CptS 580: Advanced Topics in Machine Learning

Homework #4
Due: Tuesday, November 29, 2011

Background

In the summer of 2011, Avista together with WSU researchers turned Pullman into the region's first smart grid community. As part of the project, electricity usage meters were replaced with "smart meters" that collect detailed usage information that can be analyzed by Avista (and us!) and viewed by Avista customers.

Starting in September 2011, Avista provided the WSU CASAS smart home research team with residential energy usage data. Our goal is to mine the data to find individual and community-wide trends and to correlate in-home activity data with whole-home energy usage data. The data I have sent to you includes the following:

- 8siteshour.arff: Energy usage for 8 building sites (in kilowatt hours aggregated by hour), starting April 22, 2011 and ending September 22, 2011
- 8sitesday.arff: Energy usage for the same 8 building sites and same time range, reported total usage by day
- 8sitesdaytotal.arff: Energy usage for the same time range, reported daily and totaled over the 8 building sites

Part One

The goal of part one is to use Weka to find trends in whole-home energy usage data. We will demonstrate in class how Weka can perform some of these tasks. In this homework assignment you need to use Weka to perform the following forecasting task. For each dataset, provide snapshots of the resulting graphs and/or summarize the performance measures, as appropriate. Here are the initial parameters to use.

- Use the support vector machine for regression algorithm.
- For the hour-based data set generate predictions 1 day into the future. In the case of the day-based data set these predictions should extend 7 days into the future. For each dataset report the root mean squared error and the mean absolute percentage error.
- Set the periodicity to intuitive values for each data set. Good choices would be daily for the hour-based data set and weekly for the day-based data set.

In addition, write a program (in whatever language you choose) to compute moving averages for the 8siteshour.arff dataset, where the window size is 7 days). Add this field to the 8siteshour.arff data and compare the forecast graphs for the raw data and the moving averages data (for site4 and site6).

What did you observe about the individual site data by hour and by day? Are the trends expected given the provided dates? What additional analyses do you think would be appropriate?

Part Two - *optional*

While Weka is an excellent tool, there are several other tools for analyzing time series data. Excel offers a wide variety of analysis methods and RapidMiner offers a large set of time series tools. In the second part of this assignment you will install and use one additional time series analysis tool (Excel, RapidMiner, or another of your choice) and apply the tool to the energy usage data. In particular, perform the following tasks with the new tool:

- Repeat one of the analysis experiments listed in Part One
- Perform one additional analysis (trend analysis or forecasting) not listed in Part One

Write up the results of your analysis. In addition, provide your assessment of the trend analysis tool and compare features of the tool with Weka.

Turning in the Assignment

The assignment should be mailed as a TXT or PDF file to cook@eecs.wsu.edu by 9:00am on the due date. Include your source code, sample output, and summarized results with observations. No late assignments will be accepted.