**Practical Machine Learning**

**(A case study demonstrating the applications of supervised learning – logistic regression, fisher’s linear discriminant analysis, k-nearest neighbours, artificial neural network, and unsupervised learning – k-means clustering for some botanical data.)**

1. **Practical Machine Learning (5 Cases)**
   1. Learning a 3-class flower species classifier based on some attributes of a flower using logistic regression
   2. Learning a 3-class flower species classifier based on some attributes of a flower using fisher’s linear discriminant analysis
   3. Learning a 3-class flower species classifier based on some attributes of a flower using k-nearest neighbors
   4. Learning a 3-class flower species classifier based on some attributes of a flower using artificial neural network
   5. Categorizing flowers into 3 categories based on some attributes of a flower using k-means clustering

**Data Source**-Fisher R.A. (1936) the use of multiple measurements in taxonomic problems. Annals of Eugenics, 7, Part II, 179-188

**Case 1** Consider the data set flowerspecies.xlsx which has 150 observations on four features namely sepal length, sepal width, petal length, petal width for 3 different species of flowers (50 observations for each species).Learn a 3 class classification hypothesis on given data, which predicts the species type based on the given 4 features using **logistic regression**.

Obtain the estimated classes for all the observations in the given data. Construct the observed vs predicted classification table and calculate the class wise and overall percentage of misclassification.

**Nominal Regression**

**Multinomial Logistic Regression** is the linear regression analysis to conduct when the dependent variable is nominal with more than two levels.  Thus it is an extension of logistic regression, which analyses dichotomous (binary) dependents.  Multinomial logistic regression is used to model nominal outcome variables, in which the log odds of the outcomes are modelled as a linear combination of the predictor variables.

Like all linear regressions, the multinomial regression is a predictive analysis.  Multinomial regression is used to describe data and to explain the relationship between one dependent nominal variable and one or more continuous-level (interval or ratio scale) independent variables

We set the following hypothesis:

Ho: Full model has better fit than the null model.

H1: Full model does not have a better fit than a null model.

**Model** - This indicates the parameters of the model for which the model fit is calculated.  "Intercept Only" describes a model that does not control for any predictor variables and simply fits an intercept to predict the outcome variable. "Final" describes a model that includes the specified predictor variables and has been arrived at through an iterative process that maximizes the log likelihood of the outcomes seen in the outcome variable. By including the predictor variables and maximizing the log likelihood of the outcomes seen in the data, the "Final" model should improve upon the "Intercept Only" model.  This can be seen in the differences in the -2(Log Likelihood) values associated with the models.

3)**df** - This indicates the degrees of freedom of the chi-square distribution used to test the LR Chi-Square statistic and is defined by the number of predictors in the model.

Here, we see model fit is significant χ² (8) = 317.685, *p*< .05, which indicates our full model predicts significantly better, or more accurately, than the null model. We want the *p-*value to be *less than* your established cutoff (generally 0.05) to indicate good fit.

| Case Processing Summary | | | |
| --- | --- | --- | --- |
|  |  | N | Marginal Percentage |
| X5 | setosa | 50 | 33.3% |
| versicolor | 50 | 33.3% |
| virginica | 50 | 33.3% |
|  | Valid | 150 | 100.0% |
| Missing | 0 |  |
| Total | 150 |  |
| Subpopulation | 149a |  |
| a. The dependent variable has only one value observed in 149 (100.0%) subpopulations. | | | |

| Model Fitting Information | | | | |
| --- | --- | --- | --- | --- |
| Model | Model Fitting Criteria | Likelihood Ratio Tests | | |
| -2 Log Likelihood | Chi-Square | df | Sig. |
| Intercept Only | 329.584 |  |  |  |
| Final | 11.899 | 317.685 | 8 | .000 |

1)**-2(Log Likelihood)** - This is the product of -2 and the log likelihoods of the null model and fitted "final" model. The likelihood of the model is used to test of whether all predictors' regression coefficients in the model are simultaneously zero.

2)**Chi-Square** - This is the Likelihood Ratio (LR) Chi-Square test that at least one of the predictors' regression coefficient is not equal to zero in the model. The LR Chi-Square statistic can be calculated by  -2\*L(null model) - (-2\*L(fitted model)) = 329.584 – 11.899 = 317.685, where *L(null model)* is from the log likelihood with just the response variable in the model (Intercept Only) and *L(fitted model)* is the log likelihood from the final iteration (assuming the model converged) with all the parameters.

| **Likelihood Ratio Tests** | | | | |
| --- | --- | --- | --- | --- |
| Effect | Model Fitting Criteria | Likelihood Ratio Tests | | |
| -2 Log Likelihood of Reduced Model | Chi-Square | Df | Sig.  The statistics in the Likelihood Ratio Tests table are the same types as those reported for the null and full models above in the Model Fitting Information table. Here however, each element of the model is being compared to the full model in such a way as to allow the research to determine if it (each element) should be included in the full model.  In other words, does each element (predictor) contribute meaningfully to the full effect? For instance, we see that the X1, X3 predictor displays a non-significant (*p*= .505, p=0.166) chi-square which indicates X1, X3 could be dropped from the model and the overall fit would NOT be significantly reduced. To be clear, if the *p-*value is *less than* your established cutoff (generally 0.05) for a predictor then that predictor contributes significantly to the full (final) model. |
| Intercept | 21.680 | 9.781 | 2 | .008 |
| X1 | 13.266 | 1.367 | 2 | .505 |
| X2 | 15.492 | 3.594 | 2 | .166 |
| X3 | 25.902 | 14.003 | 2 | .001 |
| X4 | 23.772 | 11.873 | 2 | .003 |
| The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0. | | | | |

