An Exposition on Multiple Regression: A Case Study for Applications

Gregory A. Bauer, Divya Kamath

Department of Engineering, San Diego University

Prof. Dallin Munger

**An Exposition on Multiple Regression: A Case Study for Applications**

A recent study involving the categorization of wines based on its quality involved both objective and subjective approaches. In the later case, wine quality is generally assessed by physiochemical and sensory tests (Cortez, Paulo et. Al, July 2008). These sensory tests are performed by human ‘expert’ taste testers which are assumed to be the final arbiter of “quality.” As part of the case study’s literature review, (Cortez, Paulo et. Al, July 2008) reviewed several Data Mining (DM) techniques that were used to map the physiochemical parameters (e.g. alcohol and density) with a sensory taste panel which often proved difficult – frequently mitigated by using an “electronic tongue.” These above referenced DM techniques involved the us of Neural Networks (NN) and Support Vector Machines (SVM) to discriminate wines into separate classifications; and not necessarily into the subjective categories of “Quality.”

(Cortez, Paulo et. Al, July 2008) methodology used both NN and SVM DM techniques to build models that support wine evaluation or “quality.” This in practical terms, involves taking the physiochemical properties from the wine data set and performing variable and model selection, using sensitivity analysis and kernel selection methods, to select the best performing model in terms of the how accurate the model is at predicting a wine’s “Quality.”

As a baseline (Cortez, Paulo et. Al, July 2008) also used a classical approach to modeling continuous data, Multiple Regression, to gauge the performance of their NN and SVM models – seeking to outperform the classical approach using novel techniques. Our approach, as an academic exercise, seeks to develop a MR model based on the same data set to illustrate how that process looks from beginning to end; something the authors left out completely or had barely a mention of.

# Methods

## Data

## Exploratory Analysis

## Multiple Regression Analysis

## Model Selection Techniques

# Results

## Discussion