

Sample title

Brandon Hosley

University of Illinois - Springfield

September 27, 2020

Overview

- 1 Q1A: What is a Naïve-Bayes classifier?
- 2 Q1B: What are the evaluation metrics for classification in machine learning?
- 3 Q2: Hastie and Tibshirani Summary

Naïve-Bayes

- Based on Bayes' Theorem

Bayes' Theorem

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

Naïve-Bayes

- Based on Bayes' Theorem
- By comparing a series of known probabilities one may infer the probability of an associated outcome/event/attribute

Bayes' Theorem

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

Bayes' Theorem Generalized

$$P(C_k|x_1, \dots, x_n) = P(C_k) \prod_{i=1}^n \frac{P(x_i|C_k)}{P(x_i)}$$

Naïve-Bayes

- Based on Bayes' Theorem
- By comparing a series of known probabilities one may infer the probability of an associated outcome/event/attribute
- Naïve means that the known variables are being treated as independent

Bayes' Theorem

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

Bayes' Theorem Generalized

$$P(C_k|x_1, \dots, x_n) = P(C_k) \prod_{i=1}^n \frac{P(x_i|C_k)}{P(x_i)}$$

Naïve-Bayes classification

- From training data we can build our classifier

Bayes' Theorem Generalized

$$P(C_k|x_1, \dots, x_n) = P(C_k) \prod_{i=1}^n \frac{P(x_i|C_k)}{P(x_i)}$$

Naïve-Bayes classification

- From training data we can build our classifier
 - From the data set we can estimate $P(x)$ and $P(C_k)$ for all of our attributes
- If these values are not similar in the test set it warns of a poor sample

Bayes' Theorem Generalized

$$P(C_k|x_1, \dots, x_n) = P(C_k) \prod_{i=1}^n \frac{P(x_i|C_k)}{P(x_i)}$$

Naïve-Bayes classification

- From training data we can build our classifier
- From the data set we can estimate $P(x)$ and $P(C_k)$ for all of our attributes
If these values are not similar in the test set it warns of a poor sample
- $P(x_i|C_k)$ is determined from the sample data and is the resource intensive aspect of developing a classifier in this method.

Bayes' Theorem Generalized

$$P(C_k|x_1, \dots, x_n) = P(C_k) \prod_{i=1}^n \frac{P(x_i|C_k)}{P(x_i)}$$

Common Evaluation Metrics

- Confusion Matrix

Example:

		True diagnosis		Total
		Positive	Negative	
Screening test	Positive	a	b	$a + b$
	Negative	c	d	$c + d$
Total		$a + c$	$b + d$	N

Common Evaluation Metrics

- Confusion Matrix
- F1 Score

Example:

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Note:

This is the Harmonic mean of Precision and Recall

Common Evaluation Metrics

- Confusion Matrix
- F1 Score
- Binary Cross Entropy

Example:

$$BCE = -(y \log(p) + (1 - y) \log(1 - p))$$

Note:

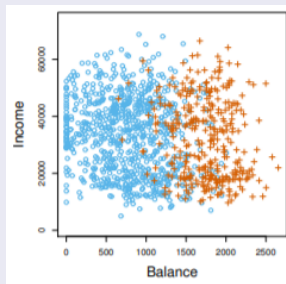
This method works well examining results given as probabilities of each classification

Tibshirani Lecture: Classification

Using available data to predict a future or current qualitative feature on a variable.

Example:

Predicting if an account will go into credit card default

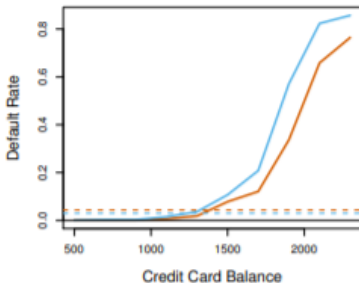


Tibshirani Lecture: Classification

- Linear Regression
 - Qualitative nature of classification limits usefulness
 - Capable of being used as linear discriminant analysis

Tibshirani Lecture: Classification

- Linear Regression
 - Qualitative nature of classification limits usefulness
 - Capable of being used as linear discriminant analysis
- Logistic Regression
 - Typically better than Linear as it favors bimodal bias
 - Multiple Logical Regression is capable of identifying underlying relationships



Tibshirani Lecture: Classification

- Bayes' Theorem
 - Covered in earlier slides

Tibshirani Lecture: Classification

- Bayes' Theorem
 - Covered in earlier slides
- Classifying to the highest density
 - Calculating decision boundaries based on their density
 - Predictions made by which side of boundaries new data falls on

