Week 5 Assignment

Sai Kumar Murarishetti
Lewis ID: L30079224

1. **On your own words, discuss (in less than one page) the differences between Multiple Regression Analysis and Multiple Discriminant Analysis.**

   - **Multiple Regression Analysis**:
     - It mainly used to model the relationship between single dependent variables and multiple independent variables.
     - Used to Predict of single dependent value.
     - It provides the information about the Strength and direction.
     - It also determines how each explanatory factor affects the response factors.
     - It also used to predict one variable from a combined knowledge of several other variables.
     - It also known as Statistical Methods.

   - **Multiple Discriminant Analysis:**
     - It mainly used in objects classification into pre-defined groups based on variables predicted.
     - It provides information on discriminant functions which are used in classification.
     - It includes discriminant functional and canonical coefficients in discriminant analysis.
     - For predicting the Single dependent values from the known independent variables.
     - The analysis is quite sensitive to outliers and the small of group must be larger than the number of predictor variables.
     - Multiple discriminant analysis also known as canonical variates analysis or canonical discriminant analysis.

2. For the data set associated with this homework (HBAT and HBAT_Test. Using X4 as the non-metric variable and (X6 up to X18) as the metric variables:

a. What does each variable represent? (go back to Week # 2)

Variable Type	Variable Description	
Non-Metric	Region	X4
Metric	Product Quality	X6
Metric	E-commerec Activities /websites	X7
Metric	Technical Support	X8
Metric	Complaint Resolution	X9
Metric	Advertising	X10
Metric	Product Line	X11
Metric	Sales force Image	X12
Metric	Competitive pricing	X13
Metric	Warranty & Claims	X14
Metric	New Products	X15
Metric	Ordering & Billing	X16
Metric	Price Flexibility	X17
Metric	Delivery Speed	X18

b. How many groups does X4 has?

2 Groups

c. Apply linear discriminant analysis to the data (HBAT) and find:
• The linear discriminant function for X4.

Linear Discriminant Function for x4			
Variable	Label	0	1
Constant		-191.92135	-194.33409
x6	x6	8.32797	7.65248
x7	x7	4.20485	1.20101
x8	x8	-2.06370	-2.09852
x9	x9	-3.62428	-3.60295
x10	x10	-1.62571	-2.03642
x11	x11	58.68681	58.34999
x12	x12	1.69711	4.70052
x13	x13	3.64274	4.22872
x14	x14	13.54926	13.38761
x15	x15	0.00591	0.28730
x16	x16	-2.84573	-2.25264
x17	x17	62.42839	64.51335
x18	x18	-101.40047	-103.57123

• By applying the LDF to the training data (HBAT): How many observations were misclassified? What are they? Find the confusion matrix and the probability of (error)misclassification.

Training Data (HBAT)

Posterior Probability of Membership in x4					
Obs	From x4	Classified into x4		0	1
3	1	0	*	0.9388	0.0612
22	1	0	*	0.9999	0.0001
24	1	0	*	0.6602	0.3398
32	1	0	*	0.5773	0.4227
38	1	0	*	0.9394	0.0606
42	0	1	*	0.4490	0.5510
60	1	0	*	0.9396	0.0604
63	0	1	*	0.3361	0.6639
64	1	0	*	0.6576	0.3424
74	1	0	*	0.7029	0.2971
81	0	1	*	0.4834	0.5166
88	0	1	*	0.3170	0.6830
94	1	0	*	0.9345	0.0655

* Misclassified observation

Finding the Confusion Matrix

Number of Observations and Percent Classified into x4			
From x4	0	1	Total
0	35 89.74	4 10.26	39 100.00
1	9 14.75	52 85.25	61 100.00
Total	44 44.00	56 56.00	100 100.00
Priors	0.39	0.61	

• By applying the LDF to the test data (HBAT_Test): How many observations were misclassified? What are they? Find the confusion matrix and the probability of (error)misclassification.

Confusion matrix of (HBAT_Test)

Observation Profile for Test Data	
Number of Observations Read	100
Number of Observations Used	100

Number of Observations and Percent Classified into x4			
From x4	0	1	Total
0	37 94.87	2 5.13	39 100.00
1	11 18.03	50 81.97	61 100.00
Total	48 48.00	52 52.00	100 100.00
Priors	0.39	0.61	

Error Count Estimates for x4			
	0	1	Total
Rate	0.0513	0.1803	0.1300
Priors	0.3900	0.6100	