CPDA SP18 Assignment 2

*Hugh Jamieson(jamieson.65)*

Introduction

A management firm wants to buy a house from a landlord. The firm wants to estimate an appropriate price of the house that they can offer to the landlord. They hire you as a machine learning engineer to do the job. You have to make an estimation based on the total land area in square feet.

(a) Import the dataset and plot.

```r
housing.data<-read.csv("/Users/hughj/Development/osu/machine-learning/osu-mach-learn/module-2/HousingDa
    header=FALSE,
    sep = ",",
    col.names = c("area","cost"))

# plot the original input dataset
plot(housing.data$area,housing.data$cost,type="p" ,
     main = "Housing Dataset",
     xlab="area in 100 sqft",
     ylab="price in 100K")
```

Housing Dataset


```
##
```
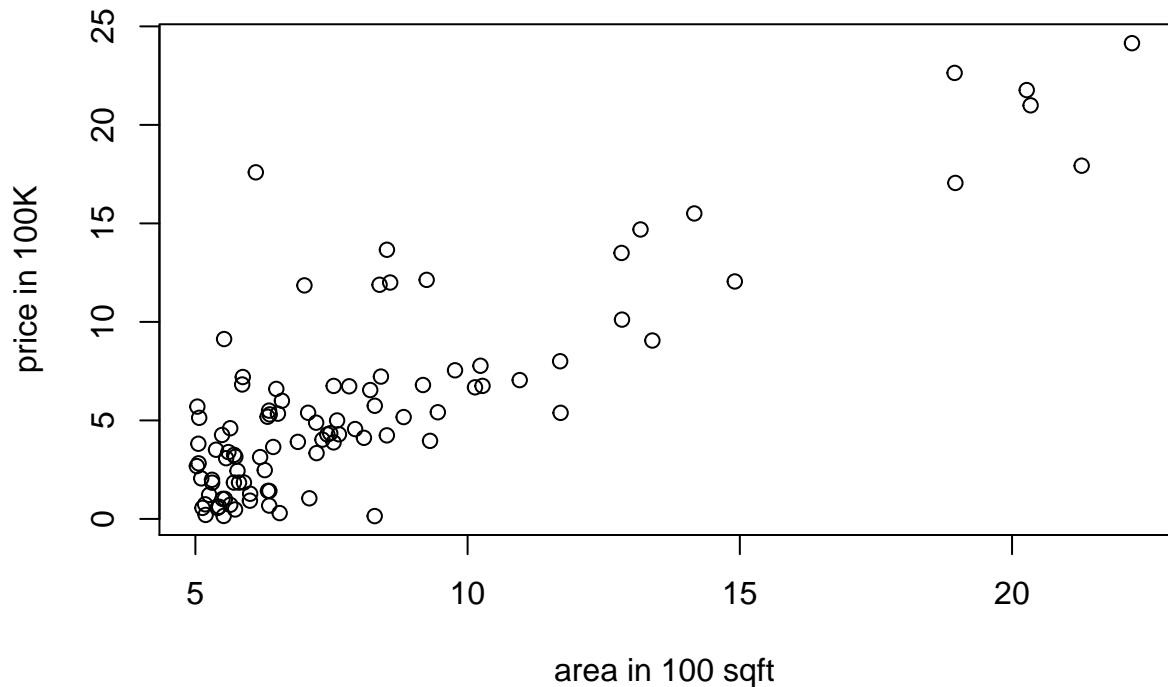
(b) Use linear least squares to find the optimal values of theta (0ls).

```r
# create the matrix for A
rows <- nrow(housing.data)
A <- as.matrix(cbind(rep(1,rows), housing.data$area))
```

```
b <- as.matrix(housing.data$cost)
# calculate the optimal values of theta:
theta <- solve(t(A)%*%A)%*%t(A) %*% b
b0 <- theta[1]
b1 <- theta[2]
# print beta0 & beta1
cat("intercept: ",b0,"\n")
```

```
## intercept:  -3.580962
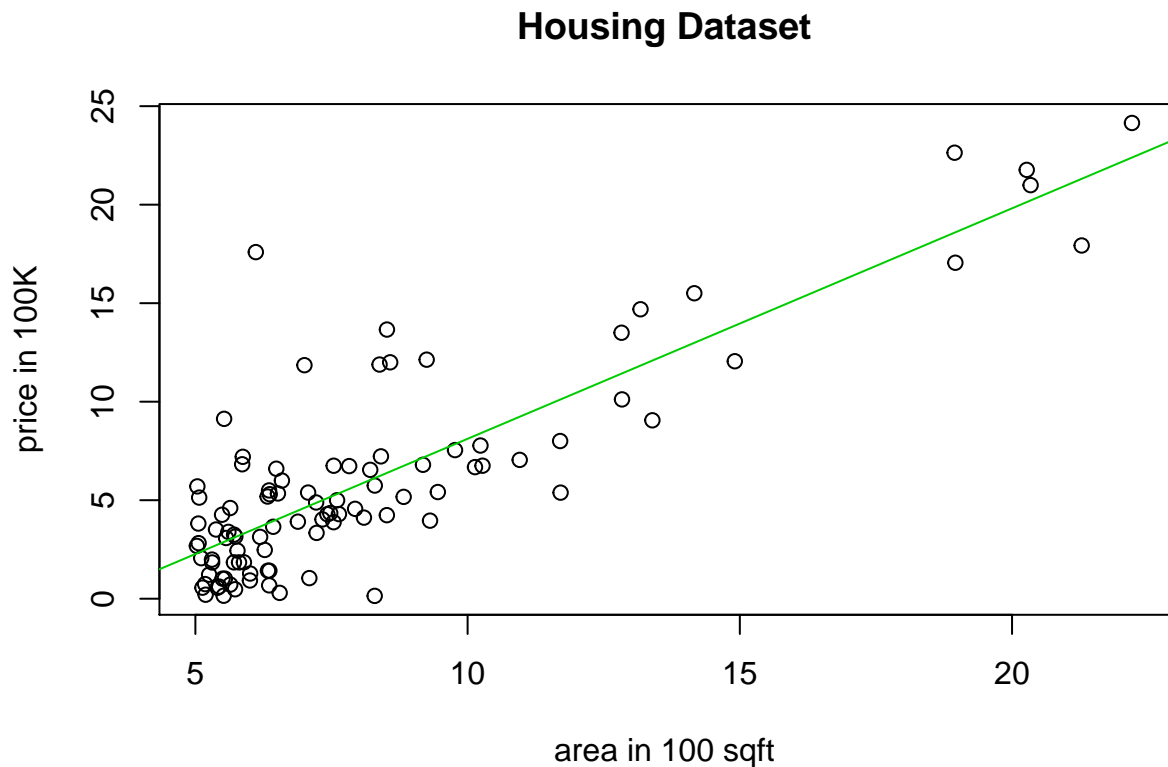```

```
cat("slope:" ,b1, "\n")
```

```
## slope: 1.169776
```

(c) Plot the hypothesis function f(x) with the data:

```
plot(housing.data$area,housing.data$cost,type="p" ,
     main = "Housing Dataset",
     xlab="area in 100 sqft",
     ylab="price in 100K")
abline(b0, b1, col=3)
```


Housing Dataset

##

(d) Find the estimated price of a house in dollars for 700 sqft home:

```
area <- 7
price <- (b1 * area + b0) * 10^5
cat("Estimated price of 700 sqft home: $",price, "\n")
```

```
## Estimated price of 700 sqft home: $ 460747.2
```