Bayes classifier

Expected loss of the sample **x**:

$$R(c_i|x) = \sum_{j=1}^{N} \lambda_{ij} P(c_j|x)$$

Bayes decision rule:

$$h^*(x) = \arg min_{c \in Y} R(c|x)$$

0-1 loss:

$$\lambda_{ij} = \begin{cases} 0, & if\ i = j \\ 1, & otherwise \end{cases}$$

$$R(c|x) = 1 - P(c|x)$$

$$h^*(x) = \arg max_{c \in Y} P(c|x)$$

Bayes classifier

Bayes' theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$h^*(x) = \arg max_{c \in Y} P(c|x) = \arg max_{c \in Y} \frac{P(x|c)P(c)}{P(x)} = \arg \ max_{c \in Y} P(x|c)P(c)$$

class-conditional/likelihood probability

prior probability

Naive Bayes

- Attribute conditional independence assumption:

$$P(x|c) = \prod_{i=1}^{d} P(x_i|c)$$

$$h^*(x) = \arg max_{c \in Y} P(c|x) = \arg max_{c \in Y} P(c) \prod_{i=1}^{d} P(x_i|c)$$

prior probability conditional probability

Naive Bayes

- Discrete Attribute

$$P(c) = \frac{|D_c|}{|D|} \qquad P(x|c) = \frac{|D_{c,x_i}|}{|D_c|}$$

Problem

Solution: Bayes Estimation

$$P_\lambda(c) = \frac{|D_c| + \lambda}{|D_c| + N_c * \lambda} \qquad P_\lambda(x|c) = \frac{|D_{c,x_i}| + \lambda}{|D| + N * \lambda}$$

$\lambda = 1$ Laplace smoothing

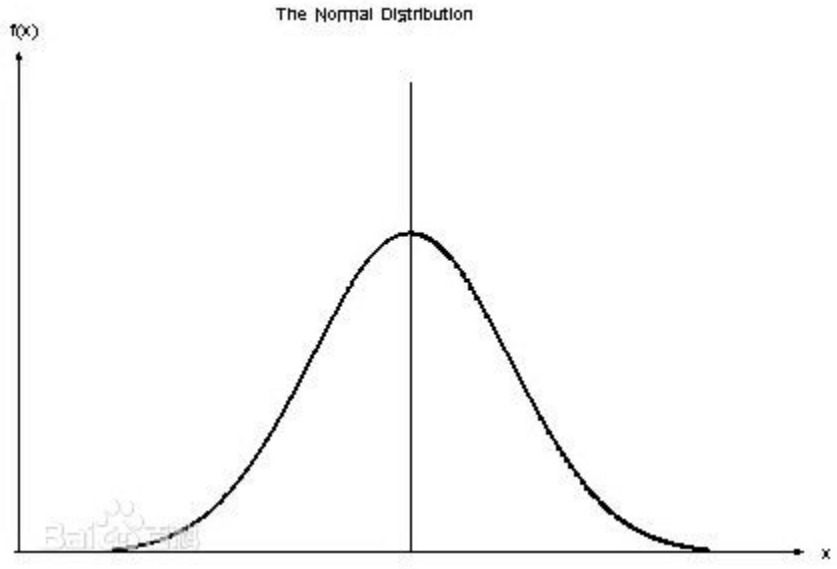
Naive Bayes

- Continuous Attribute

  Make assumptions about the distribution of each feature data

Gaussian Naive Bayes

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right)$$



Multinomial Naive Bayes

Bernoulli Naive Bayes

Naive Bayes

“我司可办理正规发票（保真）17%增值税发票点数优惠！”

$$P(\text{“垃圾邮件”}|(\text{“我”},\text{“司”},\text{“可”},\text{“办理”},\text{“正规发票”},\text{“保真”},\text{“增值税”},\text{“发票”},\text{“点数”},\text{“优惠”})) = \frac{P((\text{“我”},\text{“司”},\text{“可”},\text{“办理”},\text{“正规发票”},\text{“保真”},\text{“增值税”},\text{“发票”},\text{“点数”},\text{“优惠”})|\text{“垃圾邮件”})P(\text{“垃圾邮件”})}{P((\text{“我”},\text{“司”},\text{“可”},\text{“办理”},\text{“正规发票”},\text{“保真”},\text{“增值税”},\text{“发票”},\text{“点数”},\text{“优惠”}))}$$

Multinomial Naive Bayes

$$P((\text{“代开”},\text{“发票”},\text{“增值税”},\text{“发票”},\text{“正规”},\text{“发票”})|S)$$
$$= P(\text{“代开””}|S)P(\text{“发票”}|S)P(\text{“增值税”}|S)P(\text{“发票”}|S)P(\text{“正规”}|S)P(\text{“发票”}|S)$$
$$= P(\text{“代开””}|S)P^3(\text{“发票”}|S)P(\text{“增值税”}|S)P(\text{“正规”}|S)$$

注意这一项：$P^3(\text{“发票”}|S)$。

Bernoulli Naive Bayes

$$P((\text{“代开”},\text{“发票”},\text{“增值税”},\text{“发票”},\text{“正规”},\text{“发票”})|S)$$
$$= P(\text{“发票”}|S)P(\text{“代开””}|S)P(\text{“增值税”}|S)P(\text{“正规”}|S)$$

