Gradient descent

Jason Wang

Monday, April 20, 2015

Here, I would like to introduce a method, gradient descent, to solve the coefficient of regression.(From machine learning courses)

Our cost function: $\mathrm{J}(\theta) = \frac{1}{n}\sum_{i=1}^{n} (h_\theta(x_i) - y_i)^2$

Our goal: $minimize_\theta\ \mathrm{J}(\theta)$

Our solution;

repeat until convergence{

θ_j <- θ_j - $\alpha\frac{\partial}{\partial\theta}\mathrm{J}(\theta)$

}

```
#Gradient descent
#A method to solve coefficient while would not
#provide estimation of standard deviation
Gradient_descent <- function(X, y, theta, alpha, iteration){
    theta <- as.matrix(theta); n <- dim(as.matrix(X))[1]
    Xnew <- as.matrix(cbind(rep(1, n), X))
    outputlist <- list(); outputlist$Cost <- c(0)
    for(i in 1:iteration){
        outputlist$Cost[i] <- sum((Xnew %*% theta - y)^2)/n
        temp <- theta[1] - alpha*sum(Xnew %*% theta - y)/n
        temp2 <- theta[-1] - alpha*(t(Xnew[, -1])%*%(Xnew%*%theta - y))/n
        theta[1] <- temp
        theta[-1] <- temp2
    }
    outputlist$Final_theta <- theta
    outputlist
}
#Demo
data(mtcars)
X <- mtcars$hp
y <- mtcars$mpg
(my_theta <- Gradient_descent(X, y, c(30, 6), 0.000001, 10000)[[2]])
```

```
##             [,1]
## [1,] 29.96609367
## [2,] -0.06748127
```

```
coef(lm(mpg ~ hp, data=mtcars))
```

```
## (Intercept)          hp
## 30.09886054 -0.06822828
```

```r
X_all <- mtcars[, c(4, 6, 8, 9, 11)]
(my_theta <- Gradient_descent(X_all, y, rep(1, 6), 0.00001, 10000)[[2]])
```

```
##            [,1]
## [1,] 1.41405411
## [2,] 0.03925162
## [3,] 1.52708167
## [4,] 1.53153752
## [5,] 1.38363912
## [6,] 0.98812042
```

```r
(my_theta <- Gradient_descent(X_all, y, c(30, rep(1, 5)), 0.00005, 10000)[[2]])
```

```
##             [,1]
## [1,] 29.79016952
## [2,] -0.07558679
## [3,] -0.28763787
## [4,]  0.86275312
## [5,]  1.29575218
## [6,]  0.57582496
```

```r
#Convert vs and am to factor
mtcars$vs <- as.factor(mtcars$vs)
mtcars$am <- as.factor(mtcars$am)
#Using lm to fit model
m1 <- lm(mpg ~ hp + wt + vs + am + carb, data=mtcars)

coef(m1)
```

```
## (Intercept)          hp          wt         vs1         am1        carb
## 30.48551739 -0.02339723 -2.40937042  1.77406494  2.96876808 -0.42434830
```

```r
#New data
new_data <- mtcars[, c(1, 4, 6)]
#Scale the data
new_data_scale <- scale(new_data)[1:32, 1:3]
m2 <- lm(as.data.frame(new_data_scale))
coef(m2)
```

```
##   (Intercept)            hp            wt
##  3.813089e-17 -3.614507e-01 -6.295545e-01
```

```r
X <- new_data_scale[, 2:3]
y <- new_data_scale[, 1]
(my_theta <- Gradient_descent(X, y, c(1, 1, 1), 0.01, 10000)[[2]])
```

```
##               [,1]
## [1,]  9.542100e-17
## [2,] -3.614507e-01
## [3,] -6.295545e-01
```