MBA15

REGULARIZATION OF DATASETS USING SVD

Ben Kim
Seattle University, Seattle, WA, USA

Abstract

Regularization in machine learning is essential for preventing overfitting and reducing generalization errors. While L1 (lasso regression) and L2 (ridge regression or Tikhonov regularization) are commonly used techniques, this paper explores SVD (Singular Value Decomposition) as a method for regularization. SVD is particularly effective for high-dimensional and noisy datasets. We propose two specific methods: one that selects dimensions with singular values exceeding a threshold, and another that uses data compression techniques. For each method, we explain the implementation process and present results on execution time, memory usage, and generalization errors, evaluated through cross-validation using the random forest regression algorithm. All demonstrations are implemented in Python, and the code is provided for reference.

Conference Track

MIS and Business Analytics