Bayesian Classifiers

Naive Bayes

Naive Bayes is a probabilistic classifier based on Bayes theorem the makes the assumption that all features are independent of each other. The strong assumption of independent features gives the classifier its name "Naive" Bayes. This makes the classifier susceptible to error when features are not totally independent of each other. The normalized weather set was used on 1st Run of the Naive Bayes and the numeric set on the 2nd Run. The 2nd run had a higher correct classification percent using the numeric data set.

1st Run

Scheme: weka.classifiers.bayes.NaiveBayes

Relation: weather.symbolic

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Naive Bayes Classifier

Class

Attribute yes no

(0.63) (0.38)

=============================

outlook

sunny 3.0 4.0

overcast 5.0 1.0

rainy 4.0 3.0

[total] 12.0 8.0

temperature

hot 3.0 3.0

mild 5.0 3.0

cool 4.0 2.0

[total] 12.0 8.0

humidity

high 4.0 5.0

normal 7.0 2.0

[total] 11.0 7.0

windy

TRUE 4.0 4.0

FALSE 7.0 3.0

[total] 11.0 7.0

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 8 57.1429 %

Incorrectly Classified Instances 6 42.8571 %

Kappa statistic -0.0244

Mean absolute error 0.4374

Root mean squared error 0.4916

Relative absolute error 91.8631 %

Root relative squared error 99.6492 %

Total Number of Instances 14

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.778 0.8 0.636 0.778 0.7 0.578 yes

0.2 0.222 0.333 0.2 0.25 0.578 no

Weighted Avg. 0.571 0.594 0.528 0.571 0.539 0.578

=== Confusion Matrix ===

a b <-- classified as

7 2 | a = yes

4 1 | b = no

2nd Run

Scheme: weka.classifiers.bayes.NaiveBayes

Relation: weather

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Naive Bayes Classifier

Class

Attribute yes no

(0.63) (0.38)

===============================

outlook

sunny 3.0 4.0

overcast 5.0 1.0

rainy 4.0 3.0

[total] 12.0 8.0

temperature

mean 72.9697 74.8364

std. dev. 5.2304 7.384

weight sum 9 5

precision 1.9091 1.9091

humidity

mean 78.8395 86.1111

std. dev. 9.8023 9.2424

weight sum 9 5

precision 3.4444 3.4444

windy

TRUE 4.0 4.0

FALSE 7.0 3.0

[total] 11.0 7.0

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 9 64.2857 %

Incorrectly Classified Instances 5 35.7143 %

Kappa statistic 0.1026

Mean absolute error 0.4649

Root mean squared error 0.543

Relative absolute error 97.6254 %

Root relative squared error 110.051 %

Total Number of Instances 14

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.889 0.8 0.667 0.889 0.762 0.444 yes

