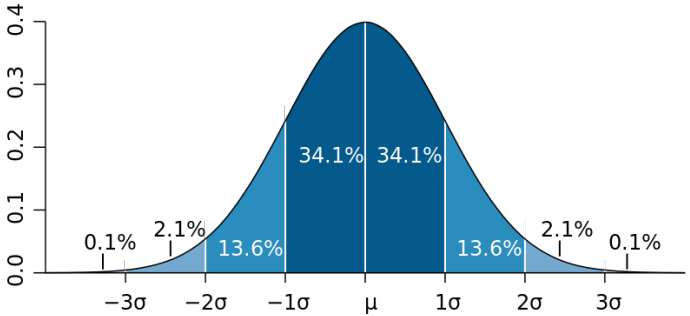
Naive Bayes Classifier

Supervised Machine Learning

Basic statistics

- Mean (average)
- Variance
- Standard deviation
- Gaussian distribution (normal distribution)



(image from wikipedia.org)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

Bayes Theorem

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \qquad \text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}.$$

Calculating most probable hypothesis based on conditional probabilities

Notations

P(A)

Prior probability of A

Probability of rain is 5 %

P(A|B)

Probability of A given B

Probability of rain if cloudy is 50%

Example of Bayes Theorem

Suppose you go to a forest. You see tree. You want to find out if it is an orange tree or apple tree.

You know:

- Forest has 30 % orange trees and 70 % apple
- You notice that tree has white flowers. 10 % of apple trees have white flowers. 60% of orange trees

Solution

O - orange

A - apple

F - flower

$$P(O|F) = \frac{P(F|O)P(O)}{P(F)} = \frac{P(F|O)P(O)}{P(O)P(F|O) + P(A)P(F|A)}$$

$$P(O|F) = \frac{0.60 * 0.30}{0.30 * 0.60 + \ 0.70 * 0.10} = 0.72$$

Naive Bayes Classifier

"Naive" because assumes independent features

$$v_{NB} = \underset{v_j \in V}{\operatorname{argmax}} P(v_j) \prod_i P(a_i | v_j)$$

Example of Naive Bayes

- Number of leaves
- Height of the tree

Class	Number of leaves	Height (m)
Orange	102	5.1
Orange	131	5.6
Orange	95	6.4
Apple	404	3.1
Apple	350	3.5
Apple	464	2.6
????	500	2.1

Applying Bayes Theorem

- L - leaves
- H- height
- O - orange
- A - apple

$$p(O) = \frac{P(O)P(L|O)P(H|O)}{P(O)P(L|O)P(H|O) + P(A)P(L|A)P(H|A)}$$

$$p(A) = \frac{P(A)P(L|A)P(H|A)}{P(O)P(L|O)P(H|O) + P(A)P(L|A)P(H|A)}$$

Getting basic statistics

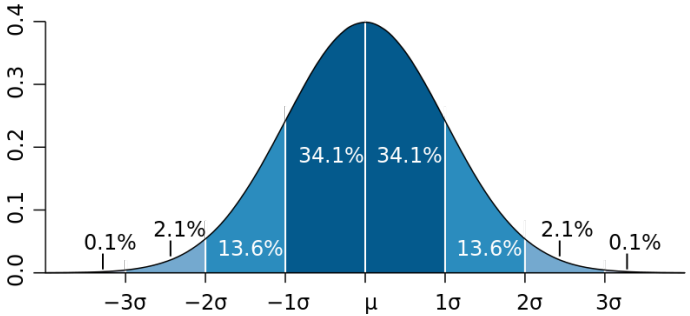
	Mean of # of leaves	Standard deviation # of leaves
Orange	109	19
Apple	406	57

	Mean height	Standard deviation of height
Orange	6	0.66
Apple	3	0.45

Using normal distribution

$$P(L|O) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)\text{^}2}{2\sigma^2}} = \frac{1}{19\sqrt{2\pi}} e^{-\frac{(500-109)\text{^}2}{2(19)^2}} = 2.3e-94$$

$$P(L|A) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)\text{^}2}{2\sigma^2}} = \frac{1}{57\sqrt{2\pi}} e^{-\frac{(500-406)\text{^}2}{2(57)^2}} = 0.0018$$



$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

$$p(O) = \frac{P(O)P(L|O)P(H|O)}{P(O)P(L|O)P(H|O) + P(A)P(L|A)P(H|A)}$$

$$= \frac{(0.30)(2.3e-94)(1.6e-08)}{(0.30)(2.3e-94))(1.6e-08) + (0.70)(0.12)(0.0018)}$$

$$\approx 0$$

$$p(A) = \frac{P(A)P(L|A)P(H|A)}{P(O)P(L|O)P(H|O) + P(A)P(L|A)P(H|A)}$$

$$= \frac{(0.70)(0.12)(0.0018)}{(0.30)(2.3e-94))(1.6e-08) + (0.70)(0.12)(0.0018)}$$

$$\approx 1$$

Advantages / Disadvantages

Advantages

- Fast
- Can make probabilistic predictions

Disadvantages

- Can’t do regression
- A lot of assumptions: gaussian and independent.