MFE204TC
ARTIFICIAL INTELLIGENCE AND DATA ANALYSIS

LECTURE 5

LONG HUANG

Xi'an Jiaotong-Liverpool University
西交利物浦大學

CONTENTS

- Review of Taught Contents
- Review Practices

Certain contents of this presentation are adopted from training material at MathWorks

TAUGHT CONTENTS

- Matlab Basics

- Data Processing
  - Importing/Exporting Data
  - Plotting Data
  - Missing Data
  - Data Smoothing
  - Outlier Detection
  - Nonuniform Data
  - Inconsistent Data
  - Filter Data
  - Smooth Data with Convolution
  - Detrending Data
  - Descriptive Statistics

TAUGHT CONTENTS

- Regression Analysis
  - Linear Correlation
  - Linear Regression
  - Interactive Fitting
  - Programmatic Fitting

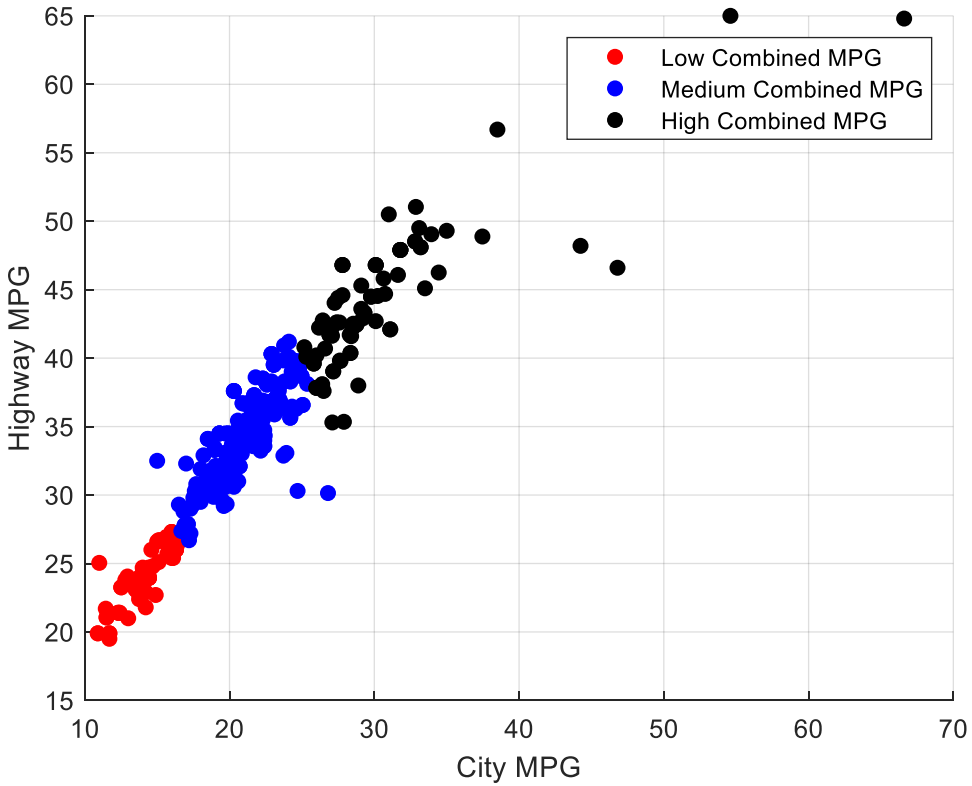
- Time Series Analysis

REVIEW QUESTION - 1

The file "fuelEconomy.txt" contains fuel economy data for different models of cars. The goal of this exercise is to divide the combined miles per gallon data into three classes and create a scatter plot of city and highway miles per gallon like the one shown below.

- Step 1 – Import the data into a table.
- Step 2 – The variable CombinedMPG contains missing values represented by NaN. Remove the rows corresponding to these missing values from the imported table.
- Step 3 – Discretize the CombinedMPG variable into three classes called "Low", "Medium", and "High" using the bin edges 0, 20, 30, and 70 (values greater than or equal to 0 and less than 20 are classified as "Low" and so on). Store the discretized result in a variable named MPGClass.
- Step 4 – Create a scatter plot with CityMPG on the x-axis and HighwayMPG on the y-axis. Observations with "Low" combined MPG should be colored red ("r"), "Medium" should be blue ("b"), and "High" should be "black" ("k").

