SOFT COMPUTING FTN

Rossmann Store Sales

Forecast sales using store, promotion, and competitor data

Student : Dimitrije Jankov

Professor : Djordje Obradovic

Problem

Rossmann is challenging us to predict 6 weeks of daily sales for 1,115 stores located across Germany.

Motivation

Reliable sales forecasts enable store managers to create effective staff schedules that increase productivity and motivation. By helping Rossmann create a robust prediction model, you will help store managers stay focused on what’s most important to them: their customers and their teams!

The Data

The data is organized in three files :

- **store.csv - historical data including Sales**
- **test.csv - historical data excluding Sales**

train.csv - supplemental information about the stores

SOFT COMPUTING FTN

The Data Fields

- **Id - an Id that represents a (Store, Date) duple within the test set**
- **Store - a unique Id for each store**
- **Sales - the turnover for any given day (this is what you are predicting)**
- **Customers - the number of customers on a given day**
- **Open - an indicator for whether the store was open: 0 = closed, 1 = open**
- **StateHoliday - indicates a state holiday. Normally all stores, with few exceptions, are closed on state holidays. Note that all schools are closed on public holidays and weekends. a = public holiday, b = Easter holiday, c = Christmas, 0 = None**
- **SchoolHoliday - indicates if the (Store, Date) was affected by the closure of public schools**
- **StoreType - differentiates between 4 different store models: a, b, c, d**
- **Assortment - describes an assortment level: a = basic, b = extra, c = extended**
- **CompetitionDistance - distance in meters to the nearest competitor store**
- **CompetitionOpenSince[Month/Year] - gives the approximate year and month of the time the nearest competitor was opened**
- **Promo - indicates whether a store is running a promo on that day**
- **Promo2 - Promo2 is a continuing and consecutive promotion for some stores: 0 = store is not participating, 1 = store is participating**
- **Promo2Since[Year/Week] - describes the year and calendar week when the store started participating in Promo2**
- **PromoInterval - describes the consecutive intervals Promo2 is started, naming the months the promotion is started anew. E.g. "Feb,May,Aug,Nov" means each round starts in February, May, August, November of any given year for that store**

Similar problems

- **Bayesian learning for sales rate prediction for thousands of retailers**
- **Newspaper Vendor Sales Prediction Using Artificial Neural Networks.**
- **Sales Forecasting for Fashion Retailing Service Industry: A Review**

Implementation steps

- Visualize the data
- Make a large set of features
- Determine the importance of each feature
- Make a prediction model (Support vector machine, Gradient boosting tree…)
- Evaluate the results

Thank you for your patiance