**Crosstabs**

| X5 \* Predicted Response Category Crosstabulation | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  | Predicted Response Category | | |  |
|  |  |  | setosa | versicolor | virginica | Total |
| X5 | setosa | Count | 50 | 0 | 0 | 50 |
| % within X5 | 100.0% | .0% | .0% | 100.0% |
| % of Total | 33.3% | .0% | .0% | 33.3% |
| versicolor | Count | 0 | 49 | 1 | 50 |
| % within X5 | .0% | 98.0% | 2.0% | 100.0% |
| % of Total | .0% | 32.7% | .7% | 33.3% |
| virginica | Count | 0 | 1 | 49 | 50 |
| % within X5 | .0% | **2.0%** | 98.0% | 100.0% |
| % of Total | .0% | .7% | 32.7% | 33.3% |
|  | Total | Count | 50 | 50 | 50 | 150 |
| % within X5 | 33.3% | 33.3% | 33.3% | 100.0% |
| % of Total | 33.3% | 33.3% | 33.3% | 100.0% |

PRE1 PCP1

The Classification Table (above) shows how well our full model correctly classifies cases. A perfect model would show only values on the diagonal--correctly classifying all cases. Adding across the rows represents the number of cases in each category in the actual data and adding down the columns represents the number of cases in each category as classified by the full model. We see that **setosa flower is classified with 100 % accuracy, versicolor flower is classified with 98% accuracy and virginica flower is classified with 98 % accuracy** .The key piece of information is the overall percentage in the lower rightcorner which shows our model (with all predictors & the constant) is **100%** accurate; which is excellent.

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

virginica 1.00

virginica 1.00

virginica 1.00

virginica 1.00

virginica 1.00

virginica 1.00

virginica 1.00

virginica 1.00

virginica 1.00

virginica 0.98

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

virginica 0.87

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

virginica 1.00

virginica 1.00

virginica 1.00

virginica 1.00

virginica 1.00

virginica 1.00

virginica 0.89

virginica 1.00

virginica 1.00

virginica 1.00

virginica 0.99

virginica 1.00

virginica 1.00

virginica 1.00

virginica 1.00

virginica 1.00

virginica 1.00

virginica 1.00

virginica 1.00

virginica 0.92

virginica 1.00

virginica 1.00

virginica 1.00

virginica 0.95

virginica 1.00

virginica 1.00

virginica 0.82

virginica 0.80

virginica 1.00

virginica 0.97

virginica 1.00

virginica 1.00

virginica 1.00

versicolor 0.80

virginica 0.97

virginica 1.00

virginica 1.00

virginica 1.00

virginica 0.67

virginica 1.00

From the observed vs predicted classification table we have the class wise percentage of misclassification as 0% for setosa,2% for versicolor and 2% for virginica as well. The overallpercentage of misclassification is 2%

Predicted classes and predicted probabilities are shown in the table.

Estimated classes for all the observations in the given data are also shown.

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

setosa 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 0.94

versicolor 1.00

versicolor 0.60

versicolor 1.00

versicolor 0.78

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 1.00

versicolor 0.72

**Case 2**- Consider the data set flowerspecies.xlsx which has 150 observations on four features namely sepal length, sepal width, petal length, petal width for 3 different species of flowers (50 observations for each species).Learn a 3 class classification hypothesis on given data, which predicts the species type based on the given 4 features using **Fishers Discriminant Analysis(FDA).**

Obtain the estimated classes for all the observations in the given data. Construct the observed vs predicted classification table and calculate the class wise and overall percentage of misclassification.

| Structure Matrix | | |
| --- | --- | --- |
|  | | |
|  | Function | |
|  | 1 | 2 |
| X3 | .706\* | .168 |
| X2 | -.119 | .864\* |
| X4 | .633 | .737\* |
| X1 | .223 | .311\* |
| Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions  Variables ordered by absolute size of correlation within function. | | |
| \*. Largest absolute correlation between each variable and any discriminant function | | |
|  | | |
|  | | |

Classification Statistics

| Prior Probabilities for Groups | | | |
| --- | --- | --- | --- |
| numeric |  | Cases Used in Analysis | |
| Prior | Unweighted | Weighted |
| 1.00 | .333 | 50 | 50.000 |
| 2.00 | .333 | 50 | 50.000 |
| 3.00 | .333 | 50 | 50.000 |
| Total | 1.000 | 150 | 150.000 |

| Functions at Group Centroids | | |
| --- | --- | --- |
| numeric | Function | |
| 1 | 2 |
| 1.00 | -7.608 | .215 |
| 2.00 | 1.825 | -.728 |
| 3.00 | 5.783 | .513 |
| Unstandardized canonical discriminant functions evaluated at group means | | |

From the structure matrix we infer that -By identifying the largest absolute correlations associated with each discriminant function the researcher gains insight into how to name each function. It provides a way to study usefulness of each variable in the discriminant function.X2 and X4 are more correlated with function 2

Functions of group centroids indicates the average discriminant score for subjects in the 2 groups. More specifically the discriminant score for each group when the variable means are entered into the discriminant equation. Functions of group centroids are mean discriminant scores for each of the dependent variable categories for each of the discriminant functions in the discriminant analysis. The closer the means the more the errors of classification. As the means of both the groups are well apart, so the discriminant function is clearly discriminating.

The table above gives the prior probabilities as 0.33.

