AMS – 559
Smart Energy in the Information Age

HW 1-Report

Submitted By:
Anuroop Katiyar
112609023

Data Pre-processing:

HomeF 2016 file contained complete data for approximately 348 days. Therefore, to maintain consistency across all the years, I have considered 348 days data across all 3 homes. The HomeF data file had per minute electricity consumption data for which I calculated load demand at half hour intervals for the data to be similar to other data files. Similarly, the weather file has also been split into 30 minutes interval instead of hourly weather data while using the temperature data.

The data has been split as in ratio of 70:20:10 for training, cross-validation and testing respectively.

Features Extraction:

From the load demand file, I have taken the load demand values and time stamp for all the homes and from the weather files I have taken the temperature feature. From the time stamps I have separated the following features:

1. Day – to model day wise energy consumption trends in a month.
2. Month – to model moth wise energy consumption trends in a year.
3. Year – to model year wise energy consumption trends from 2014-2016.
4. Hour – to model per hour energy consumption trends in a day.
5. Weekday – to model weekdays energy consumption trends in a week.
6. Holidays – to model energy consumption during holidays.
7. Temperature – Temperature from weather files is used to predict consumption

Models Used:

1. **Naïve Method:** This method simply predicts the consumption using the latest previous time slots present in training data. The Naïve method provides a good baseline for improvement on predictions using other models.

2. **Random Mean:** In this method I tried my own approach. I tried to randomly pick sample data points from my training set and calculate the mean for the random points selected to predict for one time slot. The results produced were slightly better than the Naïve approach.

3. **Linear Regression:** In this model I have taken the temperature and electricity consumption data to fit a predictive model for the given data sets. The data has been split randomly using linear regression split function.

4. **Random Forest:** Random forest operates by constructing a multitude of decision trees at training time and outputting the class i.e. the mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set.

5. **SARIMA:** Better known as Seasonal Autoregressive Integrated Moving Average is a method for time series forecasting with univariate data containing trend and seasonality. Parameters SARIMA(0,1,0)x(1,1,1,12) were used to make the prediction.

6. **Prophet:** Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. It is robust to missing data and shifts in the trend, and typically handles outliers well. I used the holidays list to improve the prediction using Prophet model.

RESULTS:

A. Data prediction over next 24 hours (48 half hour slots) or one day:

	Models	HomeB-2014(MAE)	HomeC-2015(MAE)	HomeF-2016(MAE)
1	Naive Model	1.59022	0.53233	1.36774
2	Random Mean	0.44830	0.38216	1.22775
3	Linear Regression	0.43976	1.14081	0.47853
4	Random Forest	0.47969	0.49466	1.20841
5	SARIMA	0.43044	0.43514	1.21992
6	Prophet	1.39702	0.34112	1.10706

B. **Data prediction over next hour (2 half hour slots):**

	Models	HomeB-2014(MAE)	HomeC-2015(MAE)	HomeF-2016(MAE)
1	Naive Model	0.05861	0.06549	0.09242
2	Random Mean	0.10327	0.36309	1.95870
3	Linear Regression	0.43976	1.14081	0.47853
4	Random Forest	0.09739	0.03888	0.08052
5	SARIMA	0.17652	0.05574	0.20037
6	Prophet	0.04945	0.07482	0.06769

ANALYSIS:

As expected, Naïve model produced maximum error. Overall, SARIMA and Random Forest performed considerably well as compared to other models with the lowest Mean Absolute Errors.

Note- All the graphs have been attached in separate documents.