Project 2 Report

After import train and test data sets, change the data type of Date. Get several functions which will be used later. Flatten_forecast: covert a date*stores data frame to a data frame with 3 columns(date, store, weekly_sale). Update_forecast: add forecasts to the one fold test data frame. Update_test: update forecasts in the global data frame. Shift: designed for fold 5 which contains Christmas day. Shift last five rows in a data frame, if the mean of the 2:4 columns divided by the mean of the 1 and 5 columns bigger than a threshold, then shift 1/7 of the sales.
Naïve model: Simply predict the sale with the last observation in the train data.
Linear model: Assume the sale follows a linear combination of trend and season.
My predict: After the first fold, append the previous periods test data to the current training data, filter test data frame for the month that needs predictions, backtesting starts during March 2011. Get a data frame with (num_test_dates x num_stores) rows. Create the same dataframe for the training data.
There are three models, the first is Naïve model, the second is linear model, the third is linear model with post-process with shift function.
Naïve model is for each test department: filter for the particular department in the training data, Reformat so that each column is a weekly time-series for that store's department. The dataframe has a shape (num_train_dates, num_stores). We create a similar dataframe to hold the forecasts on the dates in the testing window. Then we get the prediction and update global test dataframe.
Linear model is similar to the Naïve model, except instead of using naïve model, we use linear model to get the prediction.
Linear model with post-process is similar to the linear model. After we get the prediction from the linear model, we use shift function to do a shift.

The accuracy is:

fold	model_one	model_two	model_three
1	2078.72587	2042.40149	2042.40149
2	2589.33762	1440.08326	1440.08326
3	2253.93612	1434.71561	1434.71561
4	2823.09826	1596.9877	1596.9877
5	5156.01153	2327.63783	2029.38775
6	4218.34774	1674.18487	1674.18487
7	2269.90424	1718.57696	1718.57696
8	2143.83903	1420.81708	1420.81708
9	2221.14475	1430.8008	1430.8008
10	2372.4248	1447.03367	1447.03367
Overall Average	2812.677	1653.32393	1623.49892

The running time is 9.362102 mins.

The computer system: Macbook Pro, 2.3GHz, i5, 8GB memory.