Crosstabs

 A cross-tabulation is a table that depicts the number of times each of the possible category combinations occurred in the sample data.

| numeric \* Predicted Group for Analysis 1 Crosstabulation | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  | Predicted Group for Analysis 1 | | |  |
|  |  |  | 1.00 | 2.00 | 3.00 | Total |
| numeric | 1.00 | Count | 50 | 0 | 0 | 50 |
| % within numeric | **100.0%** | .0% | .0% | 100.0% |
| % of Total | 33.3% | .0% | .0% | 33.3% |
| 2.00 | Count | 0 | 48 | 2 | 50 |
| % within numeric | .0% | **96.0%** | 4.0% | 100.0% |
| % of Total | .0% | 32.0% | 1.3% | 33.3% |
| 3.00 | Count | 0 | 1 | 49 | 50 |
| % within numeric | .0% | 2.0% | **98.0%** | 100.0% |
| % of Total | .0% | .7% | 32.7% | 33.3% |
|  | Total | Count | 50 | 49 | 51 | 150 |
| % within numeric | 33.3% | 32.7% | 34.0% | 100.0% |
| % of Total | 33.3% | 32.7% | 34.0% | 100.0% |

* The Classification Table (above) shows how well our full model correctly classifies cases. A perfect model would show only values on the diagonal--correctly classifying all cases. Adding across the rows represents the number of cases in each category in the actual data and adding down the columns represents the number of cases in each category as classified by the full model.
* We see that **setosa flower is classified with 100 % accuracy, versicolor flower is classified with 96% accuracy and virginica flower is classified with 98 % accuracy** .The key piece of information is the overall percentage in the lower rightcorner which shows our model (with all predictors & the constant) is **100%** accurate; which is excellent.
* From the observed vs predicted classification table we have the class wise percentage of misclassification as 0% for setosa,4% for versicolor and 2% for virginica as well. The overall percentage of misclassification is 3%Predicted classes and predicted probabilities are shown in the table. Estimated classes for all the observations in the given data are also shown

74 versicolor Virginica

75 versicolor Virginica

76 versicolor Virginica

77 versicolor Virginica

78 versicolor Setosa

79 versicolor Virginica

80 versicolor Virginica

81 versicolor Virginica

82 versicolor Virginica

83 versicolor Virginica

84 versicolor Virginica

85 versicolor Virginica

86 versicolor Virginica

87 versicolor Virginica

88 versicolor Virginica

89 versicolor Virginica

90 versicolor Virginica

91 versicolor Virginica

92 versicolor Virginica

93 versicolor Virginica

94 versicolor Virginica

95 versicolor Virginica

96 versicolor Virginica

97 versicolor Virginica

98 versicolor Virginica

99 versicolor Virginica

100 versicolor Virginica

101 virginica Setosa

102 virginica Virginica

X5 Predicted

1 setosa Versicolor

2 setosa Versicolor

3 setosa Versicolor

4 setosa Versicolor

5 setosa Versicolor

6 setosa Versicolor

7 setosa Versicolor

8 setosa Versicolor

9 setosa Versicolor

10 setosa Versicolor

11 setosa Versicolor

12 setosa Versicolor

13 setosa Versicolor

14 setosa Versicolor

15 setosa Versicolor

16 setosa Versicolor

17 setosa Versicolor

18 setosa Versicolor

19 setosa Versicolor

20 setosa Versicolor

21 setosa Versicolor

22 setosa Versicolor

23 setosa Versicolor

24 setosa Versicolor

25 setosa Versicolor

26 setosa Versicolor

27 setosa Versicolor

28 setosa Versicolor

29 setosa Versicolor

30 setosa Versicolor

31 setosa Versicolor

32 setosa Versicolor

33 setosa Versicolor

| 103 | virginica | Setosa |
| --- | --- | --- |
| 104 | virginica | Setosa |
| 105 | virginica | Setosa |
| 106 | virginica | Setosa |
| 107 | virginica | Virginica |
| 108 | virginica | Setosa |
| 109 | virginica | Setosa |
| 110 | virginica | Setosa |
| 111 | virginica | Setosa |
| 112 | virginica | Setosa |
| 113 | virginica | Setosa |
| 114 | virginica | Virginica |
| 115 | virginica | Virginica |
| 116 | virginica | Setosa |
| 117 | virginica | Setosa |
| 118 | virginica | Setosa |
| 119 | virginica | Setosa |
| 120 | virginica | Virginica |
| 121 | virginica | Setosa |
| 122 | virginica | Virginica |
| 123 | virginica | Setosa |
| 124 | virginica | Virginica |
| 125 | virginica | Setosa |
| 126 | virginica | Setosa |
| 127 | virginica | Virginica |
| 128 | virginica | Virginica |
| 129 | virginica | Setosa |
| 130 | virginica | Setosa |
| 131 | virginica | Setosa |
| 132 | virginica | Setosa |
| 133 | virginica | Setosa |
| 134 | virginica | Virginica |
| 135 | virginica | Setosa |
| 136 | virginica | Setosa |
| 137 | virginica | Setosa |
| 138 | virginica | Setosa |
| 139 | virginica | Virginica |
| 140 | virginica | Setosa |
| 141 | virginica | Setosa |
| 142 | virginica | Setosa |
| 143 | virginica | Virginica |
| 144 | virginica | Setosa |
| 145 | virginica | Setosa |
| 146 | virginica | Setosa |
| 147 | virginica | Virginica |

1virginica Setosa

149 virginica Setosa150 virginica Virginica44 virginica Setosa

145 virginica Setosa

146 virginica Setosa

147 virginica Virginica

148 virginica Setosa

149 virginica Setosa

150 virginica Virginica

Total N 150 150

| 145 | virginica | Setosa |
| --- | --- | --- |
| 146 | virginica | Setosa |
| 147 | virginica | Virginica |

41 virginica Setosa

142 virginica Setosa

143 virginica Virginica

144 virginica Setosa

145 virginica Setosa

146 virginica Setosa

147 virginica Virginica

148 virginica Setosa

149 virginica Setosa

150 virginica Virginica

34 setosa Versicolor

35 setosa Versicolor

36 setosa Versicolor

37 setosa Versicolor

38 setosa Versicolor

39 setosa Versicolor

40 setosa Versicolor

41 setosa Versicolor

42 setosa Versicolor

43 setosa Versicolor

44 setosa Versicolor

45 setosa Versicolor

46 setosa Versicolor

47 setosa Versicolor

48 setosa Versicolor

49 setosa Versicolor

50 setosa Versicolor

51 versicolor Virginica

52 versicolor Virginica

53 versicolor Setosa

54 versicolor Virginica

55 versicolor Virginica

56 versicolor Virginica

57 versicolor Virginica

58 versicolor Virginica

59 versicolor Virginica

60 versicolor Virginica

61 versicolor Virginica

62 versicolor Virginica

63 versicolor Virginica

64 versicolor Virginica

65 versicolor Virginica

66 versicolor Virginica

67 versicolor Virginica

68 versicolor Virginica

69 versicolor Virginica

70 versicolor Virginica

71 versicolor Virginica

72 versicolor Virginica

73 versicolor Virginica

148 virginica Setosa

149 virginica Setosa

150 virginica Virginica

**Case 3** Consider the data set flowerspecies.xlsx which has 150 observations on four features namely sepal length, sepal width, petal length, petal width for 3 different species of flowers (50 observations for each species).Learn a 3 class classification hypothesis on given data, which predicts the species type based on the given 4 features using **K Nearest neighbours (K-NN).**

