11/5: Visualization and Synthetic Data

Overview

This class discussed the importance of visualization and synthetic data when working with probability models and any real data set.

The main example discussed was the paper/talk by T. Minka "Expectation-Propogation for the Generative Aspect Model."

See the following slides on this topic:

`http://www.stat.cmu.edu/~minka/papers/aspect.html`

Main Points

- Real data is often very difficult to understand directly
- Simplify by creating synthetic data
- Allows testing ideas and controlled experiments
- Visualization is essential—make plots to understand what is going on