

hw6-report

Jack McShane

2022-04-02

1. Let y_i be the car price (in hundreds), x_{i1} be the age, and x_{i2} be the mileage (in thousands) of the i^{th} car, consider the following model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \epsilon_i, i = 1, \dots, n$$

where ϵ_i is the random error component and is assumed to be iid $N(0, \sigma^2)$. Write down the pdf $f(y_i)$.

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \times e^{-\frac{1}{2\sigma^2}(y - (\beta_0 + \beta_1 x_1 + \beta_2 x_2))^2}$$

2. Write down the joint distribution $f(y_1, \dots, y_n)$.

$$f(y_1, \dots, y_n) = f(y_1) \times \dots \times f(y_n)$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})]^2}$$

3. What is the likelihood function $L(\beta_0, \beta_1, \beta_2, \sigma)$.

The Likelihood function occurs when data has been observed. It takes the following form:

$$L(\beta_0, \beta_1, \beta_2, \sigma) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})]^2}$$

4. Write down the log likelihood function, $l(\beta_0, \beta_1, \beta_2, \sigma) = \log L(\beta_0, \beta_1, \beta_2, \sigma)$, and negative log likelihood function $-l(\beta_0, \beta_1, \beta_2, \sigma)$.

$$\begin{aligned}
l(\beta_0, \beta_1, \beta_2, \sigma) &= \log L(\beta_0, \beta_1, \beta_2, \sigma) \\
&= \log \left[\left(\frac{1}{\sqrt{2\pi}\sigma} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})]^2} \right] \\
&= -n \log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})]^2 \\
&= -n \log(\sqrt{2\pi}) - n \log(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})]^2
\end{aligned}$$

$$\begin{aligned}
-l(\beta_0, \beta_1, \beta_2, \sigma) &= -[-n \log(\sqrt{2\pi}) - n \log(\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})]^2] \\
&= n \log(\sqrt{2\pi}) + n \log(\sigma) + \frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})]^2
\end{aligned}$$

5. The maximum likelihood estimator of $\beta_0, \beta_1, \beta_2$, and σ is

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}) = \operatorname{argmax} L(\beta_0, \beta_1, \beta_2, \sigma)$$

explain that it is equivalent to the following

$$\begin{aligned}
(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}) &= \operatorname{argmax} [l(\beta_0, \beta_1, \beta_2, \sigma)] \\
&= \operatorname{argmin} [-l(\beta_0, \beta_1, \beta_2, \sigma)]
\end{aligned}$$

The first equation, $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}) = \operatorname{argmax} [l(\beta_0, \beta_1, \beta_2, \sigma)]$, is equivalent to the original due to the fact that logarithmic functions are monotonically increasing (i.e. the value of the function is forever increasing over its range of x). This property allows us to apply a logarithmic transformation, but preserve the values of the parameters, in this case $\beta_0, \beta_1, \beta_2$ and σ that maximize the likelihood function as they will also be the values that maximize the log likelihood function.

The second function, the negative log likelihood, is simply a reflection across the x-axis, result of which is the previously maximum values of the log likelihood function now represent the minimum values of the *negative* log likelihood function. It therefore follow that the values of $\beta_0, \beta_1, \beta_2$, and σ which minimize the negative log likelihood function are the same values that maximize both the likelihood and the log likelihood functions.

6. Explain that the maximum likelihood estimator and least squared estimator of $\beta_0, \beta_1, \beta_2$ are the same.

With the maximum likelihood estimator, we are trying to find values for β_0, β_1 , and β_2 that maximize the expression: $= -n \log(\sqrt{2\pi}\sigma) - \frac{1}{2\sigma^2} \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})]^2$. The variables β_0, β_1 , and β_2 are present only in the latter term of the expression and because we are maximizing the expression in terms of these variables, it is the only term we need consider. Ignoring the multiplying constant, this simply leaves us with the term $\sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})]^2$ which mirrors the least squared estimator equation. Finally, because of the negative multiplier, we need to minimize the term to maximize the overall expression, leaving us with the least squared estimator, that is, $\argmin \sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})]^2$.

7. In class, we show that $\hat{\beta} = (X^T X)^{-1} X^T y$, carry out this computation in R, what is your $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$?

```
y
##      [,1]
## [1,]  85
## [2,] 103
## [3,]  70
## [4,]  82
## [5,]  89
## [6,]  98
## [7,]  66
## [8,]  95
## [9,] 169
## [10,]  70
## [11,]  48

X <- cbind(1, x1, x2)
X
##      [,1] [,2] [,3]
## [1,]    1    5   57
## [2,]    1    4   40
## [3,]    1    6   77
## [4,]    1    5   60
## [5,]    1    5   49
## [6,]    1    5   47
## [7,]    1    6   58
## [8,]    1    6   39
## [9,]    1    2    8
## [10,]    1    7   69
## [11,]    1    7   89

betas <- solve(t(X)%*%X)%*%t(X)%*%y
betas
##      [,1]
## [1,] 183.0352076
## [2,] -9.5042704
## [3,] -0.8214833
```

8. How do you interpret $\hat{\beta}_1$ and $\hat{\beta}_2$?

$\hat{\beta}_1$ and $\hat{\beta}_2$ are estimations for the true β_1 and β_2 . $\hat{\beta}_1$ and $\hat{\beta}_2$ are the estimated parameters which determine the contribution of our observed features (age and mileage) to the car price prediction, and which minimize the least-squared error calculation. They describe for us (excluding β_0), the line of best fit for the data.

$\hat{\beta}_1 \approx -9.5$. This means that for every year the car gets older, its value drops approximately \$950 as long as the mileage on the car does not change.

$\hat{\beta}_2 \approx -0.82$. This means that for every mile put on the car, the value of the car drops by approximately \$82 as long as the age of the car does not change.

9. Given your estimates, what is your price prediction of a used car with 4 years old, and 50,000 miles?

```
pred <- betas[1] + betas[2]*4 + betas[3]*50
pred
```

```
## [1] 103.944
```

10. Use lm command, what is your $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$? Are they the same as your answers in Question 7?

```
lm(y~x1+x2)
```

```
##
## Call:
## lm(formula = y ~ x1 + x2)
##
## Coefficients:
## (Intercept)          x1          x2
##    183.0352    -9.5043    -0.8215
```

Yes. While not **EXACT**, each of the beta coefficients are nearly identical, further solidifying $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ as the best fit line for the carprice dataset.