Naive Bayes

	文档ID	文档中的词	属于$c = China$类
训练集	1	Chinese Beijing Chinese	Yes
	2	Chinese Chinese Shanghai	Yes
	3	Chinese Macao	Yes
	4	Tokyo Japan Chinese	No
测试集	5	Chinese Chinese Chinese Tokyo Japan	?

$$P(c) = 3/4, \quad P(\overline{c}) = 1/4$$

$$P(Chinese|c) = (5+1)/(8+6) = 3/7$$
$$P(Tokyo|c) = P(Japan|c) = (0+1)/(8+6) = 1/14$$
$$P(Chinese|\overline{c}) = (1+1)/(3+6) = 2/9$$
$$P(Tokyo|\overline{c}) = P(Japan|\overline{c}) = (1+1)/(3+6) = 2/9$$

$$P(c|d_5) \propto P(c) \cdot P(Chinese|c)^3 \cdot P(Tokyo|c) \cdot P(Japan|c)$$
$$= \frac{3}{4} \cdot (\frac{3}{7})^3 \cdot \frac{1}{14} \cdot \frac{1}{14} = 0.0003$$

$$P(\overline{c}|d_5) \propto P(\overline{c}) \cdot P(Chinese|\overline{c})^3 \cdot P(Tokyo|\overline{c}) \cdot P(Japan|\overline{c})$$
$$= \frac{1}{4} \cdot (\frac{2}{9})^3 \cdot \frac{2}{9} \cdot \frac{2}{9} = 0.0001$$

Naive Bayes

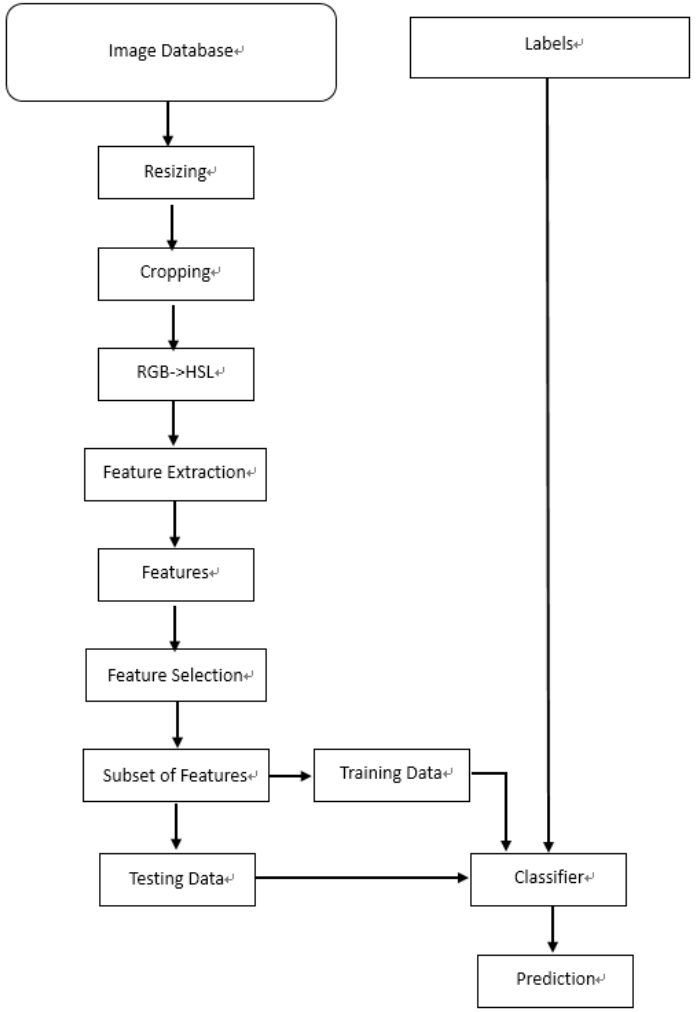
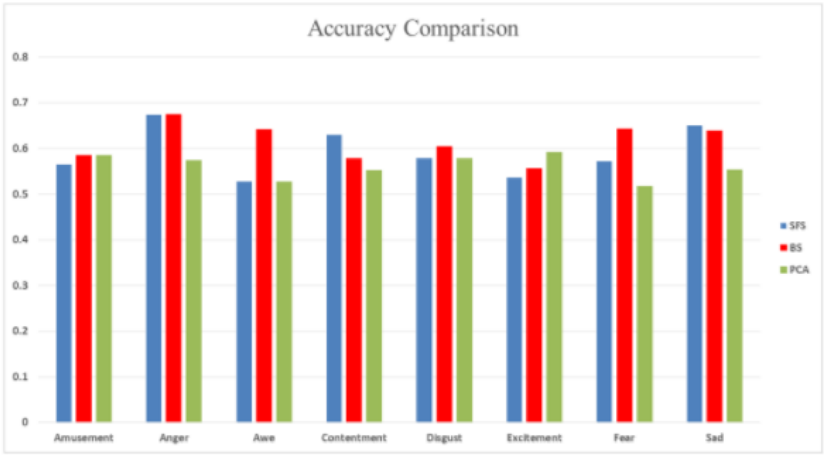


Image Database
Labels
Resizing
Cropping
RGB->HSL
Feature Extraction
Features
Feature Selection
Subset of Features
Training Data
Testing Data
Classifier
Prediction


Accuracy Comparison
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Amusement
Anger
Awe
Contentment
Disgust
Excitement
Fear
Sad
SFS
BS
PCA