Homework Set 5
Due Tuesday, September 17

I have taken the US census data from 1790 to 1990 and generated a "random subsample." This random subsample produces a noisy set of measurements, and simulates the type of census data we would have if we tried to estimate the total population from a much smaller sample. The US census data is available in Matlab (type '`load census`' in Matlab). The random subsample can be downloaded from `www.ece.rice.edu/~nowak/elec531/us_pop.mat` (use the "save as" feature on your browser). The noisy subsample population data are $\boldsymbol{y} = [y_0, \ldots, y_{20}]^T$ and we want to fit a degree $p \ll N$ polynomial to the data.

1. Assume the "noise" is Gaussian white noise and give expressions for $\mathbf{H}$ and $\boldsymbol{\theta}$ for the corresponding model $\boldsymbol{y} \sim N(\mathbf{H}\boldsymbol{\theta}, \sigma^2 I)$.

2. Verify that the moments $m_k = \sum_{n=0}^{N-1} n^k y_n$, $k = 0, \ldots, p$ are sufficient statistics for this model.

3. Derive a general expression for the projection matrix $\mathbf{P}_H$ and explain how the projection matrix can be used to obtain a polynomial fit to the data.

4. Generate and plot polynomial fits for $p = 1, 2, \ldots, 8$.

5. What value of p seems to provide the best fit to this data? How does your polynomial fit agree with the full census data?