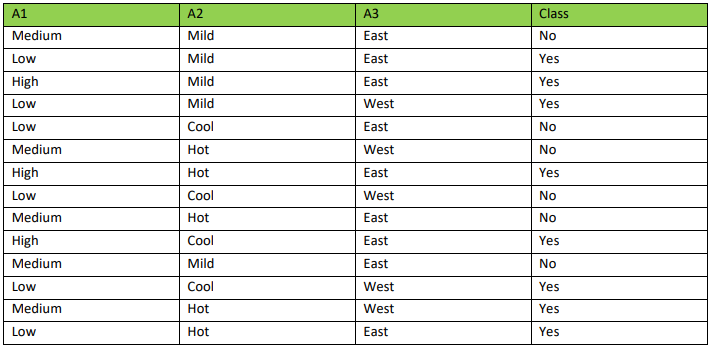
1. Naïve Bays Classification

Data



X = (A1 = Medium, A2 = Cool, A3 = East)

Class prior probabilities are:

P(C1 ) = P(class = “yes”) = 8/14 = 0.5714

P(C2 ) = P(class = “no”) = 6/14= 0.4285

Next, we compute P(X|Ci) for each class

P(A1 = Medium | class = Yes) = 1/8 = 0.125

P(A2=Cool | class = Yes) = 2/8 = 0.25

P(A3 = East | class = Yes) = 5/8 = 0.625

**P(X | class = Yes) = 0.125 \* 0.25 \* 0.625 = 0.0195**

P(A1 = Medium | class = No) = 4/6 = 0.667

P(A2=Cool | class = No) = 2/6 = 0.333

P(A3 = East | class = No) = 4/6 = 0.667

**P(X | class = No) = 0.667 \* 0.333 \* 0.667 = 0.148**

P(X|Ci )\*P(Ci )

P(X|C1 )\*P(C1 ) = P(X| class=Yes) \* P(class=Yes) = 0.0195 \* 0.5714 = 0.011

P(X|C2)\*P(C2 ) = P(X| class=No) \* P(class=No) = 0.148 \* 0.4285 = 0.063

Model predict that X belongs to Class = No

**2- Weka task output**

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes

Relation: breast-cancer

Instances: 286

Attributes: 10

age

menopause

tumor-size

inv-nodes

node-caps

deg-malig

breast

breast-quad

irradiat

Class

Test mode: 10-fold cross-validation

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

Naive Bayes Classifier

Class

Attribute no-recurrence-events recurrence-events

(0.7) (0.3)

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

age

10-19 1.0 1.0

20-29 2.0 1.0

30-39 22.0 16.0

40-49 64.0 28.0

50-59 72.0 26.0

60-69 41.0 18.0

70-79 6.0 2.0

80-89 1.0 1.0

90-99 1.0 1.0

[total] 210.0 94.0

menopause

lt40 6.0 3.0

ge40 95.0 36.0

premeno 103.0 49.0

[total] 204.0 88.0

tumor-size

0-4 8.0 2.0

5-9 5.0 1.0

10-14 28.0 2.0

15-19 24.0 8.0

20-24 35.0 17.0

25-29 37.0 19.0

30-34 36.0 26.0

35-39 13.0 8.0

40-44 17.0 7.0

45-49 3.0 2.0

50-54 6.0 4.0

55-59 1.0 1.0

[total] 213.0 97.0

inv-nodes

0-2 168.0 47.0

3-5 20.0 18.0

6-8 8.0 11.0

9-11 5.0 7.0

12-14 2.0 3.0

15-17 4.0 4.0

18-20 1.0 1.0

21-23 1.0 1.0

24-26 1.0 2.0

27-29 1.0 1.0

30-32 1.0 1.0

33-35 1.0 1.0

36-39 1.0 1.0

[total] 214.0 98.0

node-caps

yes 26.0 32.0

no 172.0 52.0

[total] 198.0 84.0

deg-malig

1 60.0 13.0

2 103.0 29.0

3 41.0 46.0

[total] 204.0 88.0

breast

left 104.0 50.0

right 99.0 37.0

[total] 203.0 87.0

breast-quad

left\_up 72.0 27.0

left\_low 76.0 36.0

right\_up 21.0 14.0

right\_low 19.0 7.0

central 18.0 5.0

[total] 206.0 89.0

irradiat

yes 38.0 32.0

no 165.0 55.0

[total] 203.0 87.0

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 205 71.6783 %

Incorrectly Classified Instances 81 28.3217 %

Kappa statistic 0.2857

Mean absolute error 0.3272

Root mean squared error 0.4534

Relative absolute error 78.2086 %

Root relative squared error 99.1872 %

Total Number of Instances 286

=== Detailed Accuracy By Class ===

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

0.836 0.565 0.778 0.836 0.806 0.288 0.701 0.837 no-recurrence-events

0.435 0.164 0.529 0.435 0.477 0.288 0.701 0.514 recurrence-events

Weighted Avg. 0.717 0.446 0.704 0.717 0.708 0.288 0.701 0.741

=== Confusion Matrix ===

a b <-- classified as

168 33 | a = no-recurrence-events

48 37 | b = recurrence-events