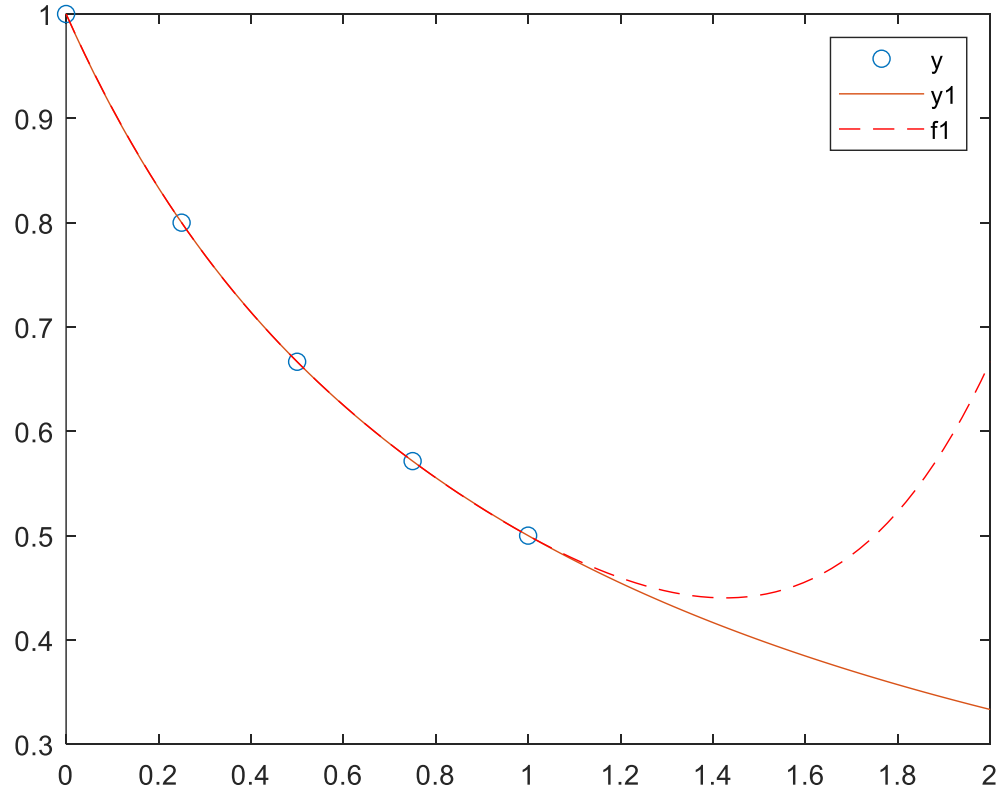
REVIEW QUESTION – 1 - SOLUTION



REVIEW QUESTION - 2

1. Create a vector of 5 equally spaced points using linspace function in the interval [0,1], and evaluate $y(x)=(1+x)^{-1}$ at those points.
2. Fit a polynomial of degree 4 to the 5 points.
3. Evaluate the original function and the polynomial fit on a finer grid of points between 0 and 2.
4. Plot the function values and the polynomial fit in the wider interval [0,2], with the points used to obtain the polynomial fit highlighted as circles.

