Employing supervised machine learning algorithms for fitting terrain data

Eina B. Jørgensen, Anna Lina P. Sjur and Jan-Adrian H. Kallmyr

September 17, 2019

Abstract

1 Introduction

The use of machine learning for problem solving has risen in popularity as large data sets have become available for analysis. There now exists many different methods in varying complexity for both supervised and unsupervised learning. All of these methods have advantages and drawbacks, as well as many similarities. This means we can get familiar with some of the central themes in machine learning by studying simple algorithms. In this report, we will implement three different supervised learning algorithms with increasing complexity, as well as the k-fold resampling technique.

2 Theory

2.1 Ordinary least squares

2.2 Ridge regression

2.3 Lasso regression

2.4 Mean squared error

2.5 Score function

2.6 k-fold cross validation

There are several methods for estimating the skill of a machine learning model. One such method is the k-fold cross-validation procedure, which is used when working with a limited data sample. The idea is to divide the data sample into k groups or folds, and then retain some of the data to use as a test set after fitting a model to the remaining data.

```
Shuffle the dataset randomly;
Divide the dataset into $k$ folds;
foreach $k$ do
    Take the $k$th fold out to use as test data set;
    Set the remaining folds as training data set;
    Fit a model to the training set;
    Evaluate the model on the test set;
    Retain the evaluation score and discard the model ;
Calculate the mean of the evaluation scores;
```

Algorithm 1: The k-fold cross-validation algorithm.

$$A = \begin{bmatrix} b_1 & c_1 & 0 & \dots & \dots & 0 \\ a_1 & b_2 & c_2 & 0 & \dots & 0 \\ 0 & a_2 & b_3 & c_3 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \ddots & a_{n-2} & b_{n-1} & c_{n-1} \\ 0 & \dots & \dots & 0 & a_{n-1} & b_n \end{bmatrix},$$

3 Results

Figure 1

4 Discussion

5 Conclusion

References