0.2 0.111 0.5 0.2 0.286 0.444 no

Weighted Avg. 0.643 0.554 0.607 0.643 0.592 0.444

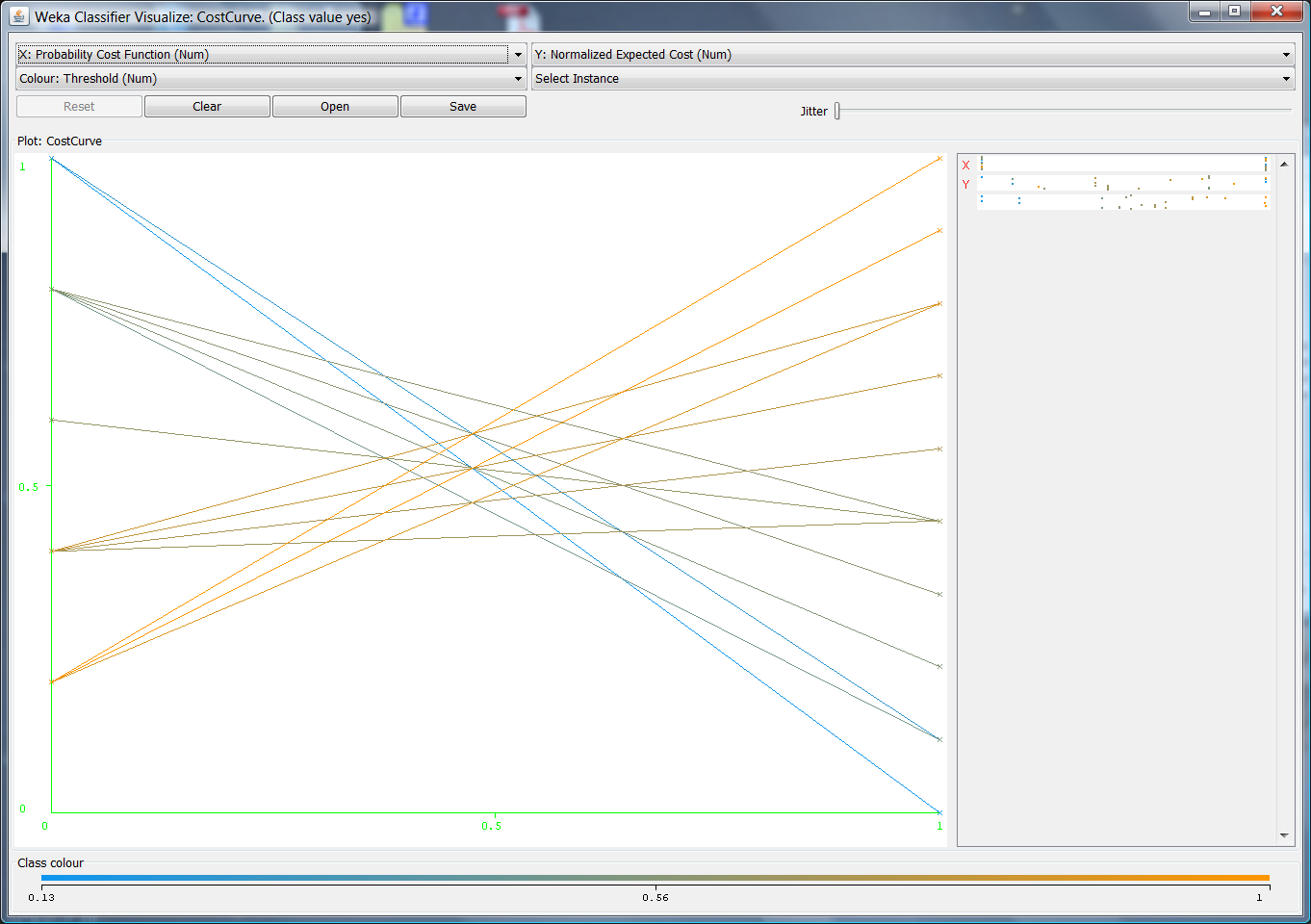
=== Confusion Matrix ===

a b <-- classified as

8 1 | a = yes

4 1 | b = no

Cost Curve Visualization for the 2nd Naive Bays run. I am not sure what it really means but it looks interesting.



BayesNet

BayesNet classifier builds a probabilistic network based on various search algorithms and quality measurements. Multiple runs were done using BayesNet changing the search function and estimator function. The 2nd run had the highest correct classification % but again it just classified everything as play-yes which is more a less a dumb classification.

1st Run

Scheme: weka.classifiers.bayes.BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5

Relation: weather

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Bayes Network Classifier

not using ADTree

#attributes=5 #classindex=4

Network structure (nodes followed by parents)

outlook(3): play

temperature(1): play

humidity(1): play

windy(2): play

play(2):

LogScore Bayes: -39.49991749238267

LogScore BDeu: -46.697734525428125

LogScore MDL: -46.72727827843424

LogScore ENTROPY: -37.49057762478084

LogScore AIC: -44.49057762478084

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 8 57.1429 %

Incorrectly Classified Instances 6 42.8571 %

Kappa statistic -0.0244

Mean absolute error 0.4671

Root mean squared error 0.5221

Relative absolute error 98.0951 %

Root relative squared error 105.8276 %

Total Number of Instances 14

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.778 0.8 0.636 0.778 0.7 0.511 yes

