Code

Problem Set 2 - Nonparametric Regression

Rich Sweeney & Charlie Murry

Deregulated electricity markets solve a series of interellated, high-frequency auctions to ensure that supply equals demand at all points in time and space. The dataset `nodelmp.csv` contains one year of locational marginal prices (LMP's) for a node in Brighton. [If you're interested, more info here (https://www.iso-ne.com/participate/support/faq/lmp)].

In this problem you will explore the relationship between these prices and temperature (var: `temp`).

Code

1. Make a binner scatter plot of the relationship between the maximum price each day and the temperature. Does the relationship look linear?

Code

2. Predict the relationship using a polynomial in temperature, for polynomials varying from degree $p = 1$ to $p = 10$. Which fits best *in sample*?

Code

3. Repeat the above exercise, using k-fold cross validation to select between polynomials of degree 1 through 10. Use $k = 10$. Which p fits best now?

Code

4. Plot the predicted lmp for $p = 1, 2, p^{cv}, 10$, where p^{cv} is your preferred degree from above.

Code

5. Predict lmp using a natural cubic spline in temperature. Use k-fold cross-validation to find the optimal number of knots, ranging from 1 to 10.

Code

6. Finally predict lmp using a lowess/ loess regression (using the default bandwidth/ span).

Plot your preferred polynomial prediction (p^{cv}), your prefered spline prediction, and the lowess prediction.

Code

7. Now plot the residuals for each of these curves against temperature. Which parts of the distribution does each do better/ worse on?

Code