exercise3

Lee Rui (BA19008042)

2020/3/26

```
rm(list = ls())
```

Read data

```
setwd("E:/RStudio/workspace/ecology")
mydata <- read.csv("exercise3/xy.csv")
mydata
```

```
##      x  y
## 1  100 55
## 2  120 60
## 3  140 62
## 4  160 64
## 5  180 68
## 6  200 70
## 7  220 80
## 8  240 85
## 9  260 90
## 10 280 95
```

```
x <- mydata$x
y <- mydata$y
```

Define cost function

```
cost <- function(X, y, theta) {
  sum( (X %*% theta - y)^2 ) / (2*length(y))
}
```

Define step and number of iterations

```
alpha <- 0.00001
num_iters <- 1000
```

```
cost_history <- rep(0,num_iters) # Save the value of cost function
theta_history <- list(num_iters) # Save theta
theta <-  matrix(c(0,0), nrow = 2) # Initial theta
X <- cbind(1,x) # Make the hypothesis function have an intercept
```

Gradient descent cycle

```
for (i in 1:num_iters) {
  theta[1] <- theta[1] - alpha * (1/length(y)) * sum(((X%*%theta)- y))
  theta[2] <- theta[2] - alpha * (1/length(y)) * sum(((X%*%theta)- y)*X[,2])
  cost_history[i] <- cost(X, y, theta)
  theta_history[[i]] <- theta
}
print(theta)
```

```
##             [,1]
## [1,] 0.02738687
## [2,] 0.37007117
```

Plot the training set data and draw all straight lines during convergence

```
plot(x,y, col=rgb(0.2,0.4,0.6,0.4), main='Linear regression by gradient descent')
for (i in c(1,3,6,10,14,seq(20,num_iters,by=10))) {
  abline(coef=theta_history[[i]], col=rgb(0.8,0,0,0.3))
}
abline(coef=theta, col='blue')
```