IS71104A Statistics and Statistical Data Mining – Coursework

Elliot Walker [SN: 3368 6408]1

1Goldsmiths, University of London

8 Lewisham Way, London SE14 6NW

**Abstract.**

# 1 Overview

# 2 Theory

## 2.1 Regression

## 2.2 Classification

## 2.3 Feature Scaling

## 2.4 Confidence Intervals

# 3 Tasks

## 3.1 Task 1 of 3

The **Auto.csv** dataset contains statistics recorded over a period of 12 years between 1970 and 1982, denominated with one response and nine inputs:

* MPG Miles per Gallon of fuel consumed

by the automobile’s engine.

* Cylinders The number of cylinders possessed

by the automobile’s engine.

* Displacement Engine displacement.
* Horsepower Horsepower produced by the

engine.

* Weight Weight (mass under gravity) of

the engine.

* Acceleration How much the automobile progresses

between two velocities.

* Year Year which the automobile was

manufactured.

* Origin Region from which the automobile

descends from. Encoded as for America (USA), for Europe and for Japan.

* Name The model of the automobile

instance/observation.

The dataset contains 397 observations, one of which possesses a duplicate (in name, year of manufacture, and cylinder count).

## 3.2 Task 2 of 3

## 3.3 Task 3 of 3

# 4 Discussion

# 5 Conclusion

# References