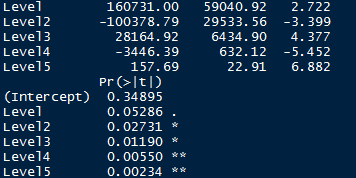
**Business problem**

Create a model that gives us the estimated salary for an employee based on the position.

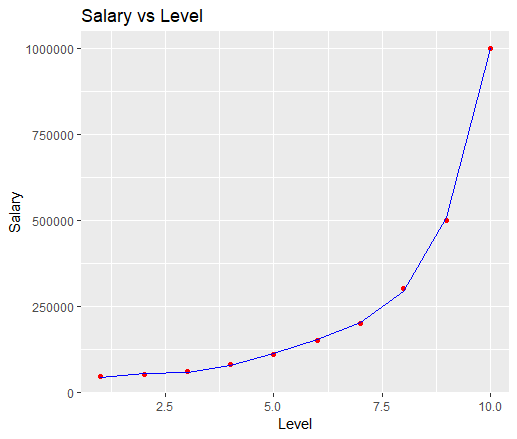
**Explaining the model**

1. Import the dataset, which is in a csv format.
2. We don’t need to encode categorical data because we already have one column in our dataset that represents the position – we subset our data set, just to have one dependent variable and one independent variable.
3. Since it’s a dataset with only 10 rows, there’s not enough data to create a training and teste sets, therefore we will not split.
4. Before creating the regressor, we need to create some columns in our dataset that represent the exponents of our polynomial model. I’ve decided to go until exponent 5, as it’s enough to have a good fitting model. Then, the regressor is created. We want to see how the salary is impacted by the position, so that means our dependent variable is Salary, and the independent is Position. The data we want to use in our regressor is our whole dataset.
5. When we analyze our regressor, we can see that the coefficients of our columns that represent the exponents increase as we increase the exponent.



1. It’s now time to test our model; the first argument is the regressor we’ve created, and the second is the Test set.

**Analyzing the results**

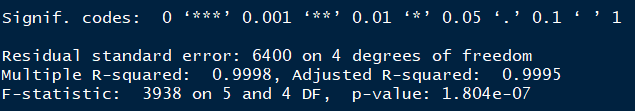


As we can see, our model fits almost perfectly on the real observations.

**Prediction**

Imagining a person has 2 years of experience as a Region Manager (level 6) and needs 2 more to jump to a Partner (7), we can consider it to be a 6.5 regarding the Level. Therefore, our model predicts a salary of $174.878.

**Evaluating the model’s performance**

****

Since it’s a model with more than one independent variable, we’ll look at the Adjusted Multiple R-Squared, which is 0.9998 (the maximum value is 1, so we can conclude that it’s a very robust model).