

Evaluating Regression Models Performance - Homework's Final Part

<pre>Call: lm(formula = Profit ~ R.D.Spend + Administration + Marketing.Spend + State, data = dataset) Residuals: Min 1Q Median 3Q Max -33504 -4736 90 6672 17338 Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 5.008e+04 6.953e+03 7.204 5.76e-09 *** R.D.Spend 8.060e-01 4.641e-02 17.369 < 2e-16 *** Administration -2.700e-02 5.223e-02 -0.517 0.608 Marketing.Spend 2.698e-02 1.714e-02 1.574 0.123 State2 4.189e+01 3.256e+03 0.013 0.990 State3 2.407e+02 3.339e+03 0.072 0.943 --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 9439 on 44 degrees of freedom Multiple R-squared: 0.9508, Adjusted R-squared: 0.9452 F-statistic: 169.9 on 5 and 44 DF, p-value: < 2.2e-16</pre>	<pre>Call: lm(formula = Profit ~ R.D.Spend + Administration + Marketing.Spend, data = dataset) Residuals: Min 1Q Median 3Q Max -33534 -4795 63 6606 17275 Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 5.012e+04 6.572e+03 7.626 1.06e-09 *** R.D.Spend 8.057e-01 4.515e-02 17.846 < 2e-16 *** Administration -2.682e-02 5.103e-02 -0.526 0.602 Marketing.Spend 2.723e-02 1.645e-02 1.655 0.105 --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 9232 on 46 degrees of freedom Multiple R-squared: 0.9507, Adjusted R-squared: 0.9475 F-statistic: 296 on 3 and 46 DF, p-value: < 2.2e-16</pre>
<pre>Call: lm(formula = Profit ~ R.D.Spend + Marketing.Spend, data = dataset) Residuals: Min 1Q Median 3Q Max -33645 -4632 -414 6484 17097 Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 4.698e+04 2.690e+03 17.464 <2e-16 *** R.D.Spend 7.966e-01 4.135e-02 19.266 <2e-16 *** Marketing.Spend 2.991e-02 1.552e-02 1.927 0.06 . --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 9161 on 47 degrees of freedom Multiple R-squared: 0.9505, Adjusted R-squared: 0.9483 F-statistic: 450.8 on 2 and 47 DF, p-value: < 2.2e-16</pre>	<pre>Call: lm(formula = Profit ~ R.D.Spend, data = dataset) Residuals: Min 1Q Median 3Q Max -34351 -4626 -375 6249 17188 Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 4.903e+04 2.538e+03 19.32 <2e-16 *** R.D.Spend 8.543e-01 2.931e-02 29.15 <2e-16 *** --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 9416 on 48 degrees of freedom Multiple R-squared: 0.9465, Adjusted R-squared: 0.9454 F-statistic: 849.8 on 1 and 48 DF, p-value: < 2.2e-16</pre>

Machine Learning A-Z

© SuperDataScience

- This is the backward elimination for our last exams.
- Here we implemented four models.
- In model 3 we exclude 'Marketing. Spend' because its p value (0.06) is bigger than 0.05 and it's not so significant but we want to be sure that its exclusion improved our model.
- R squared is biased
- By adding more variables, the R squared will grow too. It is shown in our models too.
- Adjusted R-Squared has the same characteristic but with this difference that it has the penalization factor.
- Our first three models' R-Squared changes only by a small fraction. In this case the penalization factor (our Adjusted R-Squared) overwhelmed this growth. In our case, the Adjusted R-Squared is decreased despite the fact that our variable increased.
- The Adjusted R-Squared in model 3 is bigger than the model 4 so it indicated that the model 4 is worse than our model 3.
- The model 3 is the best one.