

Task3:

1. How does the construction of regression trees differ to classification trees? How is a prediction computed in a regression tree?

In a regression tree, each leaf node represents a numeric value. This numeric value is generally the average of the subgroup of datapoints it is assigned to.

In contrast, classification trees have a binary variable or a discrete class representation in their leaves.

The decision threshold of a node is selected by the lowest sum of squared residuals.

A prediction in a regression tree is done like in a classification tree: The tree is traversed down to a leaf node based on the used features. The difference is that the leaf node is assigned to a continuous value and not to a class.

Sources:

Yisehac Yohannes & Patrick Webb: "Classification and Regression Trees, Cart: A User Manual for Identifying Indicators of Vulnerability to Famine and Chronic Food Insecurity" Intl Food Policy Research Ins, Dezember 1998

StatQuest with Josh Starmer: "Regression Trees, Clearly Explained!!!", 2019,

Url: <https://www.youtube.com/watch?v=g9c66TUylZ4>

University of Cincinnati Business Analytics R Programming Guide: "Regression Trees"

Url: https://uc-r.github.io/regression_trees

2. How can kNN be used for regression?

In kNN for regression a prediction is made via the nearest neighbors of the training sample in respect to a line in the feature space. This line represents the feature values at which a prediction should be made. The average of the kNN is taken to determine a predicted value. This results in a simple way to predict even more complex (nonlinear) functions. As in kNN for classification, this doesn't require a dedicated training step.

Max Miller: "The Basics: KNN for classification and regression: Building an intuition for how KNN models work", towards datascience, October 2018

Url: <https://towardsdatascience.com/the-basics-knn-for-classification-and-regression-c1e8a6c955>