# (520|600).666
# Information Extraction from Speech and Text

*Optional* Homework # 11

Due May 7, 2009.

**Initializations and Local Maxima in the E-M Algorithm:** This problem brings out the fact that the E-M algorithm, at best, converges to a local maximum of the likelihood, and that the point of convergence depends on the starting point. You will have to write a computer program to perform the numerical computations necesary to solve this problem.

Consider a 1-state HMM (!) that produces real-valued outputs $y_1, y_2, \ldots, y_n$, and denote the lone state by $s$. Let the HMM have two output-producing arcs $t_1$ and $t_2$, both from $s$ to $s$, which may be taken with probabilities $p(t_1) = \alpha_1$ and $p(t_2) = \alpha_2 = (1 - \alpha_1)$. Let the output densities be Gaussian.

$$\mathcal{N}(y|t_1) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{(y-\mu_1)^2}{2\sigma_1^2}\right\} \quad \text{and} \quad \mathcal{N}(y|t_2) = \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}.$$

Let $\theta = \begin{bmatrix}\alpha_1 & \mu_1 & \sigma_1^2 & \mu_2 & \sigma_2^2\end{bmatrix}$ denote the parameters of the HMM.

1. Write down the formula for $P_\theta(y_1, \ldots, y_n)$. (Hint: this HMM is just a mixture density.)

2. Plot $P_\theta(y_1)$ for $\theta = \begin{bmatrix}\frac{1}{3} & -2 & 1 & 2 & 1\end{bmatrix}$.

3. Implement the E-M algorithm for updating $\theta$ for this HMM. Make your implementation flexible, so that some parameters may be tied. E.g. we will experiment with arbitrarily fixing the variances and estimating only the means and $\alpha_1$, or with setting a minimum (floor) on the estimate of the variance, etc.

4. Let the observed data comprise the following $n = 25$ values.

| | | | | |
|---|---|---|---|---|
| +0.608 | −1.590 | +0.235 | +3.949 | −2.249 |
| +2.704 | −2.473 | +0.672 | +0.262 | +1.072 |
| −1.773 | +0.537 | +3.240 | +2.400 | −2.499 |
| +2.608 | −3.458 | +0.257 | +2.569 | +1.415 |
| +1.410 | −2.653 | +1.396 | +3.286 | −0.712 |

Set $\sigma_1^2 = \sigma_2^2 = 1$ and $\alpha_1 = \frac{1}{3}$, and obtain the sequence of iterates $\left[\mu_1^{(k)}, \mu_2^{(k)}\right]$ starting with different values $\left[\mu_1^{(0)}, \mu_2^{(0)}\right]$. Try several different starting points in each of the 4 quadrants in the region $[-4, +4] \times [-4, +4] \subset \mathbb{R}^2$.

Plot the "trajectory" of the parameters in the $x$-$y$ plane for a few illustrative starting points.

5. How many different points of convergence did you encounter in Step 4? Comment on the value of the (local) maxima at these points, and speculate upon their *regions of attraction*. By region of attraction of a point of convergence, we mean the *set* of starting points from which the E-M algorithm converges to that particular point.

6. Use a program such as MATLAB or R to plot the surface corresponding to the likelihood of the 25 data-points of Step 4 as a function of $[\mu_1, \mu_2] \in [-4, +4] \times [-4, +4]$, with $\alpha_1$, $\sigma_1^2$ and $\sigma_2^2$ fixed as above.

   Comment on the trajectories you plotted in Step 5 in light of this surface plot, and revise your speculation about the regions of attraction if necessary.

7. Repeat Step 4 but this time, make $\alpha_1$ also a free parameter. In particular, choose the same initial values of $\left[\mu_1^{(0)}, \mu_2^{(0)}\right]$ that you used to plot the illustrative trajectories of $\left[\mu_1^{(k)}, \mu_2^{(k)}\right]$ in Step 4, and plot the (new) trajectories when $\alpha_1^{(0)} = \frac{1}{3}$, but it is updated at every iteration instead of being held fixed. Comment on the point of convergence, as well as the (locally) maximum value of likelihood attained, when the HMM has more free parameters.

   What happens when you start at $\left[\mu_1^{(0)}, \mu_2^{(0)}\right]$ but an $\alpha_1^{(0)}$ that is different from $\frac{1}{3}$?

8. Let all five parameters be free, and compute the maximum likelihood that is attained for one or two different starting points.

Consider signing up for 600.603/520.701 *Current Topics in Language and Speech Processing* in Fall 2009.