0.2 0.222 0.333 0.2 0.25 0.511 no

Weighted Avg. 0.571 0.594 0.528 0.571 0.539 0.511

=== Confusion Matrix ===

a b <-- classified as

7 2 | a = yes

4 1 | b = no

2nd Run

weka.classifiers.bayes.BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.MultiNomialBMAEstimator -- -k2 -A 0.5

Correctly Classified Instances 9 64.2857 %

Incorrectly Classified Instances 5 35.7143 %

=== Confusion Matrix ===

a b <-- classified as

9 0 | a = yes

5 0 | b = no

3rd Run

Scheme: weka.classifiers.bayes.BayesNet -D -Q weka.classifiers.bayes.net.search.local.RepeatedHillClimber -- -U 10 -A 1 -P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5

Correctly Classified Instances 6 42.8571 %

Incorrectly Classified Instances 8 57.1429 %

=== Confusion Matrix ===

a b <-- classified as

6 3 | a = yes

5 0 | b = no

AODE

AODE classification averages over multiple "weaker", or less independent variable dependent, Naive Bayes functions which results in a higher accuracy than Naive Bayes. This came up with the most interesting classification results compared to the other algorithms used in this and the last assignment. The AODE run had a 50/50 split on the classification percentages and classified a 2 play-no correctly which was a rare occurrence in the other algorithms.

=== Run information ===

Scheme: weka.classifiers.bayes.AODE -F 1

Relation: weather.symbolic

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

The AODE Classifier

Class yes: Prior probability = 0.63

Class no: Prior probability = 0.38

Dataset: weather.symbolic

Instances: 14

Attributes: 5

Frequency limit for superParents: 1

Correction: laplace

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 7 50 %

Incorrectly Classified Instances 7 50 %

Kappa statistic -0.0426

Mean absolute error 0.4706

Root mean squared error 0.498

Relative absolute error 98.8198 %

Root relative squared error 100.941 %

Total Number of Instances 14

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.556 0.6 0.625 0.556 0.588 0.533 yes

0.4 0.444 0.333 0.4 0.364 0.533 no

Weighted Avg. 0.5 0.544 0.521 0.5 0.508 0.533

=== Confusion Matrix ===

a b <-- classified as

5 4 | a = yes

3 2 | b = no

ANN Classifers

Multilayer Perceptron

The Multilayer Perceptron uses back propagation to classify instances. The network can be built by hand using a GUI in Weka as shown below or generated from an algorithm or both. In this case the algorithm was used to build the network which then was explored using the GUI. This method had a high count of correctly classified play-no values which was unusual.

=== Run information ===

Scheme: weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Relation: weather.symbolic