Obtain the estimated classes for all the observations in the given data. Construct the observed vs predicted classification table and calculate the class wise and overall percentage of misclassification

Nearest Neighbor Analysis is a method for classifying cases based on their similarity to other cases. In machine learning, it was developed as a way to recognize patterns of data without requiring an exact match to any stored patterns, or cases. Similar cases are near each other and dissimilar cases are distant from each other. Thus, the distance between two cases is a measure of their dissimilarity.

Cases that are near each other are said to be “neighbors.” When a new case (holdout) is presented, its distance from each of the cases in the model is computed. The classifications of the most similar cases – the nearest neighbors – are tallied and the new case is placed into the category that contains the greatest number of nearest neighbors.

You can specify the number of nearest neighbors to examine; this value is called k.

Nearest neighbor analysis can also be used to compute values for a continuous target. In this situation, the average or median target value of the nearest neighbors is used to obtain the predicted value for the new case.

**Nearest Neighbour Analysis**

* Displays the case processing summary table, which summarizes the number of cases included and excluded in the analysis, in total and by training and holdout samples.

We have 100% training data and 0% holdout data.

| **Case Processing Summary** | | | |
| --- | --- | --- | --- |
|  |  | N | Percent |
| Sample | Training | 150 | 100.0% |
| Holdout | 0 | .0% |
| Valid | | 150 | 100.0% |
| Excluded | | 0 |  |
| Total | | 150 |  |



This chart classifies our prototypes (in red) based on their nearest neighbours in accordance with three predictor variables: setosa, virginica and versicolor .This is a 3D interactive chart that identifies the 3 nearest neighbours (K = 3) to our prototypes.

**Crosstabs**

| **X5 \* Predicted Value for X5 Crosstabulation**  The predicted data is shown in cross tabulation table:  50 cases are classified as setosa  49 out of 50 cases are classified as versicolor.  46 out of 50 cases are classified as virginica. | | | | | |
| --- | --- | --- | --- | --- | --- |
| Count | | | | | |
|  |  | Predicted Value for X5 | | | Total |
|  |  | setosa | versicolor | virginica |
| X5 | setosa | 50 | 0 | 0 | 50 |
| versicolor | 0 | 49 | 1 | 50 |
| virginica | 0 | 4 | 46 | 50 |
| Total | | 50 | 53 | 47 | 150 |

KNN(pred value) 1 2 3

setosa 0.867 0.067 0.067

versicolor 0.067 0.867 0.067

versicolor 0.067 0.733 0.200

versicolor 0.067 0.867 0.067

versicolor 0.067 0.800 0.133

versicolor 0.067 0.867 0.067

versicolor 0.067 0.667 0.267

versicolor 0.067 0.867 0.067

versicolor 0.067 0.800 0.133

versicolor 0.067 0.867 0.067

versicolor 0.067 0.867 0.067

versicolor 0.067 0.867 0.067

versicolor 0.067 0.867 0.067

versicolor 0.067 0.800 0.133

versicolor 0.067 0.867 0.067

versicolor 0.067 0.867 0.067

versicolor 0.067 0.800 0.133

versicolor 0.067 0.867 0.067

versicolor 0.067 0.600 0.333

versicolor 0.067 0.867 0.067

versicolor 0.067 0.467 0.467

versicolor 0.067 0.867 0.067

versicolor 0.067 0.467 0.467

versicolor 0.067 0.800 0.133

versicolor 0.067 0.867 0.067

versicolor 0.067 0.867 0.067

versicolor 0.067 0.800 0.133

versicolor 0.067 0.467 0.467

versicolor 0.067 0.733 0.200

versicolor 0.067 0.867 0.067

versicolor 0.067 0.867 0.067

versicolor 0.067 0.867 0.067

versicolor 0.067 0.867 0.067

virginica 0.067 0.267 0.667

versicolor 0.067 0.867 0.067

versicolor 0.067 0.667 0.267

versicolor 0.067 0.867 0.067

versicolor 0.067 0.800 0.133

versicolor 0.067 0.867 0.067

versicolor 0.067 0.867 0.067

versicolor 0.067 0.867 0.067

versicolor 0.067 0.800 0.133

versicolor 0.067 0.867 0.067

versicolor 0.067 0.867 0.067

versicolor 0.067 0.867 0.067

versicolor 0.067 0.867 0.067

versicolor 0.067 0.867 0.067

versicolor 0.067 0.867 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

setosa 0.867 0.067 0.067

versicolor 0.067 0.800 0.133

Case summaries have been shown in the above table.

