Influence Diagnostics

Shibendra Kumar Singh (BS2015)

1 DFFIT(S)

One of the important quantity is fitting of the model.
We compare the "difference of fit(s)", $\hat{\vec{y_i}} - \hat{\vec{y_i}}(i)$, which is the change in the predicted value for a point, also known as $DFFIT_i$, where $\hat{\vec{y_i}}$ and $\hat{\vec{y_i}}(i)$ are the prediction for *point* i with and without *point* i included in the regression. We want to predict $\hat{\vec{y_i}}(i)$.
The difference, $DFFIT_i$ shows how influential a point is in a statistical regression. If $DFFIT_i$ is large then the i^{th} quantity of $\hat{\vec{y_i}}$ shows too much influence on the fit. In order to measure this, we first need to *Studentize* the $DFFIT$ and it is obtained as:
$DFFITS = \frac{DFFIT_i}{\hat{\sigma}(i)\sqrt{h_i}}$, ***where*** $\hat{\sigma}(i)\sqrt{h_i} = S.E.(\hat{\vec{y_i}} - \hat{\vec{y_i}}(i))$, $\hat{\sigma}(i)$ is the standard error estimated without the *point* i, and h_i is its leverage point, i.e., h_i is the i^{th} diagonal entry of the $ortho-projection\ matrix, h = X(X'X)^{-1}X'$.

2 COVFIT(S)

The next important quantity is to estimate the covariance matrix which is the standard error matrix of $\hat{\vec{\beta}}$.
The change in the determinant of the covariance matrix of the estimates by deleting the i^{th} observation is given by $COVFIT_i$ and is measured as:
$COVFIT_i = \frac{|\hat{\sigma^2}(X'X)^{-1}|}{|\hat{\sigma^2}(i)(X(i)'X(i))^{-1}|}$
Since $COVFIT_i$ is the ratio of two numbers, if it is far away from 1 then we suspect that the *point* i is a very $influential\ point$.