SPECTF Heart Data Prediction using Bayes Classifier with modified Probability Distribution Functions and Feature Selection

Arlind Stafad

CIS 5800 Machine Learning
Final Project

INTRODUCTION

The dataset describes diagnosing of cardiac Single Proton Emission Computed Tomography (SPECT) images. Each of the patients is classified into two categories: normal and abnormal. The database of 267 SPECT image sets (patients) was processed to extract features that summarize the original SPECT images. As a result, 44 continuous feature patterns were created for each patient. The dataset had no missing attributes. The training data included 39 class 0 instances and 40 class 1 instances. The test data included 15 class 0 instances and 171 class 1 instances. There was no need to use an imputer to fill missing values or deal with them in general since there were none. The data were all continuous values.

For this dataset I built a Bayes classification algorithm and worked with three different probability distributions, Gaussian, Uniform, Rayleigh. The implementations had varied results on accuracy and time. Filter methods and probability distributions affected each in its own way. I also used support vector machines, SVM, from the sklearn library in python to test accuracy of data through SVM's and its settings.

METHODS:

For the Bayes Classifier I used three different probability distributions. The bayes theorem states that

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

OR →

$$P(y|X) = \frac{P(X|y) \cdot P(y)}{P(X)}$$

X is the data vector and holds (x1, x2, x3, ... , xn) I assumed that all features were mutually independent so

$$P(y|X) = \frac{P(x_1|y) \cdot P(x_2|y) \cdot \ldots \cdot P(x_n|y) \cdot P(y)}{P(X)}$$

This resulted in the probability of a class prediction to be:

$$y = argmax_y P(y|X) = argmax_y \frac{P(x_1|y) \cdot P(x_2|y) \cdot \ldots \cdot P(x_n|y) \cdot P(y)}{P(X)}$$

Because we are trying to find the maximum with respect to y we can get rid of the P(X):

$$y = argmax_y P(x_1|y) \cdot P(x_2|y) \cdot \ldots \cdot P(x_n|y) \cdot P(y)$$

In order to make this easier to deal with I then applied a a logarithm to the function:

$$y = argmax_y \log(P(x_1|y)) + \log(P(x_2|y)) + \ldots + \log(P(x_n|y)) + \log(P(y))$$

Bayes Classification with Gaussian Probability Distributions:

I implemented the Bayes algorithm to deal with a gaussian distribution of data and that required implementing the Gaussian distribution probability density function:

Conditional probability for the Gaussian distribution can be summarized as:

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

While computing the parameters from training data (mean and standard deviation), there was a slight decrease in performance accuracy if I used variance instead of standard deviation as there was truncation in calculating the formula which decreased the real representation of a value.

Bayes Classification with Uniform Probability Distributions:

The functional change between Gaussian and Uniform probability was the conditional probability used:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b, \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$$

Bayes Classification with Rayleigh Probability Distributions:

$$f(x;\sigma) = \frac{x}{\sigma^2} e^{-x^2/(2\sigma^2)}, \quad x \geq 0,$$

Where sigma is a scale parameter. This parameter I got through

$$\widehat{\sigma}^2 \approx \frac{1}{2N} \sum_{i=1}^{N} x_i^2$$

FEATURE SELECTION:

I used two different methods of feature selection. The Filter method and wrapper method.

Filter Method:

For the filter method I used the Pearson coefficient correlation (PCC) to get information about which variables were more correlated with a classification of an instance in the data. The Pearson correlation coefficient is used to measure the strength of a linear association between two variables, where the value $r = 1$ means a perfect positive correlation and the value $r = -1$ means a perfect negative correlation.

$$r = \frac{\sum_i (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_i (x_i - \overline{x})^2} \sqrt{\sum_i (y_i - \overline{y})^2}}$$

Based on the results from this computation I then proceeded to add features and classify them based on current number of features added. At the optimal point I got which features train data indicated as most important and used them to compute results on test data.

Wrapper Method

For the wrapper method I used a greedy approach where I select the best performing feature starting from a single feature. Then I test the selected best performing feature and run through a loop that tests that current best performing feature with a second feature from all other available features. When the best performing pair is chosen the process continues into best three performers, four, and so on up to the number of features available. What makes this algorithm greedy is that a new feature is added as long as there are still improvements in the accuracy of the algorithm. Otherwise it halts immediately, regardless of whether further combinations could yield better results. An optimization I used with the wrapper method was making choices in pairs instead of single additions in each to see the impact it will have on the performance of the algorithm.

