

# DM Assignment 12

19.3.2016

Adithya Abraham Philip

1PI13CS008

## Overview

This document consolidates the observations and answers to questions pertaining to the use of the WEKA API for outlier detection.

## Observations

### General characteristics of outliers:

All the outliers lie in the “high” region of their respective outlier attributes i.e they all have values greater than  $Q3 + IQR * OF$ , and none have values lesser than  $Q1 - IQR * OF$  for the attribute in which they are an outlier.

## Questions & Answers

### 1. What is the effect of changing OF?

**Ans:**

**Increasing OF** results in a fewer number of outliers, in both the low and high regions, for each attribute.

**Decreasing OF** results in a larger number of outliers, in both the low and high regions, for each attribute.

### 2. Can you make any general statement about the quality of outliers?

**Ans:** Most of the outliers seem to be of **MEDIUM** quality (their quality rating lies between  $Q1$  and  $Q3$  for the quality attribute).