

CSP 571 Data Preparation and Analysis

Midterm Exam - Part 1

- 1) A real-valued response variable has observations with a total sum of squares (TSS) value of 4987 - a regression analysis has produced a residual sum of squares (RSS) of 1780. Based on these values, what would be the R^2 of this regression?**

A. 0.64
- 2) Given a classification task with a single feature, we wish to perform binary classification via LDA - we have the mean feature value for observations in Class 1 as -107.41 and in Class 2 as 160.85. Assuming the same prior probabilities for each class, at what value is the decision boundary located?**

A. 26.72
- 3) Given a dataset of 26635 sample observations with 17 features, we add an additional 3854 sample observations - assuming each feature follows a Gaussian/Normal distribution, what is the resulting number of parameters that we would need to estimate for the sample covariance matrix?**

A. 306
- 4) Given a dataset with 2581 sample observations, we hold out 15 percent for a test set. The remaining data is used as a training set where we perform 4-fold cross-validation. What is the size of each validation set in this case?**

A. 548
- 5) We fit a cubic spline model for a single feature non-linear regression, resulting in a total of 28 total degrees-of-freedom. Assuming we have selected K knot points in our training dataset values, what are the total number of constraints that have been applied to the estimation?**

A. 72

6) We fit a single feature logistic regression model with 17741 training data observations, and when classifying a new observation x we obtain a value of 39 for the odds ratio. What is the probability value implied for this data point?

A. 0.98