RESULTS

Algorithm	Accuracy	Time
Bayes Gaussian (train)	84.81012658227847	Apprx. 0.002s
Bayes Gaussian (test)	82.79569892473118	Apprx. 0.002s
Bayes Uniform (train)	84.81012658227847	Apprx. 0.0015s
Bayes Uniform (test)	78.49462365591397	Apprx. 0.0016s
Bayes Rayleigh (train)	49.36708860759494	Apprx. 0.0017s
Bayes Rayleigh (test)	91.93548387096774	Apprx. 0.0016s

ALGORITHM WITH FEATURE SELECTION

Algorithm (with Filter Method)	Accuracy	Time
Bayes Gaussian (train)	88.60759493670885	Apprx. 0.94s
Bayes Gaussian (test)	82.79569892473118	
Bayes Uniform (train)	87.34177215189874	Apprx. 0.97s
Bayes Uniform (test)	79.03225806451613	
Bayes Rayleigh (train)	77.21518987341773	Apprx. 0.95s
Bayes Rayleigh (test)	91.93548387096774	

When computing the filter method we need to generate PCC scores mentioned previously. A result of the rated PCC features based on importance is printed and added from command line:

```
Features Rated in terms of importance:

f39 :  0.5218058162505883
f29 :  0.480915392533332
f25 :  0.4172977673046537
f41 :  0.4088889628523885
f23 :  0.39291284741535293
f42 :  0.39142889656853685
f24 :  0.38395263510491956
f38 :  0.3821798722495768
f27 :  0.37159920367282684
f40 :  0.3666839105746829
f3 :  0.3532012695374648
f43 :  0.3448355015966223
f28 :  0.3295248675286953
f31 :  0.3268492968359196
f21 :  0.32109880550693026
f13 :  0.3094233519299062
f15 :  0.29723911152539706
f2 :  0.2669585154683029
f5 :  0.25488245682148924
f22 :  0.25405214956274386
f19 :  0.23771461333719726
f30 :  0.2338780674210629
f7 :  0.22740786704893023
f9 :  0.21561275024811533
f35 :  0.20023032541200805
f14 :  0.1940918040381662
f32 :  0.18679839477541493
f37 :  0.17242951878521426
f0 :  0.16099288779357704
f33 :  0.15957214031156935
f11 :  0.15820798055674837
f17 :  0.15596207277426052
f26 :  0.1385697044048728
f12 :  0.13767794608151143
f1 :  0.1298302817781699
f18 :  0.10581109668620997
f10 :  0.09374987289407934
f34 :  0.08382504901658588
f20 :  0.0833026816638612
f16 :  0.07173801860658414
f4 :  0.04475640738198739
f6 :  0.043096894942594194
f36 :  0.012096326221496687
f8 :  0.009322821586624652
```

Algorithm (with Wrapper Method)	Accuracy	Time
Bayes Gaussian (train)	92.40506329113924	Apprx. 4.5s
Bayes Gaussian (test)	79.56989247311827	
Bayes Uniform (train)	91.13924050632912	Apprx. 3.5s
Bayes Uniform (test)	67.74193548387096	
Bayes Rayleigh (train)	78.48101265822784	Apprx. 0.78s
Bayes Rayleigh (test)	47.31182795698925	

The wrapper method on Gaussian and best features selected

```
The wrapper method feature selection using Sequential Forward Selection:
   [39, 25, 29, 17, 1, 4, 6, 9, 23, 2, 30, 16, 0, 32, 21, 42, 28]
The maximum accuracy is:  92.40506329113924 at  17  iterations
```

The wrapper method on Uniform and best features selected

```
The wrapper method feature selection using Sequential Forward Selection:
   [39, 25, 29, 1, 16, 17, 10]
The maximum accuracy is:  91.13924050632912 at  7  iterations
```

The wrapper method on Rayleigh and best features selected

```
The wrapper method feature selection using Sequential Forward Selection:
   [25, 39]
The maximum accuracy is:  78.48101265822784 at  2  iterations
```

SVM Accuracy on data with gamma set to auto = 0.7473118279569892