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Sigmoid Node 0

Inputs Weights

Threshold -4.597967080790813

Node 2 2.433270074007239

Node 3 2.0546443732203774

Node 4 1.364159803860347

Node 5 2.6974766889493536

Node 6 3.908322709064356

Sigmoid Node 1

Inputs Weights

Threshold 4.601251960011152

Node 2 -2.4045226373071156

Node 3 -2.0532744956144127

Node 4 -1.379986429753948

Node 5 -2.756274547604192

Node 6 -3.877948258791871

Sigmoid Node 2

Inputs Weights

Threshold -0.1550798021501342

Attrib outlook=sunny -1.323464477913686

Attrib outlook=overcast 1.6602675280399888

Attrib outlook=rainy -0.3207802552865604

Attrib temperature=hot -0.2873122456981835

Attrib temperature=mild 1.181190360097958

Attrib temperature=cool -0.7853150475848826

Attrib humidity 2.808930687905

Attrib windy 1.9190213581350706

Sigmoid Node 3

Inputs Weights

Threshold -0.18031675012278034

Attrib outlook=sunny -1.1524514010228344

Attrib outlook=overcast 1.5760227701429683

Attrib outlook=rainy -0.32578400279223824

Attrib temperature=hot -0.2760307631136823

Attrib temperature=mild 1.0450876279343007

Attrib temperature=cool -0.6318819517738498

Attrib humidity 2.4504774603875408

Attrib windy 1.678251292646871

Sigmoid Node 4

Inputs Weights

Threshold -0.3554146745674961

Attrib outlook=sunny -0.46574052680925143

Attrib outlook=overcast 1.4382073898080827

Attrib outlook=rainy -0.6194183985830608

Attrib temperature=hot -0.0670794406887232

Attrib temperature=mild 0.6337484752708613

Attrib temperature=cool -0.20814280117719502

Attrib humidity 1.982466584793048

Attrib windy 0.9946423645131915

Sigmoid Node 5

Inputs Weights

Threshold -0.06888405078498452

Attrib outlook=sunny -1.3982064219096493

Attrib outlook=overcast 1.8084944112736516

Attrib outlook=rainy -0.31997269602762973

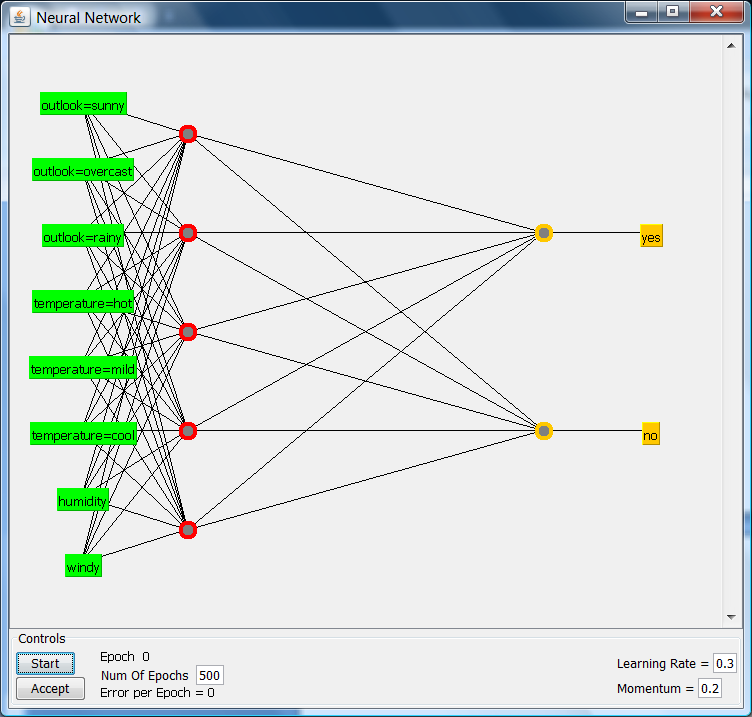
Attrib temperature=hot -0.3035821635771427

Attrib temperature=mild 1.2908528760310662

Attrib temperature=cool -0.8921466424329777

Attrib humidity 3.1090049574873424

Attrib windy 2.0747113212966872 ANN GUI

Sigmoid Node 6

Inputs Weights

Threshold 0.04399369934901554

Attrib outlook=sunny -1.80182134279014

Attrib outlook=overcast 2.2544547024444554

Attrib outlook=rainy -0.40095717506501327

Attrib temperature=hot -0.41558677311306397

Attrib temperature=mild 1.589170285947685

Attrib temperature=cool -1.2545441906677217

Attrib humidity 4.119310666164331

Attrib windy 2.740851006387263

Class yes

Input

Node 0

Class no

Input

Node 1

Time taken to build model: 0.36 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 10 71.4286 %

Incorrectly Classified Instances 4 28.5714 %

Kappa statistic 0.3778

Mean absolute error 0.287

Root mean squared error 0.5268

Relative absolute error 60.2616 %

Root relative squared error 106.7798 %

Total Number of Instances 14

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.778 0.4 0.778 0.778 0.778 0.778 yes

0.6 0.222 0.6 0.6 0.6 0.778 no

Weighted Avg. 0.714 0.337 0.714 0.714 0.714 0.778

=== Confusion Matrix ===

a b <-- classified as

7 2 | a = yes

2 3 | b = no

Voted Perceptron

The voted perceptron classifier is based on a perceptron classifier where a kernel function is used to model a non-linear classifier. The voted perceptron adds a voted function to calculate weights for the vectors. Multiple runs were done by changing the iterations and the seed parameters. The most interesting result from this was an equal correct classification percent 71.4 between runs 1 and 3 but the looking at the confusion matrix the classifications were different.

