Strathmore
UNIVERSITY

INSTITUTE OF MATHEMATICAL SCIENCES
MASTER OF SCIENCE IN STATISTICAL SCIENCES
END OF SEMESTER EXAMINATION
STA 8316: MACHINE LEARNING AND PATTERN RECOGNITION

DATE: 25th April, 2022 Time: 2.5 Hours

Instructions
1. This examination consists of **FIVE** questions and an Appendix to one of the questions.
2. Answer **Question ONE (COMPULSORY)** and any other **TWO** questions.

Question 1 (20 Marks)

a) Explain the significance of the concept of *Bias-variance trade-off* in a statistical learning algorithm.

(**4** Marks)

b) Distinguish between the following concepts in classification algorithms: *Sensitivity and specificity.*

(**4** Marks)

c) A Receiver Operating characteristic function is a useful tool in predictive modeling. Explain how an ROC functions can be developed for a given statistical learning procedure and describe its application in predictive modeling.

(**6** Marks)

d) Bagging, boosting, random forests and stochastic gradient boosting algorithms are ensemble methods that are often used to enhance the quality of a decision tree. Briefly describe each approach, clearly highlighting the enhancement that they give.

(**6** Marks)

Question 2 (20 Marks**)**

a) Suppose that Y is a discrete (outcome) random variable taking values $1,2,\dots,g$ and that $X_1, X_2, \dots, X_k$ are k predictor variables that explain the variability in Y. Assuming a prior probability $\pi_i = P(Y = i)$, that the density of the random vector $\boldsymbol{X} = (X_1, X_2, \dots, X_k)'$ given $Y = i$ is multivariate normal with mean vector $\boldsymbol{\mu_i}$ and with a variance covariance matrix Σ, derive expressions for the linear and quadratic score functions that are used in Linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA), respectively.

[**12** Marks]

b) Consider a data set of 144 observations of household cats. The data contains the cats' gender, body weight and height. The researcher would like to model and accurately predict the gender of a cat based on previously observed values.
To verify and test our model's performance, they split the data into training (60%) and test sets (40%). Two models were entertained:

- Model 1: A logistic regression model with body weight as predictors
- Model 2: A logistic regression model with body weight and height as predictors

The confusion matrices for these two models are also presented in Table 1

Table 1 Confusion matrix for Model 1 (left) and 2 (right)

	Predicted status	
Actual status	Female	Male
Female	12	10
Male	13	22

	Predicted status	
Actual status	Female	Male
Female	9	15
Male	13	20

i) From the confusion matrices above, compare the 3 models. Compare your results based on model accuracy.

[**4** Marks]

ii) For the best fitting model, compute the following measures: sensitivity, specificity and the false positive rate.

[**6** Marks]

Question 3 (20 Marks**)**

Ridge regression, Lasso regression and elastic net regression are three common penalized regression techniques in the machine learning literature.

a) For each regression approach write the cost function associated with the optimization used to estimate the regression parameters in the model.

[**6** Marks]

b) Compare and contrast Lasso regression with ridge regression model in terms of interpretability, prediction accuracy, and variable selection.

[**6** Marks]

c) Optimize the cost function for the ridge regression model and estimate parameters of the ridge regression, their mean and bias.

[**6** Marks]

Question 4 (20 Marks**)**

a) Describe the purpose and objective of *Principal Components Analysis* (PCA) and give any 3 examples of areas in which its finds application.

(**5** Marks)

b) Cluster analysis is a commonly employed unsupervised learning procedure. Distinguish between Agglomerative clustering and Divisive clustering algorithms.

(**3** Marks)

c) Explain how the Partitioning around Medoids (PAM) approach works.

(**4** Marks)

d) A random sample of 74 cars was selected. For each car the following variables were measured: **headroom** [Headroom (in.)], **trunk** [Trunk space (cu. ft.)], **weight** [Weight (lbs.)], **length** [Length (in.)], **turn** [Turn Circle (ft.)], and **displacement** [Displacement (cu. in.)].
Based on the results of the PCA analysis given in the Appendix:

i. Explain how many principal components you would select and why

(**2** Marks)

ii. Explain what each of the selected component(s) describes;

(**2** Marks)

i. Comment on the results of the 10 cars considered on the basis each of the components selected;

(**2** Marks)

ii. Comment on the correlation circle and it's significance.

