Linear Regression

November 14, 2017

Contents

Linear Regression

The goal of Linear Regression is to model a linear relationship between an independent (X) and a dependent (Y) variable. If we know the petal width can we estimate the petal length? (The iris dataset is installed as part of the base R installation and can therefore be referenced directly, as in the sample code for this section.)

First, fit a linear model.

```r
model <- lm(iris$Petal.Length ~ iris$Petal.Width)
```

Summarize the model.

```r
summary(model)
```

```
##
## Call:
## lm(formula = iris$Petal.Length ~ iris$Petal.Width)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.33542 -0.30347 -0.02955  0.25776  1.39453
##
## Coefficients:
##                  Estimate Std. Error t value
## (Intercept)       1.08356    0.07297   14.85
## iris$Petal.Width  2.22994    0.05140   43.39
##                  Pr(>|t|)
## (Intercept)        <2e-16 ***
## iris$Petal.Width   <2e-16 ***
## ---
## Signif. codes:
##   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1  ' ' 1
##
## Residual standard error: 0.4782 on 148 degrees of freedom
## Multiple R-squared:  0.9271, Adjusted R-squared:  0.9266
## F-statistic:  1882 on 1 and 148 DF,  p-value: < 2.2e-16
```

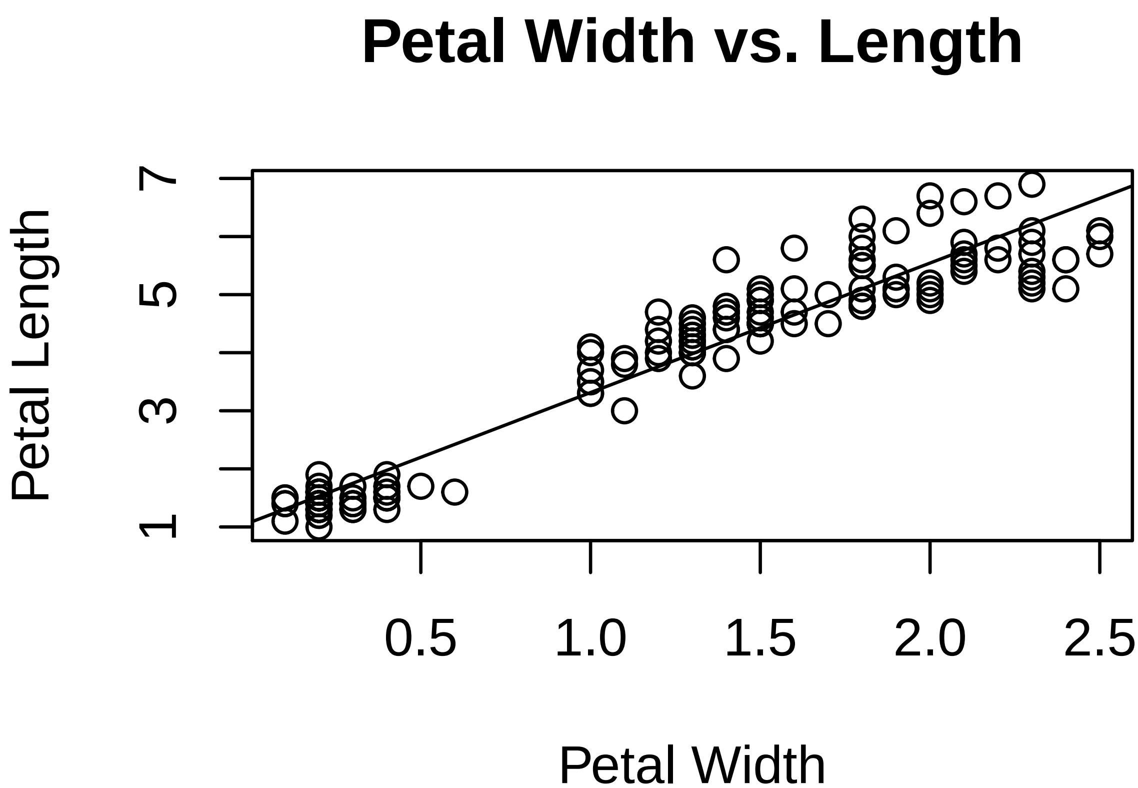
So, our predictive equation is:
Petal Length = 1.08356 + 2.22994(Petal Width)

These results tell us that:

- the intercept is significantly different than 0 ($P<0.05$)
- the slope is significantly different than 0 ($P <0.05$)
- the model is a good fit for the data ($R^2 = 0.9271$). This means that 93% of the variation in petal length is explained by the petal width.

We can also add a linear regression line directly in the plot of the two variables:

```
# par(mar=(c(5,4,4,2)))
plot(iris$Petal.Width, iris$Petal.Length, main = "Petal Width vs. Length",
    xlab = "Petal Width", ylab = "Petal Length")
abline(model)
```


Petal Width vs. Length
Petal Length
1
3
5
7
0.5
1.0
1.5
2.0
2.5
Petal Width