The predicted category probabilities are also shown in the table.

versicolor 0.067 0.867 0.067

versicolor 0.067 0.867 0.067

virginica 0.067 0.067 0.867

virginica 0.067 0.133 0.800

virginica 0.067 0.067 0.867

virginica 0.067 0.133 0.800

virginica 0.067 0.067 0.867

virginica 0.067 0.067 0.867

versicolor 0.067 0.600 0.333

virginica 0.067 0.133 0.800

virginica 0.067 0.133 0.800

virginica 0.067 0.067 0.867

virginica 0.067 0.133 0.800

virginica 0.067 0.067 0.867

virginica 0.067 0.067 0.867

virginica 0.067 0.133 0.800

virginica 0.067 0.067 0.867

virginica 0.067 0.067 0.867

virginica 0.067 0.133 0.800

virginica 0.067 0.067 0.867

virginica 0.067 0.067 0.867

versicolor 0.067 0.533 0.400

virginica 0.067 0.067 0.867

virginica 0.067 0.267 0.667

virginica 0.067 0.067 0.867

virginica 0.067 0.267 0.667

virginica 0.067 0.067 0.867

virginica 0.067 0.133 0.800

virginica 0.067 0.200 0.733

virginica 0.067 0.267 0.667

virginica 0.067 0.067 0.867

virginica 0.067 0.400 0.533

virginica 0.067 0.067 0.867

virginica 0.067 0.067 0.867

virginica 0.067 0.067 0.867

versicolor 0.067 0.667 0.267

versicolor 0.067 0.467 0.467

virginica 0.067 0.067 0.867

virginica 0.067 0.067 0.867

virginica 0.067 0.133 0.800

virginica 0.067 0.333 0.600

virginica 0.067 0.067 0.867

virginica 0.067 0.067 0.867

virginica 0.067 0.067 0.867

virginica 0.067 0.133 0.800

virginica 0.067 0.067 0.867

virginica 0.067 0.067 0.867

virginica 0.067 0.067 0.867

virginica 0.067 0.200 0.733

virginica 0.067 0.133 0.800

virginica 0.067 0.067 0.867

virginica 0.067 0.267 0.667

**Case 4** Consider the data set flowerspecies.xlsx which has 150 observations on four features namely sepal length, sepal width, petal length, petal width for 3 different species of flowers (50 observations for each species).Learn a 3 class classification hypothesis on given data, which predicts the species type based on the given 4 features using **Artificial Neural network (ANN).**

Obtain the estimated classes for all the observations in the given data. Construct the observed vs predicted classification table and calculate the class wise and overall percentage of misclassification.

**Neural Networks**

A computational Neural network is a set of nonlinear data modelling tools consisting of input and output layers and 1 or 2 hidden layers. The connections between neurons in each layer have associated weights which are iteratively adjusted by the training algorithm to minimize error and provide accurate predictions.

Neural networks are the preferred tool for many predictive data mining applications because of their power, flexibility, and ease of use. Predictive neural networks are particularly useful in applications where the underlying process is complex, such as:

Forecasting consumer demand to streamline production and delivery costs.

Predicting the probability of response to direct mail marketing to determine which households on a mailing list should be sent an offer.

Scoring an applicant to determine the risk of extending credit to the applicant.

Detecting fraudulent transactions in an insurance claims database.

Neural networks used in predictive applications, such as the multilayer perceptron (MLP) and radial basis function (RBF) networks, are supervised in the sense that the model-predicted results can be compared against known values of the target variables. The Neural Networks option allows you to fit MLP and RBF networks and save the resulting models for scoring.

**What Is a Neural Network?**

The term neural network applies to a loosely related family of models, characterized by a large parameter space and flexible structure, descending from studies of brain functioning. As the family grew, most of the new models were designed for non biological applications, though much of the associated terminology reflects its origin. Specific definitions of neural networks are as varied as the fields in which they are used.

A neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. It resembles the brain in two respects:

Knowledge is acquired by the network through a learning process.

Interneuron connection strengths known as synaptic weights are used to store the knowledge.

**Network Performance.** Displays results used to determine whether the model is "good". *Note*: Charts in this group are based on the combined training and testing samples or only on the training sample if there is no testing sample.

• **Classification results.** Displays a classification table for each categorical dependent variable by partition and overall. Each table gives the number of cases classified correctly and incorrectly for each dependent variable category. The percentage of the total cases that were correctly classified is also reported.

| **Case Processing Summary**   * Displays the case processing summary table, which summarizes the number of cases included and excluded in the analysis, in total and by training, testing, and holdout samples. * We have 66.7% training data and 33.3% testing data. | | | |
| --- | --- | --- | --- |
|  |  | N | Percent |
| Sample | Training | 100 | 66.7% |
| Testing | 50 | 33.3% |
| Valid | | 150 | 100.0% |
| Excluded | | 0 |  |
| Total | | 150 |  |

Note that the covariates are standardised so that none are numerically dominant. As a first trial a small hidden layer is adopted, this may be increased in succeeding applications of the procedure.

| **Network Information** | | | |
| --- | --- | --- | --- |
| Input Layer | Covariates | 1 | X1 |
| 2 | X2 |
| 3 | X3 |
| 4 | X4 |
|  | Number of Unitsa | 4 |
| Rescaling Method for Covariates | Standardized |
| Hidden Layer(s) |  | Number of Hidden Layers | 1 |
| Number of Units in Hidden Layer 1a | 5 |
| Activation Function | Hyperbolic tangent |
| Output Layer | Dependent Variables | 1 | X5 |
| Number of Units | | 3 |
| Activation Function | | Softmax |
| Error Function | | Cross-entropy |
| a. Excluding the bias unit | | | |

**Network Structure.** Displays summary information about the neural network.

• **Description.** Displays information about the neural network, including the dependent variables, number of input and output units, number of hidden layers and units, and activation functions.

•

These algorithms are known as supervised networks in the sense that the model predicted results can be compared against known values of the target variables.

One of the primary advantages of neural networks when compared to classical statistical techniques is their flexibility, and lack of distributional assumptions. Neural networks can be used to predict both categorical and continuous outcomes.

Neural networks with optimal architecture obtained after algorithm Training processTable 1

* A neural network works by taking the values of predictor or input fields and feeding them into the algorithm as an input layer.
* The input layer is used to create a hidden layer containing unseen nodes or units where each node is some function of the input fields.
* The output layer contains the responses or predictions. The network is continually rebuilt or refined so that the synaptic weights in the nodes correctly predict the outcome.

**Diagram.** Displays the network diagram as a non-editable chart. Note that as the number of covariates and factor levels increases, the diagram becomes more difficult to interpret.

