

CEE322 Autumn 2014
Lab 3: Linear Regression

Part 1:

The dataset Plant.csv contains the operating conditions of a combined cycle power plant recorded as hourly averages over the course of a handful of years. The data was taken at times when the plant was running at full load. The variables in the data are:

- Temperature (T) in °C
- Ambient Pressure (AP) in millibar
- Relative Humidity (RH)
- Exhaust Vacuum (V) in cm Hg
- Net electrical energy output (PE) in MW

The owner is interested in what best explains the energy output of the plant. Can you provide some insight based on the data provided?

- 1) How well can temperature explain the energy output?
- 2) What kind of relationship, if any, is there between relative humidity and energy output?
- 3) What combination of variables would you use to best explain the output of the plant in this dataset?

Part 2:

A researcher has measured a number of rock samples, applying techniques from the analysis of fibers. She believes this approach may yield new ways of understanding rock strength, and she's trying to publish some of her results. Your advisor, with whom she collaborates, would like you to evaluate the data and comment on what relationships you find and postulate one model. Justify your conclusions based on the data provided. Keep in mind that we have relatively few data points in this set.

The dataset is Rocks.csv. The researcher emailed over this description of the columns.

Uniaxial compressive strength (ucs)
Quartz Content in % (qtz)
Plagioclase Content in % (plag)
Feldspar Content in % (k.fds)
Hornblende Content in % (hb)
Grain Size in pixels/mm (gs)
Grain Area in pixels/mm² (ga)
Shape Factor (close to 1 for circular, close to 0 if elongated) (sf)
Aspect Ratio as ratio major/minor axis lengths (ar)

Hint: Navigate through an EDA to identify appropriate relationships. Make sure to diagnose any fits you come up with.