**A Shiny demonstration of Bagged Loess regression on variable number of Bootstrap sample.**

**Introduction:**

The very basic idea of bagging technique in prediction analysis is [1]:

1. To resample (bootstrap [4]) cases and recalculate predictions;

2. To average the results by taking means or majority vote.

Bagging is especially useful in predicting non-linear dependencies. It gives similar bias than other biased estimators but reduces variance.

An example of applying bagging to Loess [3] prediction is given. The purpose of this example was to show how changing the number of bootstrap samples affects the smoothiness of the resulting prediction curve.

**Methods:**

*Data Collection*

Variables that play a key role in meteorology are ozone level, Sun radiation, wind and air temperature. Dataset 'ozone' relates these four quantities together and was taken from 'ElemStatLearn' R-package. [2].

*Statistical Modeling*

To relate ozone level to air temperature depth a standard Loess regression model [3] was used. Then the following procedure was performed:

1. The 'ozone' dataset rows were resampled with replacement.

2. Loess regression models were calculated on bootstrap samples.

3. Bootstrap Loess Models were applied to the entire dataset.

4. Predicted air temperature values were stored in a matrix.

5. The averages of the predicted values against the ozone level values were calculated.

6. The original data, bootstrapped Loess curves and the averaged (bagged) prediction curve were superimposed on a single ozone-temperature plot.

*Demonstration*

The above described approach was demonstrated by a Shiny application.

The application contains two R files: UI.R and Server.R located in its folder.

UI.R defines a user interface - HTML page layout with two separate areas.

The input area (left side) contains a slider control that the user can use to set the number of bootstrap samples. The number of samples varies from 5 to 50 with the default value of 10 and step 5.

The output area (right side) contains a scatter ozone-temperature plot along with bootstrapped Loess curves (grey color) and their bagged average (red color) .

In order to run this application the user should perform the following:

1. Install the following R packages: shiny, ElemStatLearn:

install.packages('shiny')

install.packages('ElemStatLearn')

2. Set the working folder to the location with the downloaded UI.R and Server.R files.

3. Start the application using the following R commands:

library(shiny)

runApp()

4. Explore different bagged Loess prediction models by varying the number of Bootstrap

samples with the slider control.

**Conclusions:**

Our demonstration suggests that using bagging technique in statistical prediction modelling can significantly reduce prediction variance while keeping appropriate level of bias for modelling non-linear dependencies.

**References**

1. Coursera, John Hopkins University “Practical Machine Learning” URL: <https://class.coursera.org/predmachlearn-003/auth/auth_redirector?type=login&subtype=normal>

2. "The Elements of Statistical Learning” Page. URL: <http://statweb.stanford.edu/~tibs/ElemStatLearn/> Accessed 8/14/2014

3. Wikipedia, Local regression (LOESS) Page. URL: <http://en.wikipedia.org/wiki/LOESS>

4. Wikipedia, Bootstrapping (statistics) PAge. URL: <http://en.wikipedia.org/wiki/Bootstrapping_%28statistics%29> Accessed 8/14/2014

5. R Core Team (2014). ”R: A language and environment for statistical computing.” URL: [http://www.R-project.org](http://www.r-project.org/)

6. Shiny by RStudio. Tutorial, articles, documentation. URL: <http://shiny.rstudio.com/> Accessed 8/13/2014