• **Synaptic weights.** Displays the coefficient estimates that show the relationship between the units in a given layer to the units in the following layer. The synaptic weights are based on the training sample even if the active dataset is partitioned into training, testing, and holdout data. Note that the number of synaptic weights can become rather large and that these weights are generally not used forinterpreting network results.

| **Model Summary** | | |
| --- | --- | --- |
| Training | Cross Entropy Error | 5.852 |
| Percent Incorrect Predictions | 2.0% |
| Stopping Rule Used | 1 consecutive step(s) with no decrease in errora |
| Training Time | 0:00:00.031 |
| Testing | Cross Entropy Error | 4.104 |
| Percent Incorrect Predictions | 6.0% |
| Dependent Variable: X5 | | |
| a. Error computations are based on the testing sample. | | |
| The Classification Table (below) shows how well our full model correctly classifies cases. A perfect model would show only values on the diagonal--correctly classifying all cases. Adding across the rows represents the number of cases in each category in the actual data and adding down the columns represents the number of cases in each category as classified by the full model. | | |

• **Model summary.** Displays a summary of the neural network results by partition and overall, including the error, the relative error or percentage of incorrect predictions, the stopping rule used to stop training, and the training time.

* The error is the sum-of-squares error when the identity, sigmoid, or hyperbolic tangent activation function is applied to the output layer. It is the cross-entropy error when the softmax activation function is applied to the output layer. Relative errors or percentages of incorrect predictions are displayed depending on the dependent variable measurement levels.
* If any dependent variable has scale measurement level, then the average overall relative error (relative to the mean model) is displayed. If all dependent variables are categorical, then the average percentage of incorrect predictions is displayed. Relative errors or percentages of incorrect predictions are also displayed for individual dependent variables.
* **Gradient Descent**: This method must be used with online or mini batch training .It can also be used with batch training.
* The training type like batch determines how the network processes the records. Batch training directly minimizes the total error. It updates the synaptic weights only after passing all training data records that is it uses information from all records in the training data set.

| **Classification** | | | | | |
| --- | --- | --- | --- | --- | --- |
| Sample | Observed | Predicted | | | |
| setosa | versicolor | virginica | Percent Correct |
| Training | setosa | 36 | 0 | 0 | 100.0% |
| versicolor | 0 | 33 | 1 | 97.1% |
| virginica | 0 | 1 | 29 | 96.7% |
| Overall Percent | 36.0% | 34.0% | 30.0% | 98.0% |
| Testing | setosa | 14 | 0 | 0 | 100.0% |
| versicolor | 0 | 14 | 2 | 87.5% |
| virginica | 0 | 1 | 19 | 95.0% |
| Overall Percent | 28.0% | 30.0% | 42.0% | 94.0% |
| Dependent Variable: X5 | | | | | |

For the testing data

* 100% of the cases under setosa
* 87.5% of the cases under versicolor.
* 95% of the units under virginica.

Are correctly classified

So overall 94% of the total cases were correctly classified.

For training data the percentage of correct classification is more.

For the training data

* 100% of the cases under setosa
* 97.1% of the cases under versicolor.
* 96.7% of the units under virginica.

are correctly classified

So overall 98% of the total cases were correctly classified.

| **Case Summariesa** | | |
| --- | --- | --- |
|  |  | Predicted Value for X5 |  |
| 1 | | setosa | 58 | versicolor | 116 | virginica |
| 2 | | setosa | 59 | versicolor | 117 | virginica |
| 3 | | setosa | 60 | versicolor | 118 | virginica |
| 4 | | setosa | 61 | versicolor | 119 | virginica |
| 5 | | setosa | 62 | versicolor | 120 | virginica |
| 6 | | setosa | 63 | versicolor | 121 | virginica |
| 7 | | setosa | 64 | versicolor | 122 | virginica |
| 8 | | setosa | 65 | versicolor | 123 | virginica |
| 9 | | setosa | 66 | versicolor | 124 | virginica |
| 10 | | setosa | 67 | versicolor | 125 | virginica |
| 11 | | setosa | 68 | versicolor | 126 | virginica |
| 12 | | setosa | 69 | versicolor | 127 | virginica |
| 13 | | setosa | 70 | versicolor | 128 | virginica |
| 14 | | setosa | 71 | versicolor | 129 | virginica |
| 15 | | setosa | 72 | versicolor | 130 | virginica |
| 16 | | setosa | 73 | virginica | 131 | virginica |
| 17 | | setosa | 74 | versicolor | 132 | virginica |
| 18 | | setosa | 75 | versicolor | 133 | virginica |
| 19 | | setosa | 76 | versicolor | 134 | versicolor |
| 20 | | setosa | 77 | versicolor | 135 | virginica |
| 21 | | setosa | 78 | virginica | 136 | virginica |
| 22 | | setosa | 79 | versicolor | 137 | virginica |
| 23 | | setosa | 80 | versicolor | 138 | virginica |
| 24 | | setosa | 81 | versicolor | 139 | virginica |
| 25 | | setosa | 82 | versicolor | 140 | virginica |
| 26 | | setosa | 83 | versicolor | 141 | virginica |
| 27 | | setosa | 84 | virginica | 142 | virginica |
| 28 | | setosa | 85 | versicolor | 143 | virginica |
| 29 | | setosa | 86 | versicolor | 144 | virginica |
| 30 | | setosa | 87 | versicolor | 145 | virginica |
| 31 | | setosa | 88 | versicolor | 146 | virginica |
| 32 | | setosa | 89 | versicolor | 147 | virginica |
| 33 | | setosa | 90 | versicolor | 148 | virginica |
| 34 | | setosa | 91 | versicolor | 149 | virginica |
| 35 | | setosa | 92 | versicolor | 150 | virginica |
| 36 | | setosa | 93 | versicolor | Total | N |
| 43 | | setosa | 100 | versicolor |
| 44 | | setosa | 101 | virginica |
| 45 | | setosa | 102 | virginica |
| 46 | | setosa | 103 | virginica |
| 47 | | setosa | 104 | virginica |
| 48 | | setosa | 105 | virginica |
| 49 | | setosa | 106 | virginica |
| 50 | | setosa | 107 | versicolor |
| 51 | | versicolor | 108 | virginica |
| 52 | | versicolor | 109 | virginica |
| 53 | | versicolor | 110 | virginica |
| 54 | | versicolor | 111 | virginica |
| 55 | | versicolor | 112 | virginica |
| 56 | | versicolor | 113 | virginica |
| 57 | | versicolor | 114 | virginica |
|  | |  | 115 | virginica |

