BAYESIAN LEARNING: NAIVE BAYES

Dr. Hilman F. Pardede

Research Center for Informatics
Indonesian Institute of Sciences

The materials are compiled from the following resources:

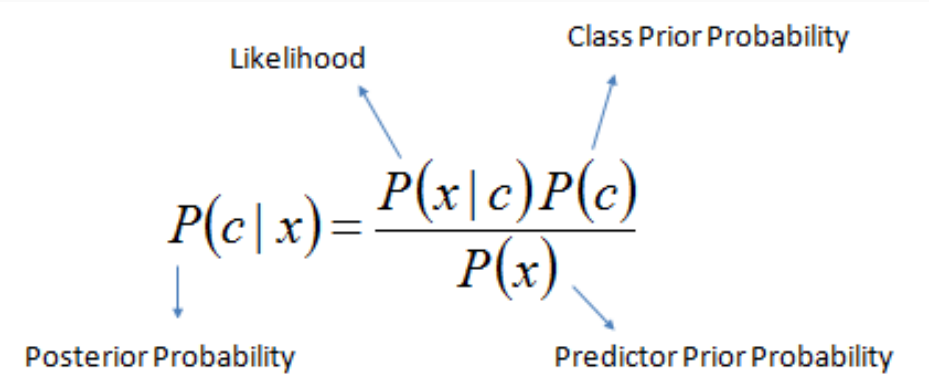
- https://github.com/joaquinvanschoren/ML-course
- https://www.cse.iitk.ac.in/users/piyush/courses/ml_autumn16/ML.html
- http://sli.ics.uci.edu/Classes/2015W-273a

NAIVE BAYES BASICS

NAIVE BAYES

- Predict the probability that a point belongs to each class, using Bayes' Theorem, assuming that the features are independent from each other.
- Very fast. They work by only extracting statistics from each feature.

BAYES THEOREM



- $P(c|x)$ is the posterior probability of class (target) given predictor (attribute).
- $P(c)$ is the prior probability of class.
- $P(x|c)$ is the likelihood which is the probability of predictor given class.
- $P(x)$ is the prior probability of predictor (evidence or marginal likelihood).
- Naive Bayes assumes that all features are conditionally independent from each other, in which case:
  $P(\mathbf{x}|c) = P(x_1|c) \times P(x_2|c) \times ... \times P(x_n|c)$

NAIVE BAYES EXAMPLES: SPAM DETECTION

- $M << 2^N$ parameters
- $y \in 2\{spam, notspam\}$
- $X =$ observed words in email
  - Ex: [“the” ... “probabilistic” ... “lottery”...]
  - “1” if word appears; “0” if not
- 1000’s of possible words: 2^1000 parameters?
- $\#$ of atoms in the universe: $>> 2^{270}$
- Model words given email type as independent
- Some words more likely for spam (“lottery”)
- Some more likely for real (“Tugas kuliah”)
- Only 1000’s of parameters now

COMPUTING PROBABILITIES FOR BAYES CLASSIFIERS

- True or not? Players will play if weather is sunny.

COMPUTING PROBABILITIES FOR BAYES CLASSIFIERS

- True or not? Players will play if weather is sunny.

$P(x \mid c) = P(Sunny \mid Yes) = 3/9 = 0.33$

Frequency Table		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3

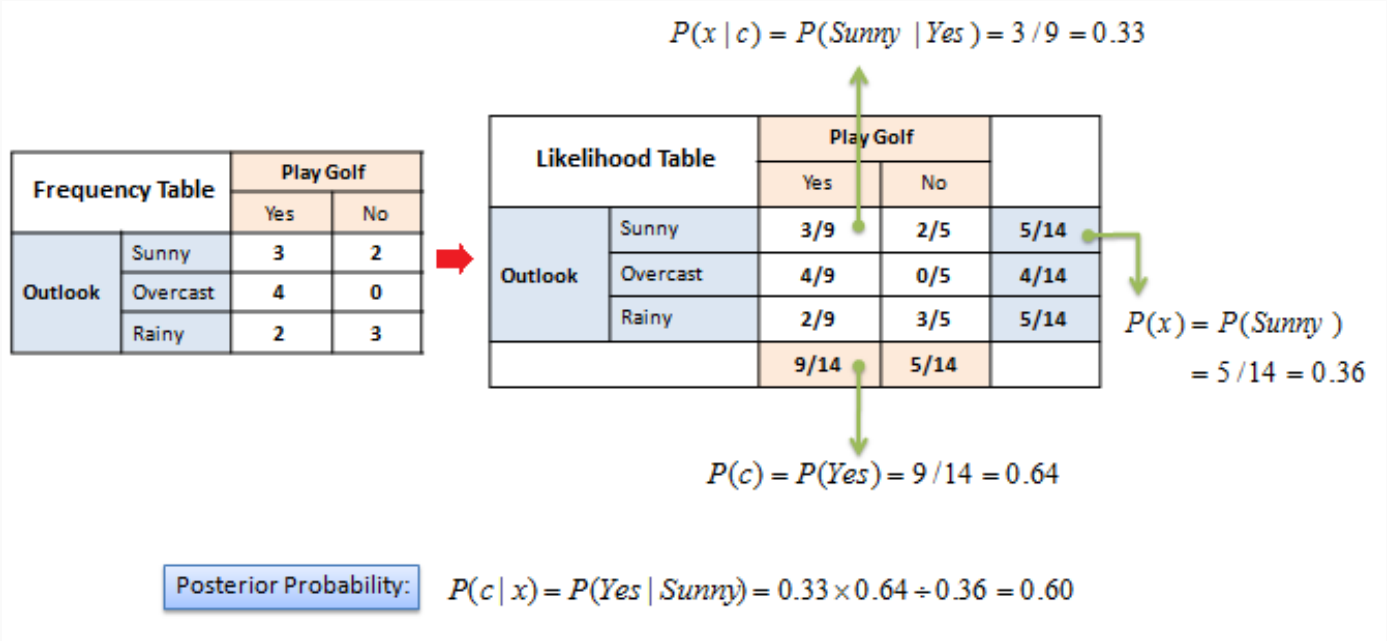
Likelihood Table		Play Golf		
		Yes	No	
Outlook	Sunny	3/9	2/5	5/14
	Overcast	4/9	0/5	4/14
	Rainy	2/9	3/5	5/14
		9/14	5/14	

$P(x) = P(Sunny) = 5/14 = 0.36$

$P(c) = P(Yes) = 9/14 = 0.64$

Posterior Probability: $P(c \mid x) = P(Yes \mid Sunny) = 0.33 \times 0.64 \div 0.36 = 0.60$

- True or not? Players will play if weather is sunny.



- Compute the posterior for every class and predict the class with highest probability

GAUSSIAN NAIVE BAYES

GAUSSIAN NAIVE BAYES

- Computes mean μ_c and standard deviation σ_c of the feature values per class
- It then fits a Gaussian distribution around the mean
  $p(x = v \mid c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{(v-\mu_c)^2}{2\sigma_c^2}}$
- Prediction are made using Bayes' theorem, by computing the joint probability given all features
  $p(c \mid \mathbf{x}) = \frac{p(c)\ p(\mathbf{x}|c)}{p(\mathbf{x})}$

OTHER NAIVE BAYES METHODS

- BernoulliNB
  - Assumes binary data
  - Feature statistics: Number of non-zero entries per class
- MultinomialNB
  - Assumes count data
  - Feature statistics: Average value per class

  Mostly used for text classification (bag-of-words data)