Homework 7: Locally Weighted Regression

Due October 25, 2018 at 2:10 PM

This homework will investigate locally weighted regression models to predict the percent body fatgiven a variety of information about body measurements and estimated body fat percentage.

The file hwkdataNEW.mat contains data related to body measurements and estimated body fat percentage. The matrix x contains 14 predictor variables:

1. Age (yrs)
2. Weight (lbs)
3. Height (inches)
4. Adiposity Index (kg/m^2)
5. Neck circumference (cm)
6. Chest circumference (cm)
7. Abdomen circumference (cm)
8. Hip circumference (cm)
9. Thigh circumference (cm)
10. Knee circumference (cm)
11. Ankle circumference (cm)
12. Extended bicep circumference (cm)
13. Forearm circumference (cm)
14. Wrist circumference (cm)

The column vector y contains the output variable, percent body fat. The correct citation for these data is: Penrose, K., Nelson, A., and Fisher, A. (1985), "Generalized Body Composition Prediction Equation for Men Using Simple Measurement Techniques", *Medicine and Science in Sports and Exercise*, 17(2), 189.

1. Briefly describe the locally weighted regression algorithm. Discuss the pros and cons of locally weighted regression over ordinary least squares regression. Discuss how weights are determined (Gaussian kernels) and how the kernel bandwidth is optimized. Discuss the importance of standardization for the weight calculation.

2. Divide the data into training, test, and validation data sets. You **must** use the same training, test, and validation data sets that you used in previous assignments.

3. Build three different locally weighted polynomial regression models: zero order (kernel regression), first order (linear regression), and second order (quadratic regression). For each model, find the optimal kernel bandwidth from values of $h = [0.15, 0.25, 0.5, 0.75, 1.0, 1.5, 2.0]$ to minimize the test RMSE.

4. Compare the validation performance of your three LWR models with all the models we've looked at before. Comment on the results.

For this homework, prepare a written report in IEEE format. Include any plots and tables that will support your findings. Make sure you correctly label your figures and tables and refer to them in the text. Include an appropriate citation for the data, both in the text and in the list of references. Your report should include **at a minimum** an abstract, introduction, methodology, results (and discussion!), conclusions, and references. Note that the methodology section of this report (and every report!) should describe the algorithm that you're using – not the implementation in MATLAB. Include all your code in an appendix (single column) at the end of the report. Convert your report to .pdf before submitting it through Canvas.