Maths6397 HW3 Due on Tuesday 4/12/2016

Part1 : DATA oil.gas.xlsx

Prices are listed for 1411 days Separate data into training set (first 1000 days) and a test set (last 411 days)

Use only the training data to perform 2 distinct standard linear regressions to predict separately 2 output variables "En" and "In" on day n

E = Conventional Gasoline Prices

I = Heating Oil Prices

using the same 4 Input variables Bn, Fn, Gn, Hn
B=Crude Oil Prices: West Texas ; F= Propane Prices: Mont Belvieu, Texas,
G= Henry Hub Natural Gas Spot Price; H= Crude Oil Prices: BrentEurope

Give the coefficients of the 4 regressions including the constant terms.

For each output variable Yn, call predYn the predicted value computed by the regression . Plot on the same graph the curves Yn and predYn to visualize the prediction errors

For each output variable Y compute separately the standard deviations STDtrain and STDtest of the prediction errors on the training set and on the test set

For each output variable Y compute mY = mean value of Y, and evaluate regression performance by STDstrain/mY and STDtest/mY

Part 2 DATA metals.xlsx

Prices are listed for 838 days 'Separate data into training set (first 600 days) and a test set (last 238 days)

Use only the training data to perform a **polynomial kernel non linear regression** to predict on day "n" the **future** output variable "J_{n+1}" = price of aluminium on day "n+1"

The input variables will be the following 4 day "n" values
Dn= steel price ; En= nickel price; In= copper price; Jn= aluminium price

All variables D,E,I,J should first be normalized by the rescaling D/mD , E/mE, I/mI, J/mJ where mD, mE, mI, mJ are the respective mean values computed on the training set

Try only a polynomial kernel of degree r = 2.

For each day n, call $predJ_{n+1}$ the predicted value of J_{n+1} computed by the nonlinear regression using the input values Dn En In Jn .

Explain as explicitly as possible how $predJ_{n+1}$ is computed in terms of Dn En In Jn .

Plot on the same graph the curves J_{n+1} and $predJ_{n+1}$ to visualize the prediction errors

Compute separately the standard deviations STDtrain and STDtest of the prediction errors on the training set and on the test set

Evaluate regression performance by STDstrain/mJ and STDtest/mJ

Part 3: Same data as Part2

Same problem as part 2 but implement a **non linear kernel ridge regression** (polynomial kernel of degree 2.) Optimize the performance by a proper choice of the coefficient λ involved in the ridge regression