Landon Buell

January 2022

Min-Max Feature Variance: A new Preprocessing Tool for Feature Selection

1. Abstract

It is known across all branches of machine learning and artificial intelligence that feature engineering is critically important to the success of any model. Often, developers are confronted with the problem of choosing a set features that are able to accurately and reliably represent a data sample in compact, and low dimensional format. This process can be lengthy and costly for developers to execute, but due to it importance cannot be overlooked. To aid in the feature engineering process for classification problems, we have designed and implemented a technique for isolating a subset of features in a design matrix. This process highlights features that exhibit a low variance within samples of each class, and a high variance over samples in all classes.

1. Introduction
2. Implementation
   1. Setting up + Given Information
      * 1. Design Matrix “X”, Label Vector “Y”
        2. X has shape (n x p). Y has shape (n x 1)
        3. There are k unique classes – [0, 1, 2, … , k-1]
        4. X can have any number of dimension, we choose 2 for this example
      1. Pseudo -Code
3. Practical Example
4. Conclusion