**DSI -06 Homework 6:**

Introduction with Statistical Learning with Applications to Python (ISLP) pg. 327

9. This question uses the variables dis (the weighted mean of distances to five Boston employment centers) and nox (nitrogen oxides concentration in parts per 10 million) from the Boston data. We will treat dis as the predictor and nox as the response.

1. Use the poly() function from the ISLP.models module to fit a cubic polynomial regression to predict nox using dis. Report the regression output, and plot the resulting data and polynomial fits.
2. Plot the polynomial fits for a range of different polynomial degrees (say, from 1 to 10), and report the associated residual sum of squares.
3. Perform cross-validation or another approach to select the optimal degree for the polynomial, and explain your results.
4. Use the bs() function from the ISLP.models module to fit a regression spline to predict nox using dis. Report the output for the fit using four degrees of freedom. How did you choose the knots? Plot the resulting fit.
5. Now fit a regression spline for a range of degrees of freedom, and plot the resulting fits and report the resulting RSS. Describe the results obtained.
6. Perform cross-validation or another approach in order to select the best degrees of freedom for a regression spline on this data. Describe your results.

Additional Practice Questions:

Explain the importance of a training and test set to a non-technical audience.

How would you describe this exercise in an interview to both a technical and non-technical interviewer? What are the key insights you would want to show?

Can you think of a business context where this exercise would have applications?