
Data Mining & Machine Learning

Yong Zheng

Illinois Institute of Technology
Chicago, IL, 60616, USA

ILLINOIS TECH

College of Computing

HW2

- Common Mistakes: KNN
 - Normalization in KNN should be applied to all columns, not a single column
 - Model is built from training set only. In KNN, the best parameters (K, distance) should be learned from training set only
 - Students do not know how to use training/testing sets
 - Students do not know how to build confusion matrix to calculate F1 score

HW2

- Common Mistakes: KNN Python Coding
 - You cannot drop columns without strong reasons
 - Loan term is the label, not other columns
 - Some students simply assigned 0 to all missing values
 - Some students used linear regression rather than KNN
 - Didn't change distance measures
 - Do not know how to proceed with Cross validation