

Problem

The initial task was to predict soil water permeability for forest harvesting operations.

Implementation idea

The predictions were done using the k-nearest neighbor algorithm. The prediction performance of 5-nearest neighbor was estimated using the spatial leave-one-out cross validation with dead zone radiuses from 0m to 200m with 10m intervals. The prediction performance was calculated in C-index. In addition, the data was prepared by normalizing it using the z-score.

The implementation structures in the following way:

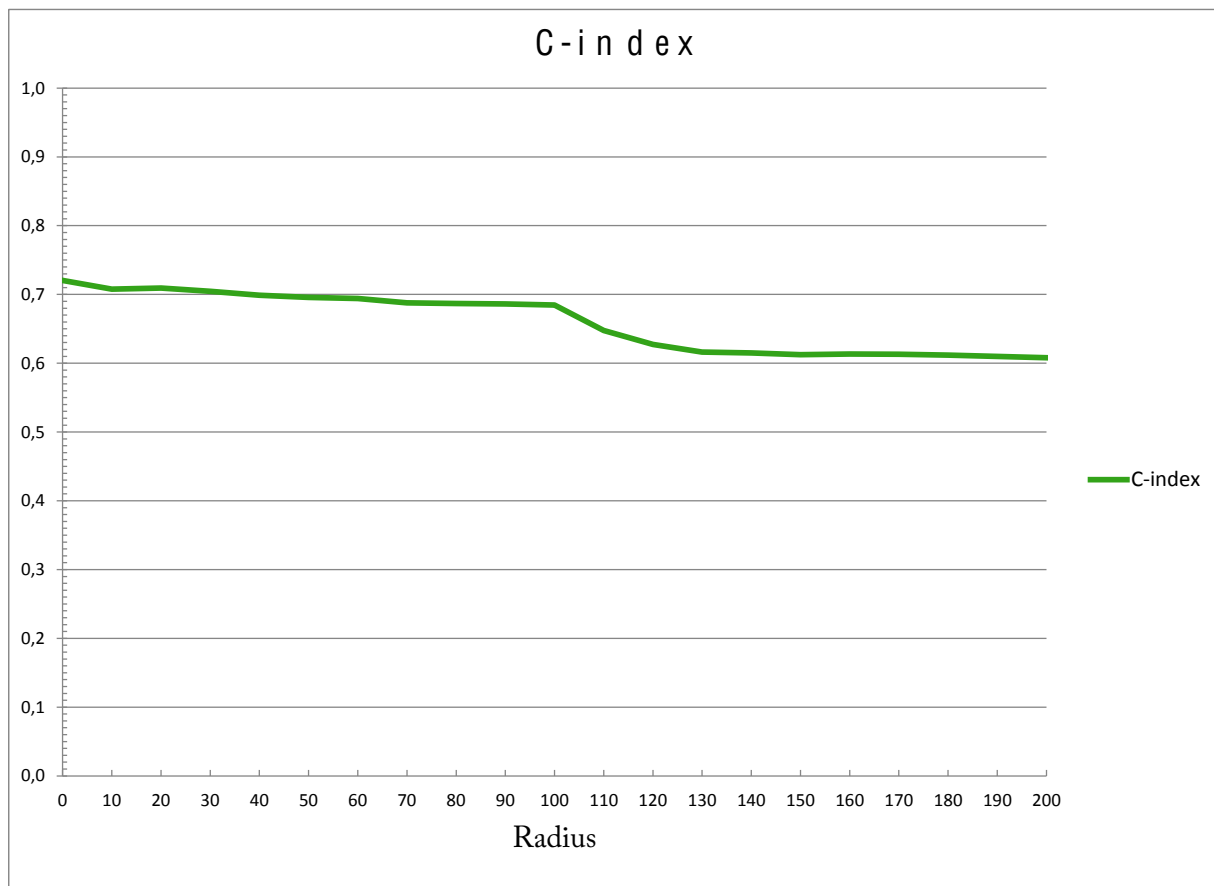
1. Read the data
2. Convert numerical values into floats
3. Normalize the data using the z-score
4. Run the knn algorithm using spatial leave-one-out cross validation
 - a. Determine the test instance
 - b. Calculate the euclidean distances and remove the data points inside the dead zone
 - c. Get the 5-nearest neighbors using the reduced data set
 - d. Get the prediction by calculating the mean value of the 5-nearest neighbors
 - e. Calculate the c-index for each prediction
 - f. Print the C-indexes and their corresponding radiuses

Description of the data set

The dataset has 1691 data points which divided into three different files:

1. COORDINATES.csv = Each row has two (some kind of) coordinate values. Subtracting these from each other yields the distance of those points.
2. INPUT.csv = Each row has 95 features that are extracted from aerial photographs from the target area
3. OUTPUT.csv = Each row has one value that represents manually measured water permeability sample point from Pomokaira area

Results



The image represents the results of the algorithm. It can be concluded that the model performs worse as the radius grows.