Homework 02 - STAT440

Joseph Sepich (jps6444)

09/04/2020

Problem 1

Which of the following is an appropriate variable name?

- (a) 1st_var
- (b) first_var
- (c) first.var

first_var or choice b is the appropriate variables name of the three choices. Variables cannot start with a number and using a dot in the variable name can be confused with function syntax.

Problem 2

Recall that if $x := (x_1, ..., x_d) \in R^d$, then the euclidean norm of x is $||x||_2 = \sqrt{\Sigma_{i=1}^d x_i^2}$. Let

$$V = [v_1, v_2, v_3, v_4, v_5] = \begin{vmatrix} 1 & 2 & 4 & -1 & 0 \\ 2 & 1 & -4 & 1 & 3 \\ 3 & 0 & 1 & -1 & 5 \end{vmatrix}$$

Create matrix V in R:

```
mat_v <- matrix(c(1, 2, 3, 2, 1, 0, 4, -4, 1, -1, 1, -1, 0, 3, 5), nrow = 3, ncol=5)
mat_v
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    1    2    4   -1    0
## [2,]    2    1   -4    1    3
## [3,]    3    0    1   -1    5
```

Use R to do the following

2a

Create a matrix D made out of the norm of all pairwise distances of the column vectors of V. That is, the ij^{th} entry of D is $||v_i - v_j||_2$.

```r
l2_norm <- function(vec) {
  sqrt(sum(vec^2))
}

num_cols <- dim(mat_v)[2]
mat_d <- matrix(1:25, nrow = num_cols, ncol = num_cols)
for (i in 1:num_cols) {
  for (j in 1:num_cols) {
    mat_d[i, j] <- l2_norm(mat_v[,i] - mat_v[,j])
  }
}
mat_d
```

```
##          [,1]     [,2]     [,3]     [,4]     [,5]
## [1,] 0.000000 3.316625 7.000000 4.582576 2.449490
## [2,] 3.316625 0.000000 5.477226 3.162278 5.744563
## [3,] 7.000000 5.477226 0.000000 7.348469 9.000000
## [4,] 4.582576 3.162278 7.348469 0.000000 6.403124
## [5,] 2.449490 5.744563 9.000000 6.403124 0.000000
```

2b

Use D to compute the average and standard deviation of these distances. Be careful not to double count.

```r
dists <- mat_d[upper.tri(mat_d,diag=TRUE)]
print('Average:')
```

```
## [1] "Average:"
```

```r
print(mean(dists))
```

```
## [1] 3.63229
```

```r
print('Standard Dev:')
```

```
## [1] "Standard Dev:"
```

```r
print(sd(dists))
```

```
## [1] 3.140712
```

2c

Find vectors y_j so that the j^{th} of D_{y_j} is the average distance from v_j to all other points. Report these numbers.