Kernel Density Estimators

...your name here...

September 3, 2015

In this lab, we will be exploring the performance of kernel estimators on the distribution shown in figure 1. What follows is a brief introduction to kernel density estimators. If you are familiar with KDEs, feel free to skip ahead to the tasks.

Kernel Density Estimators

Suppose that we have $(X_1, ...X_n)$ drawn iid from some distribution with unknown density f. We will assume that $X_i \in \mathbb{R}$ to keep things simple. Further suppose that we wish to estimate f at the point x. One natural estimator might be

$$\hat{f}(x) = \frac{\# X_i \in \mathcal{N}_\lambda(x)}{n\lambda}$$

where $\mathcal{N}(x)$ is some neighborhood about x of width λ. Repeating over a range of x gives us an idea as to the shape of f. We can get a smoother estimator using a kernel function $K(\cdot)$. A kernel function $K : \mathbb{R} \to \mathbb{R}$ is any function that satisfies $\int_{\mathbb{R}} K(u)du = 1$. Typically, one also chooses K to be symmetric. Some popular choices of kernels include the "box kernel": $K(u) = 1\{|u| \leq 0.5\}$, the "Gaussian kernel": $K(u) = \phi(u)$ where $\phi(\cdot)$ is the standard normal density, and the "cosine kernel": $K(u) = \frac{\pi}{4} \cos\left(\frac{\pi}{2}u\right) \cdot 1\{|u| \leq 1\}$. The kernel density estimator of f at the point x is then given by

$$\hat{f}(x) = \frac{1}{n} \sum_i K_\lambda(X_0 - x) = \frac{1}{n\lambda} \sum_i K\left(\frac{X_0 - x}{\lambda}\right).$$

Note that we have incorporated the notion of neighborhood "width" into our estimator by using a scaled kernel $K_\lambda(u) = \frac{1}{\lambda} K\left(\frac{u}{\lambda}\right)$. Intuitively, λ governs how well our estimate $\hat{f}$ fits a specific sample. Smaller values of λ allow for an estimate that fits a specific dataset well, but result in estimates that are more variable across new samples $(X_1', ...X_n')$ drawn from the same distribution. This tradeoff between producing both estimates that are accurate for a sepcific sample and consistent over new samples is known as the bias variance tradeoff. You will explore this tradeoff below.

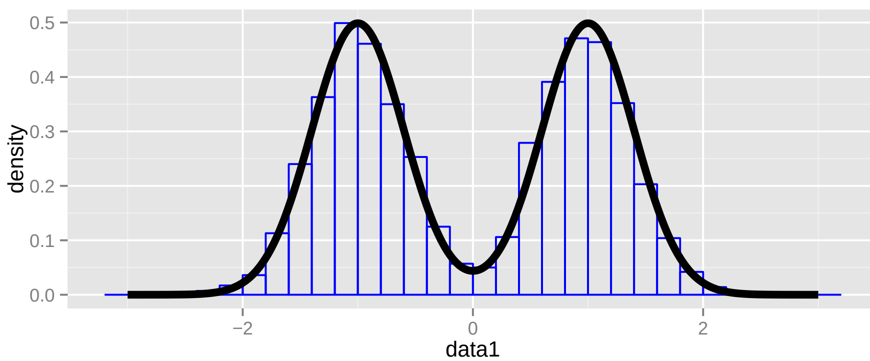

Figure 1: Raw data from a point mixture of normals

Tasks

For all tasks, include a little text in this document about what you are doing. When you're done, you'll have an R library and knitted pdf document.

1. Complete the Kernel function and plot it.
2. Complete the EstimateDensity function and try fitting it to your data with some different bandwidhts.
3. Explore how the bias and variance changes as a function of the bandwidth using the PerformSimulations function.