3.1 (due Mar. 3, 2009)

- Implement the EM algorithm for Gaussian mixtures in one dimension.
- For $K = 2, 4, 8, 16$ and $\sigma = 1/10, 1, 10$ (that is $4 \times 3 = 12$ possibilities) consider the distribution

$$f(x) = \frac{1}{K} \sum_{i=1}^{K} \mathcal{N}(x | i, \sigma).$$

- Generate samples from $f(x)$ and run the EM algorithm on the data. Let $g(x)$ be the mixture learned by the EM algorithm. The $L_1$-distance of $f$ and $g$ is

$$\|f - g\|_1 = \int_{-\infty}^{\infty} |f(x) - g(x)| \, dx.$$ 

How many samples are needed to achieve $\|f - g\|_1 \leq 1/1000$?

- E-mail (by 11:59pm on due Mar. 3, 2009) to the TA and instructor: your code, and answer to the above question for the 12 choices of $K$ and $\sigma$. Clearly describe the methodology used to obtain the answer.