}{0.5} = 0.667$$

$$P(H_B|E_R) = \frac{P(E_R|H_B)P(H_B)}{P(E_R)} = \frac{0 \cdot 0.333}{0.5} = 0$$

$$P(H_M|E_R) = \frac{P(E_R|H_M)P(H_M)}{P(E_R)} = \frac{0.5 \cdot 0.333}{0.5} = 0.333$$

Applying Bayes Theorem to a General Classification Problem

$$P(H|E) \rightarrow P(h^{(i)}|x^{(i)})$$

$$P\left(x^{(i)} \middle| h^{(i)}\right) = P(x_1^{(i)} | h^{(i)}) P(x_2^{(i)} | h^{(i)})\ P\left(x_3^{(i)} | h^{(i)}\right) \ldots P(x_n^{(i)} | h^{(i)})$$

$$P\left(x^{(i)} \middle| h^{(i)}\right) = \prod_{j=0}^{n} P(x_j^{(i)} | h^{(i)})$$

x_1(Color)	x_2(Size)	y(Fruit)
Red	Big	Apple
Orange	Big	Orange
Red	Small	Cherry
Green	Big	Apple

$$\boldsymbol{P}(\boldsymbol{h} = \textbf{Apple}|[\textbf{Red}, \textbf{Big}])$$

$$= \frac{P(\text{Red}|\text{Apple})P(\text{Big}|\text{Apple})P(\text{Apple})}{P([\text{Red},\text{Big}])}$$

outlook	temp	humidity	windy	play
sunny	hot	high	FALSE	no
sunny	hot	high	TRUE	no
overcast	hot	high	FALSE	yes
rainy	mild	high	FALSE	yes
rainy	cool	normal	FALSE	yes
rainy	cool	normal	TRUE	no
overcast	cool	normal	TRUE	yes
sunny	mild	high	FALSE	no
sunny	cool	normal	FALSE	yes
rainy	mild	normal	FALSE	yes
sunny	mild	normal	TRUE	yes
overcast	mild	high	TRUE	yes
overcast	hot	normal	FALSE	yes
rainy	mild	high	TRUE	no

x_1(Color)	x_2(Size)	y(Fruit)
Red	Big	Apple
Orange	Big	Orange
Red	Small	Cherry
Green	Big	Apple

$$\boldsymbol{P}(\boldsymbol{h} = \mathbf{Orange} | [\mathbf{O}, \mathbf{S}]) = \frac{P(\mathrm{O}|\mathrm{Orange})P(\mathrm{S}|\mathrm{Orange})P(\mathrm{Orange})}{P([\mathrm{O},\mathrm{S}])}$$

$$\boldsymbol{P}(\boldsymbol{h} = \textbf{Orange} | [\mathrm{O}, \boldsymbol{S}]) = \frac{1 \cdot 0 \cdot 0.25}{P([\mathrm{O}, S])} = 0$$

You have estimated $P(h = \text{Orange}|[0, S])$ as 0 because you have never seen a small orange fruit before. This is bad practice

Laplace Smoothing

$$P\left(x_j^{(i)} \middle| h^{(i)}\right) = \frac{N_{x_j^{(i)},h^{(i)}} + \alpha}{N_{x_j^{(i)}} + \alpha k}$$

where:

$N_{x_j^{(i)},h^{(i)}}$ is the number of $h^{(i)}$ with feature values $x_j^{(i)}$

$N_{x_j^{(i)}}$ is the total number of examples with feature values $x_j^{(i)}$

k is the number of possible values for $x_j^{(i)}$

α is the smoothing parameter

Naïve Bayes on Continuous Features

To convert continuous feature values of to discrete values, map the features into ranges.

$$c_j^{(i)} = \begin{cases} \text{high} & if & x_j^{(i)} \geq 100 \\ \text{medium} & if & 100 > x_j^{(i)} \geq 50 \\ \text{low} & if & x_j^{(i)} < 50 \end{cases}$$