Quiz 6
MSAN 601

October 7, 2016

Question 1 (6 pts)

Suggest two criteria for model selection. Compare and contrast their merits.

Answer

1. Mallows C_p criterion:
   The value of Mallows C_p criterion can be calculated as $C_p = \frac{SSE_p}{MSE(X_1,X_2,....,X_{p-1})} - (n - 2p)$

   The expected value of C_p, when there is no bias, is approximately equal to p, where the number of predictors in the model is p - 1.

   Models with $C_p >> p$ are said to have bias and the ones with $C_p < p$ have no bias.

   The criterion is easy to calculate and interpret, however, it requires carefully choosing the predictors and their form as well as the interaction terms so that the MSE is unbiased.

2. PRESS:
   PRESS estimates SSE in a way that while calculating $\hat{Y}_i$, the i^{th} term is not included. A small PRESS indicates a smaller SSE and suggest that the model is good. However, this method requires calculation of n regression models.

Question 2 (4 pts)

What is a leverage value? What is the typical quantitative measure by which to identify leverage values? Is a leverage value an outlier? Is a leverage value influential?

Answer

A leverage value for i^{th} data point is the distance between its X value and the mean of all X values. Typically, a high value greatly affects the prediction of $\hat{Y}$. A value greater than 0.5 is considered high whereas a value between 0.2 and 0.5 is considered moderate. A leverage value greater than $2p/n$ is considered to be an outlier in X, where $p-1$ is predictors and n is the number of observations. A leverage value need not always be influential.

Question 3 (4 pts)

Are all outlying cases with respect X and Y influential? How would one measure influence?

Answer

No, not all outlying cases influential. Measures such as $DFFITS$, $Cook's\ Distance$ and $DFBETAS$ can be used to determine the influence.

Question 4 (3 pts)

Explain how VIF is used to assess multicollinearity. Walk through the mechanics (not necessarily using math, but address the underlying mathematical assessment of multicollinearity).

Answer

VIF is calculated for a predictor by regressing it over other $p-2$ predictors. It has a term $1-R^2$ in its denominator where R^2 is the coefficient of multiple determination for that predictor. If there exists no multicollinearity, VIF is 1.

In case there exists a linear relation between the predictors, we obtain VIF > 1. A very high (infinite) value indicates perfect multicollinearity. Usually, a value up to 10 is acceptable.

Question 5 (6 pts)

What issue(s) arise with autocorrelated residuals? Write out the null and alternative hypotheses for the Durbin-Watson test in the context of first-order autoregressive model. Will D be small or large with first-order autocorrelated observations? Explain.

Answer

If autocorrelation exists, the regression coefficients lose the property of having minimum variance among other estimators. MSE and $s\{b_k\}$ can underestimate the variance of error terms and true standard deviation of the estimated regression coefficients respectively. It also affects results of t and F-tests and confidence interval estimation.

Durbin-Watson test too test positive autocorrelation:
First-order autoregressive model: $Y_t = \beta_0 + \beta_1 X_t + \epsilon_t$, where, $\epsilon_t = \rho\epsilon_{t-1} + u_t$
H_0: $\rho = 0$ (no autocorrelation), H_a: $\rho > 0$ (autocorrelation exists)

For interpreting the test statistic D, upper and lower bounds are obtained. H_0 is rejected if D is less than the lower bound and cannot be rejected is D is greater than the upper bound.

Small values of D indicate positive autocorrelation among the error terms.

Question 6 (4 pts)

Give an example of when you might choose to use ridge regression over linear regression. Discuss some of the limitations of ridge regression.

Answer

Ridge regression is a good substitute for linear regression when the number of predictors is high and there is a possible multicollinearity.

Limitations: the estimators obtained are biased and are heavily penalized, thus, reducing their magnitude; in ridge, ordinary inference procedures cannot be applied and so the distributional properties cannot be known; and choice of the penalty term/ biasing constant is subjective.

Question 7 (6 pts)

What is the purpose of the added-variable plot? The fitted regression line of an added-variable plot always goes through which point? Be very specific and explain. Provide an example of another measure that can be used in lieu of the added-variable plot.

Answer

Added-Variable plots provide information about the marginal importance of a predictor variable X_k, given the other predictor variables are already in the model.

The fitted regression line of an added-variable plot always goes through the origin. In this case, the response and predictor are residuals (of Y on other predictors and X on other predictors). Expected value of residuals is always 0 and we know that the regression line passes through the point $(\bar{X}, \bar{Y})$.

We can also use the coefficient of partial determination to get information about the marginal contribution of predictors.

Question 8 (6 pts)

You are provided with a standardized linear regression model:

$$\hat{Y}^* = b_1 X_1^* + b_2 X_2^*$$

where the ${b_1}^*$ indicates a standardized regression model. Interpret b_1^*.

Answer

An increase of one standard deviation of X_1 leads to an increase of Y by b_1^* on average.