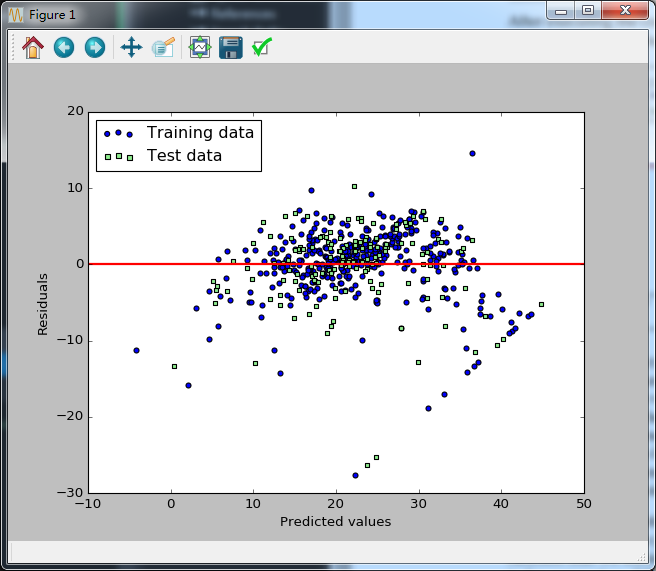
Evaluating the performance of linearregression models

1. Use all variables in the dataset and train a multiple regression model:
2. Plot the residuals(the differences or vertical distances between the actual and predicted values) versus the predicted values to diagnose our regression model. Those **residual plots** are a commonly used graphical analysis for diagnosing regression models to detect nonlinearity and outliers, and to check if the errors are randomly distributed



In the case of a perfect prediction, the residuals would be exactly zero, which we will probably never encounter in realistic and practical applications. However, for a good regression model, we would expect that the errors are randomly distributed and the residuals should be randomly scattered around the centerline. If we see patterns in a residual plot, it means that our model is unable to capture some explanatory information, which is leaked into the residuals as we can slightly see in our preceding residual plot. Furthermore, we can also use residual plots to detect outliers, which are represented by the points with a large deviation from the centerline

Another useful quantitative measure of a model's performance is the so-called **Mean Squared Error** (**MSE**), which is simply the average value of the SSE cost function that we minimize to ft the linear regression model. The MSE is useful to for comparing different regression models or for tuning their parameters via a grid search and cross-validation:



Reference:

1. Python machine learning