Energy Efficiency
Fall 2023

Background

Alternative building designs area evaluated for energy efficiency – both heating and cooling.

Dataset

Use the EnergyEfficiency.xlsx spreadsheet. This dataset is available through the University of California – Irvine machine learning repository:

https://archive.ics.uci.edu/ml/datasets/Energy+efficiency

Note: The Excel file has the data and a description.

Energy Efficiency dataset

Data Set Information:

We perform energy analysis using 12 different building shapes simulated in Ecotect. The buildings differ with respect to the glazing area, the glazing area distribution, and the orientation, amongst other parameters. We simulate various settings as functions of the afore-mentioned characteristics to obtain 768 building shapes. The dataset comprises 768 samples and 8 features, aiming to predict two real valued responses. It can also be used as a multi-class classification problem if the response is grouped.

Attribute Information:

The dataset contains eight attributes (or features, denoted by X1...X8) and two responses (or outcomes, denoted by y1 and y2). The aim is to use the eight features to predict each of the two responses.

Specifically:

X1 Relative Compactness
X2 Surface Area
X3 Wall Area
X4 Roof Area
X5 Overall Height
X6 Orientation (2=North, 3=East, 4=South, 5=West)
X7 Glazing Area
X8 Glazing Area Distribution (1=Uniform, 2=North, 3=East, 4=South, 5=West)

y1 Heating Load
y2 Cooling Load

Note that Orientation and Glazing Area Distribution are coded. These must be converted to dummy variables.

Assignment

What's due:

PowerPoint presentation due before class on Monday, December 11, 2023. Expected length of presentation is 15-20 minutes, approximately 15-25 slides. Please send me the slides at least one hour before class. You can describe the slides from your seat; I'll advance the slides.

Outline

Using the Energy Efficiency dataset, perform an analysis of the following aspects of the data.

1. Visualization (10%)
    a. Generate scatterplots to display relationships (5%)
    b. Generate histograms for each variable (5%)
2. Statistical analysis: Generate statistical analysis that is appropriate (20%)
    a. Descriptive statistics (5%)
    b. Correlation (5%)
    c. Linear regression for each Y variable (cooling load and heating load) (5%)
    d. VIF analysis (5%)
3. Run additional modeling analysis using the techniques that you've learned so far (60%)
    a. Assign categories for each cooling load and heating load (A for highest quartile, B for second highest quartile, C for third highest quartile, D for lowest quartile) (10%)
    b. Using a loop, create five (5) perceptrons for each Y variable (cooling load and heating load) (assign 1 for categories A and B, -1 for categories C and D). Report the accuracies of all ten (10) models. (10%)
    c. Create a support vector machine for each Y variable (cooling load and heating load; use categories A, B, C, D) . Generate one graph for each SVM. (10%)
    d. Using a loop for 1 to 5 hidden nodes, create a neural network with multiple outputs (cooling load and heating load) (use actual numeric values for loads, continuous output neural network, two outputs). Report the accuracies for all five (5) models. (10%)
    e. Create two K-nearest neighbor analyses (cooling load and heating load) (use categories A, B, C, D) (10%)
    f. Create two Naïve Bayes analyses (cooling load and heating load) (use categories A, B, C, D) (10%)
4. Identify a list of lessons learned (10%)
    a. Which technique worked best? Show the accuracies of each model when possible. (5%)
    b. What insights do you have for builders of energy efficient buildings? (5%)