Predictive Analytics
Regression and Classification

Lecture 2 : Part 3

Sourish Das

Chennai Mathematical Institute

Aug-Nov, 2020

How to compare two models?

Model 1 $\texttt{mpg}=\beta_0 + \beta_1\texttt{wt}+\epsilon$

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	37.2851	1.8776	19.8576	0
wt	-5.3445	0.5591	-9.5590	0

Model 2 $\texttt{mpg}=\beta_0 + \beta_1\texttt{wt}+\beta_2\texttt{hp}+\epsilon$

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	37.2273	1.5988	23.2847	0.0000
wt	-3.8778	0.6327	-6.1287	0.0000
hp	-0.0318	0.0090	-3.5187	0.0015

Model 3 $\texttt{mpg}=\beta_0 + \beta_1\texttt{hp}+\beta_2\texttt{hp}^2 + \epsilon$

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	40.4091	2.7408	14.7438	0e+00
hp	-0.2133	0.0349	-6.1148	0e+00
I(hp^2)	0.0004	0.0001	4.2746	2e-04

compare models: RMSE

- The main purpose of the predictive model is to make accurate prediction. So we compare them based on their prediction accuracy.

- **Root Mean Square Error**

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2},$$

where y_i actual target value, and $\hat{y}_i$ predicted value.

- You choose the model with lower RMSE.

compare models: coefficient of determination

- R-**squared (the coefficient of determination)**

  $$R^2 = 1 - \frac{SS_{res}}{SS_{tot}},$$

  where $SS_{res} = \sum_i (y_i - \hat{y}_i)^2$, $SS_{tot} = \sum_i (y_i - \bar{y})^2$,
  $\bar{y} = \frac{1}{n} \sum_i y_i$ and $SS_{reg} = \sum_i (\hat{y}_i - \bar{y})^2$
  Choose the model with higher R^2.

- "**R squared**" is the proportion of the variance of the target variable that is predictable from the feature variable(s).

compare models: coefficient of determination

Result For OLS estimator, $SS_{tot} = SS_{reg} + SS_{res}$, and $0 \leq R^2 \leq 1$.

Note : The result is specific for OLS estimator. For other estimators like Bayes or LASSO, or Bootstrap estimator this result is not necessarily true.

compare models: adjusted R-squared

- In least squares regression, the **R-squared** increases in the number of feature increase the value of R^2.

- **R-squared** alone cannot be used for comparison of models with very different numbers of feature variables.

- In order to solve the problem of **R-squared**, the **adjusted R-square** are used:

$$R^2_{adjusted} = 1 - (1 - R^2)\frac{n-1}{n-p-1},$$

  where p is the number of features in the model and n is the sample size.

compare models: Akaike Information criteron

- AIC is founded on information theory.
- AIC is defined as:

$$AIC = 2p - 2\ln(L(\hat{\beta}|\mathbf{y}, \mathbf{X})),$$

  where p is the number of features; $L(\hat{\beta}|\mathbf{y}, \mathbf{X})$ is the likelihood function of regression model evaluated at MLE or OLS estimator of β.

- Given a set of models, our preferred model is the one with the minimum AIC value.

Result Show that OLS estimator of β is also MLE.

compare models: Bayesian Information criteron

- Like AIC, BIC also is founded on information theory.

- BIC is defined as:

$$BIC = p \ln(n) - 2 \ln(L(\hat{\beta}|\mathbf{y}, \mathbf{X})),$$

  where p is the number of features; n is the sample size, $L(\hat{\beta}|\mathbf{y}, \mathbf{X})$ is the likelihood function of regression model evaluated at MLE or OLS estimator of β,

- Given a set of models, our preferred model is the one with the minimum AIC value.

Result Show that OLS estimator of β is also MLE.

compare models

Model 1 $\texttt{mpg}=\beta_0 + \beta_1\texttt{wt}+\epsilon$

RMSE = 2.95 , R-sqrd = 0.75 , adj R-sqrd = 0.74

AIC = 166.0294 BIC= 170.4266

Model 2 $\texttt{mpg}=\beta_0 + \beta_1\texttt{wt}+\beta_2\texttt{hp}+\epsilon$

RMSE = 2.47 , R-sqrd = 0.83 , adj R-sqrd = 0.81

AIC = 156.6523 BIC= 162.5153

Model 3 $\texttt{mpg}=\beta_0 + \beta_1\texttt{hp}+\beta_2\texttt{hp}^2 + \epsilon$

RMSE = 2.93 , R-sqrd = 0.76 , adj R-sqrd = 0.74

AIC = 167.6023 BIC= 173.4652

- Out of the three models, which one you would like to choose and why?
- Yes - we will like to choose the Model 2.

compare models

Model 1 mpg=$\beta_0 + \beta_1$wt+ϵ

```
RMSE =  2.95 ,  R-sqrd =  0.75 , adj R-sqrd =  0.74
AIC =  166.0294  BIC= 170.4266
```

Model 2 mpg=$\beta_0 + \beta_1$wt+$\beta_2$hp+ϵ

```
RMSE =  2.47 ,  R-sqrd =  0.83 , adj R-sqrd =  0.81
AIC =  156.6523  BIC= 162.5153
```

Model 3 mpg=$\beta_0 + \beta_1$hp+$\beta_2$hp$^2 + \epsilon$

```
RMSE =  2.93 ,  R-sqrd =  0.76 , adj R-sqrd =  0.74
AIC =  167.6023  BIC= 173.4652
```

- Out of the three models, which one you would like to choose and why?
- Yes - we will like to choose the Model 2.

In the next part of this lecture...

- We will try to understand the complexity of models.