Stat 154 HW04

Mokssh Surve

2/24/2020

Problem 5

Setting up data matrices and parameters for parts a) & b)

```
# extracting and processing data in order to make suitable for regresisonal analysis
dat <- read.csv("Advertising.csv", row.names = 1)
X_a <- as.matrix(dat[, 1:3])
y <- dat$sales
n <- nrow(X_a) # no. of individuals/data points
p <- ncol(X_a) # no. of characterisitcs/features

# modifying matrix for part b) by swapping columns radio and newspaper
X_b <- X_a
X_b[, c(2,3)] <- X_b[, c(3,2)]
colnames(X_b)[c(2,3)] <- colnames(X_b)[c(3,2)]

# setting up output matrices
Z_a <- matrix(1, n, p+1) # basis of orthogonal vectors (z values)
G_a <- matrix(0, p, p) # diagonal matrix with eigen values (gamma)
B_a <- rep(0,p) # regression coefficients (beta)

Z_b <- matrix(1, n, p+1)
G_b <- matrix(0, p, p)
B_b <- rep(0,p)
```

5a) $\textbf{sales} = \beta_0 + \beta_1\textbf{TV} + \beta_2\textbf{radio} + \beta_3\textbf{newspaper} + \epsilon$

```
#using loop method

for (i in 1:p)
{
  # normalising z vector - taking the sum of the squares of each column till i
  z_norm <- apply(Z_a[, 1:i, drop=FALSE], 2, function(c) sum(c*c))
  # modifying G
  G_a[1:i, i] <- t(Z_a[, 1:i]) %*% X_a[, i] * (1/(z_norm))

  g <- as.matrix(G_a[1:i, i])
  Z_a[, i+1] <- X_a[, i] - (Z_a[, 1:i] %*% g)

  # noting the regression coefficient for pertaining iteration
  B_a[i] <- (sum(y * Z_a[, i+1])) / (sum(Z_a[, i+1] * Z_a[, i+1]))
}
```

```r
# coefficient for NewsPaper () \beta_{3}
B_a[3]
```

```
## [1] -0.001037493
```

```r
# sanity check
reg <- lm(sales ~. , data=dat)
reg$coefficients[4]
```

```
##    newspaper
## -0.001037493
```

5b) $\textbf{sales} = \beta_0 + \beta_1\textbf{TV} + \beta_2\textbf{newspaper} + \beta_3\textbf{radio} + \epsilon$

```r
#using loop method

for (i in 1:p)
{
  # normalising z vector - taking the sum of the squares of each column till i
  z_norm <- apply(Z_b[, 1:i, drop=FALSE], 2, function(c) sum(c*c))
  # modifying G
  G_b[1:i, i] <- t(Z_b[, 1:i]) %*% X_b[, i] * (1/(z_norm))

  g <- as.matrix(G_b[1:i, i])
  Z_b[, i+1] <- X_b[, i] - (Z_b[, 1:i] %*% g)

  # noting the regression coefficient for pertaining iteration
  B_b[i] <- (sum(y * Z_b[, i+1])) / (sum(Z_b[, i+1] * Z_b[, i+1]))
}

# coefficient for Radio () \beta_{3}
B_b[3]
```

```
## [1] 0.18853
```

```r
# sanity check
reg <- lm(sales ~. , data=dat)
reg$coefficients[3]
```

```
##   radio
## 0.18853
```