1st Run

Scheme: weka.classifiers.functions.VotedPerceptron -I 1 -E 1.0 -S 1 -M 10000

Relation: weather.symbolic

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

Test mode: 2-fold cross-validation

=== Classifier model (full training set) ===

VotedPerceptron: Number of perceptrons=4

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 10 71.4286 %

Incorrectly Classified Instances 4 28.5714 %

Kappa statistic 0.3171

Mean absolute error 0.3219

Root mean squared error 0.485

Relative absolute error 67.5911 %

Root relative squared error 98.6723 %

Total Number of Instances 14

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.889 0.6 0.727 0.889 0.8 0.589 yes

0.4 0.111 0.667 0.4 0.5 0.589 no

Weighted Avg. 0.714 0.425 0.706 0.714 0.693 0.589

=== Confusion Matrix ===

a b <-- classified as

8 1 | a = yes

3 2 | b = no

2nd Run

Scheme: weka.classifiers.functions.VotedPerceptron -I 10 -E 1.0 -S 1 -M 10000

Relation: weather.symbolic

Correctly Classified Instances 9 64.2857 %

Incorrectly Classified Instances 5 35.7143 %

=== Confusion Matrix ===

a b <-- classified as

8 1 | a = yes

4 1 | b = no

3rd Run

Scheme: weka.classifiers.functions.VotedPerceptron -I 1 -E 1.0 -S 10 -M 10000

Relation: weather.symbolic

Correctly Classified Instances 10 71.4286 %

Incorrectly Classified Instances 4 28.5714 %

=== Confusion Matrix ===

a b <-- classified as

9 0 | a = yes

4 1 | b = no

Winnow

The documentation is rather sparse on Winnow, but it in general it is a linear learning based classifier. Three runs are given below using the Winnow class. The number of iterations and seeds were changed for each run. The 1st run had the highest correct classification % with 1 iteration and 1 seed. The Winnow classifier became less accurate with subsequent runs as the iterations and seeds were increased. That being said the 1st run on default settings just predicted everything as yes which will always lead to a 64% correct classification with this dataset.

1st Run

Scheme: weka.classifiers.functions.Winnow -I 1 -A 2.0 -B 0.5 -H -1.0 -W 2.0 -S 1

Relation: weather.symbolic

Instances: 14

Attributes: 5

outlook

temperature

humidity

windy

play

Test mode: 2-fold cross-validation

=== Classifier model (full training set) ===

Winnow

Attribute weights

w0 8.0

w1 1.0

w2 2.0

w3 4.0

w4 2.0

w5 2.0

w6 1.0

w7 1.0

Cumulated mistake count: 7

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 9 64.2857 %

Incorrectly Classified Instances 5 35.7143 %

Kappa statistic 0

Mean absolute error 0.3571

Root mean squared error 0.5976

Relative absolute error 75 %

Root relative squared error 121.5772 %

Total Number of Instances 14

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

1 1 0.643 1 0.783 0.5 yes

0 0 0 0 0 0.5 no

Weighted Avg. 0.643 0.643 0.413 0.643 0.503 0.5

=== Confusion Matrix ===

a b <-- classified as

9 0 | a = yes

5 0 | b = no

2nd Run

Scheme: weka.classifiers.functions.Winnow -I 1 -A 2.0 -B 0.5 -H -1.0 -W 2.0 -S 10

Relation: weather.symbolic

Correctly Classified Instances 6 42.8571 %

Incorrectly Classified Instances 8 57.1429 %

a b <-- classified as

6 3 | a = yes

5 0 | b = no

3rd Run

Scheme: weka.classifiers.functions.Winnow -I 10 -A 2.0 -B 0.5 -H -1.0 -W 2.0 -S 1

Relation: weather.symbolic

Correctly Classified Instances 4 28.5714 %

Incorrectly Classified Instances 10 71.4286 %

a b <-- classified as

4 5 | a = yes

5 0 | b = no