Project 2

Benjamin Hendrick

February 14, 2016

Part A: Data Pre-processing and Model Evaluation

Problem 1

The formula for *min-max normalization* from the text is:

$$v_i' = \frac{v_i - min_A}{max_A - min_A}(newMax_A - newMin_A) + newMin_A$$

The R function for this formula may be written as:

```
minMaxNorm <- function (x,l,u){
  normalized <- (x-min(x))/(max(x)-min(x))*(u-l)+l
  return(normalized)
}
```

where `x` is the data being normalized, `l` is the new lower bound, and `u` is the new upper bound.

The formula for *z-score normalization* from the text is:

$$v_i' = \frac{v_i - \bar{A}}{\sigma_A}$$

The R function for this formula may be written as:

```
zScoreNorm <- function (x,type){
  if (type == 1) {
    z <- (x-mean(x))/sd(x)
    return(z)
  }
  if (type == 2) {
    (x - mean(x)) / (length(x)^-1 * sum(abs(x-mean(x))))
  }
}
```

where `x` is the data being normalized, and `type` is the type of z-score normalization being selection (1 uses standard deviation and 2 uses mean absolute deviation.)

Problem 2

Load the data from the text.

```
data <- c(200,300,400,600,1000)
```

Part A

```
minMaxNorm(data,0,1)
```

```
## [1] 0.000 0.125 0.250 0.500 1.000
```

Part B

```
zScoreNorm(data,1)
```

```
## [1] -0.9486833 -0.6324555 -0.3162278  0.3162278  1.5811388
```

Part C

```
zScoreNorm(data,2)
```

```
## [1] -1.2500000 -0.8333333 -0.4166667  0.4166667  2.0833333
```

The normalized data, using all three methods in parts A, B, and C, may be summarized in the following table.

Orignal Data	Min-Max Normalization	Z-Score Normalization (s.d.)	Z-Score Normalization (m.a.d.)
200	0.000	-0.9486833	-1.2500000
300	0.125	-0.6324555	-0.8333333
400	0.250	-0.3162278	-0.4166667
600	0.500	0.3162278	0.4166667
1000	0.100	1.5811388	2.0833333