**Case 5** Consider the data set flowerspecies.xlsx which has 150 observations on four features namely sepal length, sepal width, petal length, petal width for 3 different species of flowers (50 observations for each species).Learn a 3 class classification hypothesis on given data, which predicts the species type based on the given 4 features using **K Means Cluster Analysis**

Obtain the estimated classes for all the observations in the given data. Construct the observed vs predicted classification table and calculate the class wise and overall percentage of misclassification.

**K Means Cluster Analysis**

Cluster analysis is a type of data classification carried out by separating the data into groups. The aim of cluster analysis is to categorize n objects in k (k>1) groups, called clusters, by using p (p>0) variables. As with many other types of statistical, cluster analysis has several variants, each with its own clustering procedure. There are two main sub-divisions of clustering procedures. In the first procedure the number of clusters is pre-defined. This is known as the K-Means Clustering method. When the number of the clusters is not predefined we use Hierarchical Cluster analysis.

* This procedure attempts to identify relatively homogeneous groups of cases based on selected characteristics, using an algorithm that can handle large numbers of cases. However, the algorithm requires the user to specify the number of clusters. We can specify initial cluster centers if we know this information.
* We can select one of two methods for classifying cases, either updating cluster centers iteratively or classifying only. We can save cluster membership, distance information, and final cluster centers. Optionally, we can specify a variable whose values are used to label case wise output. We can also request analysis of variance F statistics.
* While these statistics are opportunistic (the procedure tries to form groups that do differ), the relative size of the statistics provides information about each variable's contribution to the separation of the groups.

| **Initial Cluster Centers** | | | |
| --- | --- | --- | --- |
|  | Cluster | | |
|  | 1 | 2 | 3 |
| X1 | 7.70 | 5.70 | 4.90 |
| X2 | 3.80 | 4.40 | 2.50 |
| X3 | 6.70 | 1.50 | 4.50 |
| X4 | 2.20 | .40 | 1.70 |

The initial cluster centers are the variable values of the (k=3) well-spaced observations.

They are vectors with their values based on the four variables, (Sepal length, sepal width, petal length, petal width) which refer to setosa (first cluster), versicolor (second cluster), and virginica (third cluster).

| Iteration Historya | | | |
| --- | --- | --- | --- |
| Iteration | Change in Cluster Centers | | |
| 1 | 2 | 3 |
| 1 | 1.226 | 1.205 | 1.141 |
| 2 | .175 | .000 | .121 |
| 3 | .070 | .000 | .047 |
| 4 | .050 | .000 | .033 |
| 5 | .000 | .000 | .000 |
| a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 5. The minimum distance between initial centers is 3.824. | | | |

In early iterations, the cluster centers shift quite a lot. By the 3 rd iteration, they have settled down to the general area of their final location, and the last two iterations are minor adjustments.

The iteration history shows the progress of the clustering process at each step.

| **Final Cluster Centers** | | | |
| --- | --- | --- | --- |
|  | Cluster   * The final cluster centers are computed as the mean for each variable within each final cluster. * The final cluster centers reflect the characteristics of the typical case for each cluster. | | |
|  | 1 | 2 | 3 |
| X1 | 6.85 | 5.01 | 5.90 |
| X2 | 3.07 | 3.43 | 2.75 |
| X3 | 5.74 | 1.46 | 4.39 |
| X4 | 2.07 | .25 | 1.43 |

| Number of Cases in each Cluster  A large number of cases were assigned to the third cluster. | | |
| --- | --- | --- |
| Cluster | 1 | 38.000 |
| 2 | 50.000 |
| 3 | 62.000 |
|  | Valid | 150.000 |
| Missing | .000 |

**Crosstabs**

| **X5 \* Cluster Number of Case Crosstabulation**  We can see that all flowers of type setosa are in second cluster, versicolor is predominantly in 3rd cluster  And virginica in first cluster. | | | | | |
| --- | --- | --- | --- | --- | --- |
| Count | | | | | |
|  |  | Cluster Number of Case | | |  |
|  |  | 1 | 2 | 3 | Total |
| X5 | setosa | 0 | 50 | 0 | 50 |
| versicolor | 2 | 0 | 48 | 50 |
| virginica | 36 | 0 | 14 | 50 |
|  | Total | 38 | 50 | 62 | 150 |

Let us see the following example of clustering method.

* Example. What are some identifiable groups of television shows that attract similar audiences within each group? With k-means cluster analysis, you could cluster television shows (cases) into k homogeneous groups based on viewer characteristics. This process can be used to identify segments for marketing. Or you can cluster cities (cases) into homogeneous groups so that comparable cities can be selected to test various marketing strategies

**Quick Cluster**

K-means cluster analysis is a tool designed to assign cases to a fixed number of groups (clusters) whose characteristics are not yet known but are based on a set of specified variables. It is most useful when we want to classify a large number (thousands) of cases.

A good cluster analysis is:

• **Efficient-** Uses as few clusters as possible.

• **Effective-**Captures all statistically and commercially important clusters.

**Principles**

The K-Means Cluster Analysis procedure begins with the construction of initial cluster centers. We can assign these our self or have the procedure select k well-spaced observations for the cluster centers.

After obtaining initial cluster centers, the procedure:

• Assigns cases to clusters based on distance from the cluster centers.