(**2** Marks)

Question 5 (20 Marks**)**

a) Decision trees are procedures employed extensively in machine learning and statistical literature. Briefly explain how these procedures work and also mention aspects of their efficiency and reliability.

(**7** Marks)

b) A major problem associated with regression trees is instability. Explain what you understand by this problem.

(**3** Marks)

c) Regression tree can be seen as a kind of additive model or a piecewise constant regression model. Mathematically or using an appropriate example, clearly explain this assertion.

(**5** Marks).

d) Describe the recursive partitioning algorithm and its utility in decision trees

(**5** Marks)

APPENDIX

Table 2 Correlation Matrix

	headroom	trunk	weight	length	turn	displacement
headroom	1.0000000	0.6620111	0.4834558	0.5162955	0.4244646	0.4744915
trunk	0.6620111	1.0000000	0.6722057	0.7265956	0.6010595	0.6086350
weight	0.4834558	0.6722057	1.0000000	0.9460086	0.8574429	0.8948958
length	0.5162955	0.7265956	0.9460086	1.0000000	0.8642612	0.8351400
turn	0.4244646	0.6010595	0.8574429	0.8642612	1.0000000	0.7767647
displacement	0.4744915	0.6086350	0.8948958	0.8351400	0.7767647	1.0000000

Table 3 Eigen-values

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	4.50151930	75.0253217	75.02532
Dim.2	0.80149921	13.3583202	88.38364
Dim.3	0.30817531	5.1362552	93.51990
Dim.4	0.22411069	3.7351781	97.25508
Dim.5	0.12361234	2.0602056	99.31528
Dim.6	0.04108315	0.6847191	100.00000

