

Statistical Learning Methods

Questions Lecture 3

Tobias Famos

June 18, 2022

Questions

1. What is a Linear Simple Regression Model?
2. When do you use a simple linear regression?
3. How do you determine, which linear function to use?
4. Given the RSS, how do we compute the MSE?
5. State model with two attributes, (x_1 and x_2)
6. Interpret the following output and give the model to it.

```
> summary(mod)
lm(formula = price ~ area)

Coeff:      Estimate Std. Error t value Pr(>|t|)
(Intercept) 74.7073    34.0511   2.194 0.059551 .
area        3.5243     0.6346   5.553 0.000539 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

Residual standard error: 32.82 on 8 degrees of freedom
Multiple R-squared:  0.794,    Adjusted R-squared:  0.7683 
F-statistic: 30.84 on 1 and 8 DF,  p-value: 0.000539
```

7. Give the confidence intervall for the area (above)
8. Explain R^2

Answers

1. The linear simple regression model approximates data with a linear function $f(x)$.
2.
 - Either when the dependence of Y to X is linear.
 - Or as the first approximation of data as the model is quite easy and quick.
3.
 - The **least squares method** can be used to test different functions and get a numeric error.
 - We choose the function in a way, that it minimizes the RSS (Residual Sum of Squares).
4. The Media Standard Error (MSE) can be computed form the Residual Sum of Squares (RSS) by dividing by n
5. $y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \epsilon$
6.
 - Model: price = $74.7 + 3.52 \cdot \text{area}$
 - The Model is significant (p-value: 0.05 %) is smaller than 5%
 - 8 degrees of Freedom means there were 9 data points
 - The area is highly significant, the intercept is almost significant
7.
 - Defined by $\hat{\beta}_i \pm 2 \cdot \hat{\sigma}_{\beta_i}$
 - Thus: $3.52 \pm 2 \cdot 0.63 = [2.26, 4.78]$
8.
 - It states how much of the variance is explained in the model.
 - Formula: $R^2 = 1 - \frac{RSS}{TSS}$ Where TSS = Total sum of squares, meaning the variance of the sample.