

Machine Learning

Assignment 10.1

Submitted By: Ranji Raj

January 13, 2021

Comparing supervised and unsupervised learning

Supervised learning	Unsupervised learning
T: Assign labels/classes to instances	T: Assign similar instances to same cluster
E: Classify/predict output from input data	E: Learns inherent structures from data
P: Scores indicated by confusion matrix, SSE	P: Internal and external cluster validity measures
Classification: Naive Bayes, Logistic regression, Decision Tree, Perceptron, Neural Networks, k NN Regression: Linear regression, Locally weighted regression, Regression Tree	Clustering: k -Means, Fuzzy-C-Means, k NN, Hierarchical agglomerative clustering Association rules: Apriori algorithm, Eclat, Frequent pattern-growth

Table 1: Supervised vs. Unsupervised learning

Practical task solvable

Classification

Given different parameters like, *pregnancies*, *Glucose level*, *Blood pressure*, *Skin thickness*, *Insulin*, *BMI*, *Age* can determine whether a patient is diabetic or non-diabetic using a Decision Tree model.

Regression

Given different parameters like, *Mileage*, *Vehicle age*, *Vehicle age group*, *make*, *color set*, *state*, *make model*, *lot sale day*, *overage* can be used to predict the total cost of an

automotive using a linear regression model.

Clustering

Given different parameters like, *Day*, *distance(km)*, *duration*, *day category*, *last workout*, *habits* can be used to cluster different age groups based upon physical activity using *k*-Means algorithm.

Association rule

Census data - Every government has tonnes of census data. This data can be used to plan efficient public services(education, health, transport) as well as help public businesses (for setting up new factories, shopping malls, and even marketing particular products). This application of association rule mining and data mining has immense potential in supporting sound public policy and bringing forth an efficient functioning of a democratic society.