



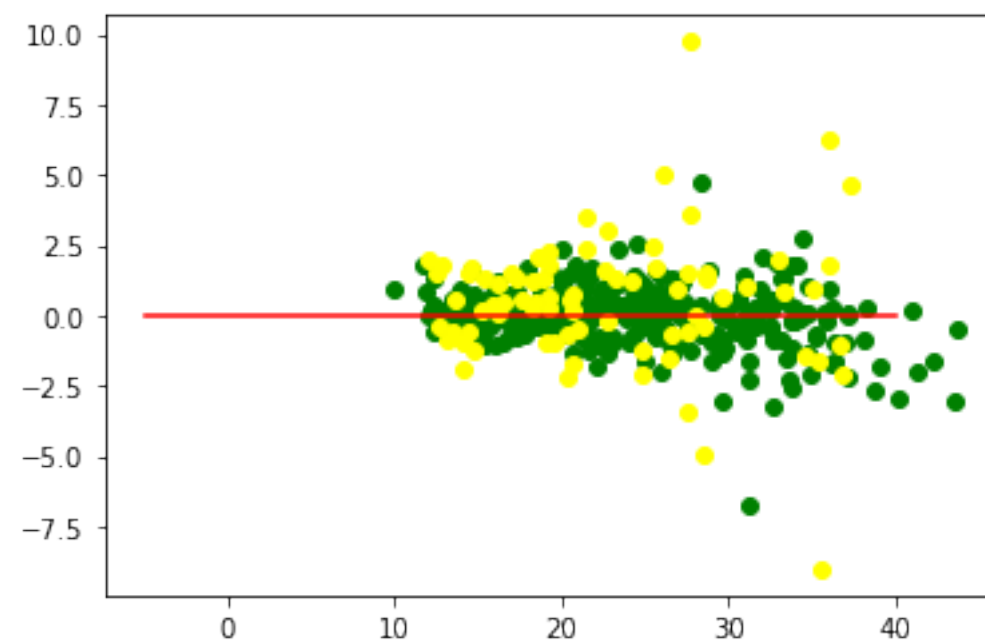
Mean Squared Error: 5.6100
RMSE: 2.3685
Mean Absolute Error: 1.6553

```
In [27]: # Diagnostic Plot (errors vs. predicted)
%pylab inline
#Predicted vs. errors plot -> demonstrates an issue with this fit (high bias)
plt.scatter(rf_reg.predict(X_train), rf_reg.predict(X_train)-y_train, color = "green")
plt.plot([-5,40],[0,0], color = "red")

#place testing data on the plot as well
plt.scatter(rf_reg.predict(X_test), rf_reg.predict(X_test)-y_test, color = "yellow")
```

Populating the interactive namespace from numpy and matplotlib

Out[27]: <matplotlib.collections.PathCollection at 0x7f9f1bbcd940>



MSE, RMSE and MAE are lower for Random forest in comparison to Linear regression for Train data. However, test data seems to perform worse than train data. This could have been a result of overfitting during training.

As indicated by the graph as well, the random forest predicts with higher accuracy in comparison to Linear regression at least on train data. However, the model on the test data seems to be performing almost similarly.