• Updates the locations of cluster centers based on the mean values of cases in each cluster.

These steps are repeated until any reassignment of cases would make the clusters more internally variable or externally similar

| Case Summariesa | | | |
| --- | --- | --- | --- |
|  |  | X5 | Cluster Number of Case  The case summaries have been shown above.All cluster number of cases,have been shown in the last column. |
|  | 1 | setosa | 2 |
| 2 | setosa | 2 |
| 3 | setosa | 2 |
| 4 | setosa | 2 |
| 5 | setosa | 2 |
| 6 | setosa | 2 |
| 7 | setosa | 2 |
| 8 | setosa | 2 |
| 9 | setosa | 2 |
| 10 | setosa | 2 |
| 11 | setosa | 2   | 46 | setosa | 2 | | --- | --- | --- | | 47 | setosa | 2 | | 48 | setosa | 2 | | 49 | setosa | 2 | | 50 | setosa | 2 | | 51 | versicolor | 3 | | 52 | versicolor | 3 | | 53 | versicolor | 1 | | 54 | versicolor | 3 | | 55 | versicolor | 3 | | 56 | versicolor | 3 | | 57 | versicolor | 3 | | 58 | versicolor | 3 | | 59 | versicolor | 3 | | 60 | versicolor | 3 | | 61 | versicolor | 3 | | 62 | versicolor | 3 | | 63 | versicolor | 3 | | 64 | versicolor | 3 | | 65 | versicolor | 3 | | 66 | versicolor | 3 | | 67 | versicolor | 3 | | 68 | versicolor | 3 | | 69 | versicolor | 3 | | 70 | versicolor | 3 | | 71 | versicolor | 3 | | 72 | versicolor | 3 | | 73 | versicolor | 3 | | 74 | versicolor | 3 | | 75 | versicolor | 3 | | 76 | versicolor | 3 | | 77 | versicolor | 3 | | 78 | versicolor | 1 | | 79 | versicolor | 3 | | 80 | versicolor | 3 | | 81 | versicolor | 3 | | 82 | versicolor | 3 | | 83 | versicolor | 3 | | 84 | versicolor | 3 | | 85 | versicolor | 3 | | 86 | versicolor | 3 | | 87 | versicolor | 3 | | 88 | versicolor | 3 | | 89 | versicolor | 3 | | 90 | versicolor | 3 | | 91 | versicolor | 3 | | 92 | versicolor | 3 | | 93 | versicolor | 3 | |
| 12 | setosa | 2 |
| 13 | setosa | 2 |
| 14 | setosa | 2 |
| 15 | setosa | 2 |
| 16 | setosa | 2 |
| 17 | setosa | 2 |
| 18 | setosa | 2 |
| 19 | setosa | 2 |
| 20 | setosa | 2 |
| 21 | setosa | 2 |
| 22 | setosa | 2 |
| 23 | setosa | 2 |
| 24 | setosa | 2 |
| 25 | setosa | 2 |
| 26 | setosa | 2 |
| 27 | setosa | 2 |
| 28 | setosa | 2 |
| 29 | setosa | 2 |
| 30 | setosa | 2 |
| 31 | setosa | 2 |
| 32 | setosa | 2 |
| 33 | setosa | 2 |
| 34 | setosa | 2 |
| 35 | setosa | 2 |
| 36 | setosa | 2 |
| 37 | setosa | 2 |
| 38 | setosa | 2 |
| 39 | setosa | 2 |
| 40 | setosa | 2 |
| 41 | setosa | 2 |
| 42 | setosa | 2 |
| 43 | setosa | 2 |
| 44 | setosa | 2 |
| 45 | setosa | 2 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
| | 94 | versicolor | 3 | | --- | --- | --- | | 95 | versicolor | 3 | | 96 | versicolor | 3 | | 97 | versicolor | 3 | | 98 | versicolor | 3 | | 99 | versicolor | 3 | | 100 | versicolor | 3 | | 101 | virginica | 1 | | 102 | virginica | 3 | | 103 | virginica | 1 | | 104 | virginica | 1 | | 105 | virginica | 1 | | 106 | virginica | 1 | | 107 | virginica | 3 | | 108 | virginica | 1 | | 109 | virginica | 1 | | 110 | virginica | 1 | | 111 | virginica | 1 | | 112 | virginica | 1 | | 113 | virginica | 1 | | 114 | virginica | 3 | | 115 | virginica | 3 | | 116 | virginica | 1 | | 117 | virginica | 1 | | 118 | virginica | 1 | | 119 | virginica | 1 | | 120 | virginica | 3 | | 121 | virginica | 1 | | 122 | virginica | 3 | | 123 | virginica | 1 | | 124 | virginica | 3 | | 125 | virginica | 1 | | 126 | virginica | 1 | | 127 | virginica | 3 | | 128 | virginica | 3 | | 129 | virginica | 1 | | 130 | virginica | 1 | | 131 | virginica | 1 | | 132 | virginica | 1 | | 133 | virginica | 1 | | 134 | virginica | 3 | | 135 | virginica | 1 | | 136 | virginica | 1 | | 137 | virginica | 1 | | 138 | virginica | 1 | | 139 | virginica | 3 | | 140 | virginica | 1 | | 141 | virginica | 1 | |  | | 142 | virginica | 1 | | --- | --- | --- | | 143 | virginica | 3 | | 144 | virginica | 1 | | 145 | virginica | 1 | | 146 | virginica | 1 | | 147 | virginica | 3 | | 148 | virginica | 1 | | 149 | virginica | 1 | | 150 | virginica | 3 | | N | 150 | 150 | | a. Limited to first 150 cases. | | | | |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  | Advantages:   * When used with ecological data it produces nice discrete groups that are usually easy to interpret.   Disadvantages:   * There needs to be a certain amount of trial and error in choosing the number of clusters. * The implementation of the procedure in SPSS is restricted to measuring distances between samples using Euclidean distance. * This becomes important when we have presence/absence data. |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |