Homework 3 (due November 8, 2016)

Problem 1: [Clustering:]
Consider the Galaxy data in the previous homework. Using the DPM, obtain the posterior distribution of the number of clusters and an estimate of the clustering configuration using the method described in class.

Problem 2: [Function estimation:]
Generate $x_{ij} \sim U(0,1), i = 1,\ldots,100; j = 1,\ldots,7$ and draw $y_i \sim N(f(x_i),1)$, where $x_i = (x_{i1},\ldots,x_{i7})^T$ and

$$f(x) = 10\sin(\pi x_1 x_2) + 10\cos(\pi(x_3 x_4 + x_5)) + 20(x_6 - 0.5)^2 + 10x_7$$

(i) Fit the function using a Gaussian process with squared exponential covariance kernel $c(x,x') = \exp\{-\kappa||x - x'||^2\}$ with a suitably chosen discrete uniform prior on κ.
(ii) Plot the predicted f vs. x_1 when $x_2 = \cdots = x_7 = 0.5$. Overlay the true f and the 95% pointwise credible intervals.