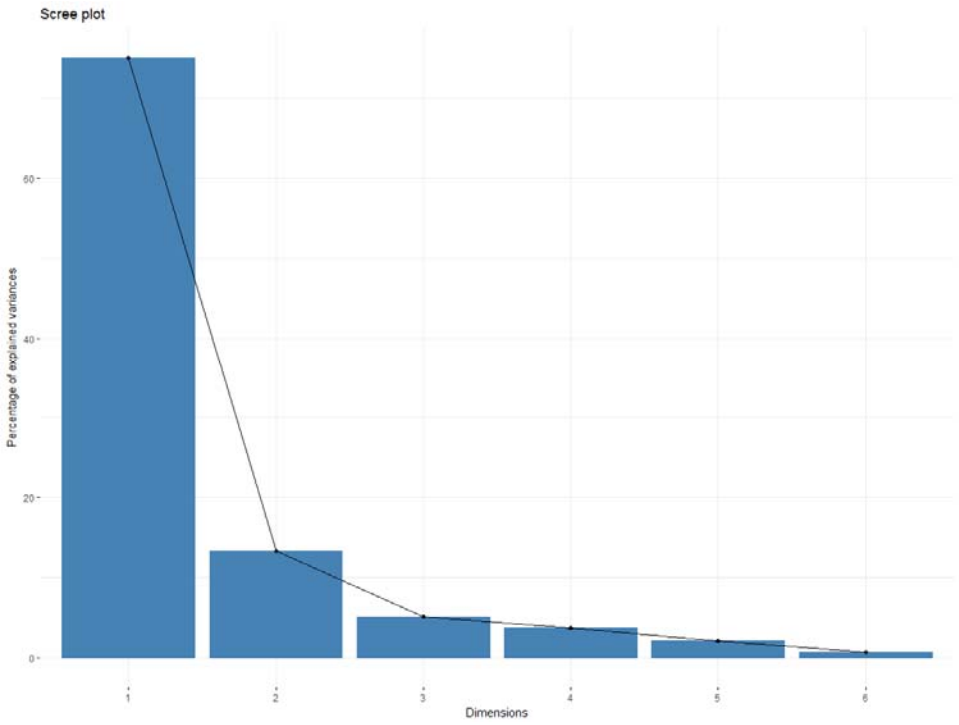

Figure 1 Scree-plot

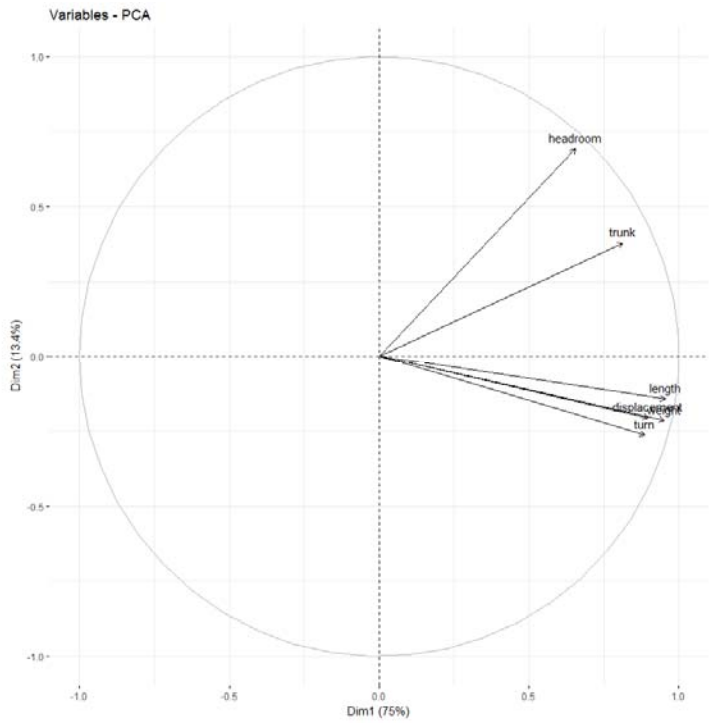

Figure 2 Correlation circle

Table 4 Summary of results

Eigenvalues

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6
Variance	4.502	0.801	0.308	0.224	0.124	0.041
% of var.	75.025	13.358	5.136	3.735	2.060	0.685
Cumulative % of var.	75.025	88.384	93.520	97.255	99.315	100.000

Individuals (the 10 first)

	Dist	Dim.1	ctr	cos2	Dim.2	ctr	cos2	Dim.3	ctr	cos2
AMC Concord	1.222	-0.842	0.213	0.475	-0.518	0.452	0.180	-0.085	0.032	0.005
AMC Pacer	1.229	-0.043	0.001	0.001	-0.440	0.326	0.128	0.829	3.014	0.455
AMC Spirit	1.748	-1.581	0.750	0.818	0.600	0.607	0.118	0.083	0.030	0.002
Buick Century	1.930	1.082	0.351	0.314	1.458	3.586	0.571	0.518	1.176	0.072
Buick Electra	3.354	3.272	3.214	0.952	0.359	0.217	0.011	0.001	0.000	0.000
Buick LeSabre	2.761	2.491	1.862	0.813	0.915	1.412	0.110	-0.630	1.741	0.052
Buick Opel	2.351	-1.206	0.436	0.263	0.121	0.025	0.003	1.064	4.961	0.205
Buick Regal	1.542	0.453	0.062	0.086	-1.014	1.735	0.433	-1.059	4.922	0.472
Buick Riviera	1.912	1.844	1.021	0.930	0.071	0.009	0.001	-0.147	0.095	0.006
Buick Skylark	1.167	0.966	0.280	0.685	-0.059	0.006	0.003	0.566	1.402	0.235

Variables

	Dim.1	ctr	cos2	Dim.2	ctr	cos2	Dim.3	ctr	cos2
headroom	0.655	9.536	0.429	0.692	59.741	0.479	0.293	27.901	0.086
trunk	0.813	14.688	0.661	0.379	17.905	0.144	-0.428	59.333	0.183
weight	0.951	20.108	0.905	-0.216	5.807	0.047	0.037	0.435	0.001
length	0.955	20.280	0.913	-0.144	2.577	0.021	-0.060	1.172	0.004
turn	0.887	17.478	0.787	-0.264	8.687	0.070	0.014	0.060	0.000
displacement	0.898	17.911	0.806	-0.206	5.283	0.042	0.185	11.099	0.034