REVIEW QUESTION – 2 SOLUTION

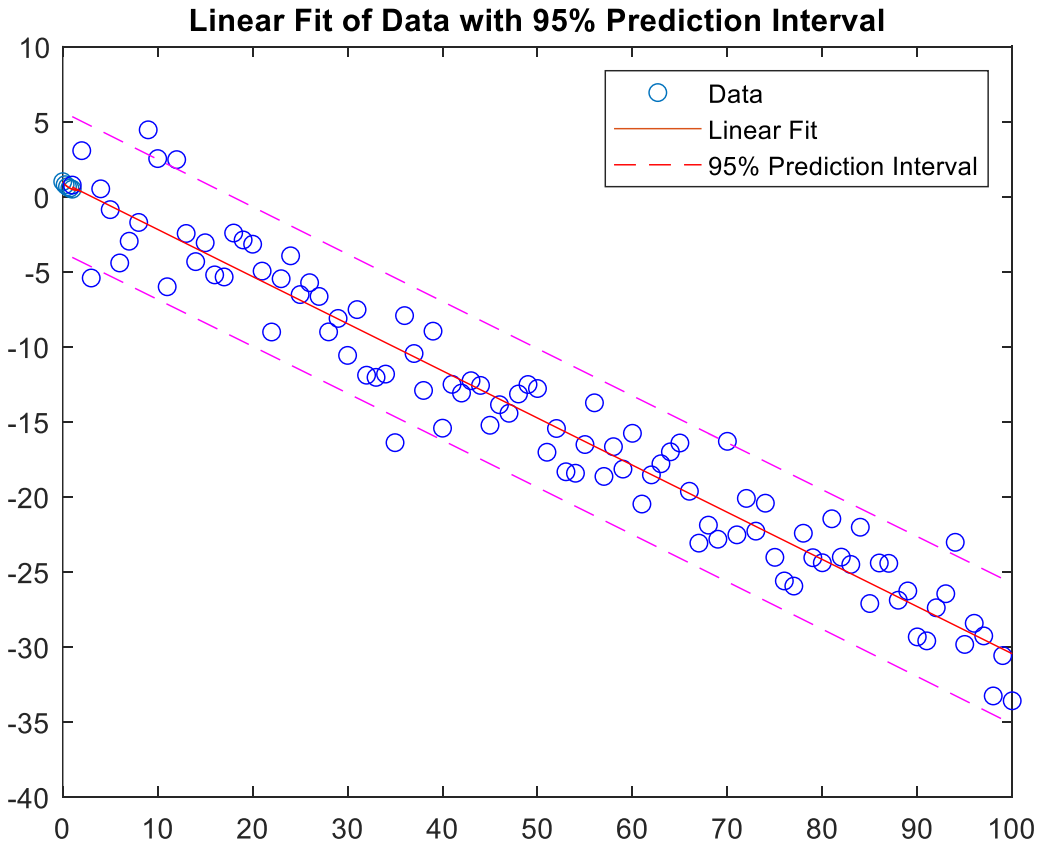


REVIEW QUESTION - 3

Fit a linear model to a set of data points and plot the results, including an estimate of a 95% prediction interval.

1. Create a few vectors of sample data points (x,y). Use polyfit to fit a first degree polynomial to the data. Specify two outputs to return the coefficients for the linear fit as well as the error estimation structure
2. Evaluate the first-degree polynomial fit in p at the points in x. Specify the error estimation structure as the third input so that polyval calculates an estimate of the standard error. The standard error estimate is returned in delta.
3. Plot the original data, linear fit, and 95% prediction interval $y \pm 2\Delta$.

REVIEW QUESTION – 3 SOLUTION



REVIEW QUESTION - 4

Use the following code to create an array of monthly counts of airline passengers, measured in thousands, for the period January 1949 through December 1960.

1. Please create an array of dates from the monthly data and use it along with the Y data to create the time series object.
2. Plot the timeseries data.

```
% 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960
y = [112 115 145 171 196 204 242 284 315 340 360 417 % Jan
118 126 150 180 196 188 233 277 301 318 342 391 % Feb
132 141 178 193 236 235 267 317 356 362 406 419 % Mar
129 135 163 181 235 227 269 313 348 348 396 461 % Apr
121 125 172 183 229 234 270 318 355 363 420 472 % May
135 149 178 218 243 264 315 374 422 435 472 535 % Jun
148 170 199 230 264 302 364 413 465 491 548 622 % Jul
148 170 199 242 272 293 347 405 467 505 559 606 % Aug
136 158 184 209 237 259 312 355 404 404 463 508 % Sep
119 133 162 191 211 229 274 306 347 359 407 461 % Oct
104 114 146 172 180 203 237 271 305 310 362 390 % Nov
118 140 166 194 201 229 278 306 336 337 405 432 ]; % Dec
% Source:
% Hyndman, R.J., Time Series Data Library,
% http://www-personal.buseco.monash.edu.au/~hyndman/TSDL/.
% Copied in October, 2005.
```

REVIEW QUESTION – 4 - SOLUTION