CL 688: AI in process engineering
Coding Assignment 1

PROBLEM I. Here, we want to study least squares and regularized least squares.

Write a code to sample N=10 points from a corrupted sinusoid

$$y=\sin(2\pi x)+\gamma N(0,1)$$

where γ is a multiplier that governs the size of the noise in the signal y .
Also, obtain an independent test set.

1) Fit a polynomial of order M by minimizing the squares of the prediction errors
2) Study effect of changing M in the range [0 15] for the training set based on $N=10, \gamma=0.1$. Obtain the SSE for the test set also and plot sum of squared prediction errors (SSE) for the training and test set with respect to M. What did you conclude
3) Study effect of changing size of the training set N in the range of [10 1000] for $M=9, \gamma=0.1$. Obtain the SSE for the test set also and plot SSE for the training and test set with respect to N. What did you conclude
4) Study effect of change in the amplitude of the noise mulitplier γ in the range of [0.1 0.5] for M= 10, N=100. Obtain the SSE for the test set also and plot SSE for the training and test set with respect to γ . What did you conclude

PROBLEM II. Here, we want to repeat the above exercise using regularized least squares. Assume $N=10, M=9, \gamma=0.1$ as default values. As a first step, obtain a plot of SSE for various values of the regularization parameter %lambda in the range [0 1]
Then repeat sub-problems 1-4 above.

PROBLEM III. Theory tells us that minimization of SSE is essentially MLE with Gaussianity and regularization is MAP with Gaussianity. For a truly Bayesian approach we should marginalize the predictions as

$$p(t|x,\mathbf{x},\mathbf{t}) = \int p(t|x,\mathbf{w})p(\mathbf{w}|\mathbf{x},\mathbf{t})\,\mathrm{d}\mathbf{w}$$

Using this approach find mean of the above density when the new data points *x* correspond to the test set

Please upload your solution in form of a pdf file on Moodle
Deadline: Sept 11,2020