Homework Assignment 4

Due in class, Thursday November 17

SDS 383C Statistical Modeling I

1. **Robust statistics**

   (a) Let Y be a random variable $Y = \mu + \epsilon$ where $\epsilon \sim N(0, 1)$ and μ is a constant. In this question you will compute the elastic net estimator of μ which minimizes:

   $$\frac{1}{2}(Y-\mu)^2 + \lambda|\mu| + \frac{\alpha}{2}\mu^2,$$

   where $\alpha, \lambda > 0$. Calculate $\hat{\mu}$.

   (b) Calculate the Sensitivity curve for the sample median.

   (c) (Extra credit) Assume that the data is generated from iid Uniform($[0, \theta]$). Is the sensitivity curve of the median bounded? Explain your answer.

2. **Expectation Maximization and k-means**

   (a) Derive the E and M steps for a Gaussian Mixture Model with two components with means μ_1, μ_2 and the same variance σ^2. The mixture proportions are $\pi, 1-\pi$.

   (b) Show that if σ has a known value and we take $\sigma \to 0$, the EM algorithm coincides with 2-means clustering.

3. **Expectation Maximization and multinomials**

   Let $\mathbf{y}_{\text{obs}} = (y_1, y_2, y_3)^T = (38, 34, 125)^T$ be observed counts from a multinomial population with probabilities $(1/2 - \theta/2, \theta/4, 1/2 + \theta/4)$.

   (a) Derive the MLE of θ.

   (b) Now we will solve the same problem using EM. In order to put this in the unobserved data framework, we will pretend that the true data is $(y_1, y_2, y_3, y_4)^T$ sampled from a multinomial with probabilities $(1/2 - \theta/2, \theta/4, 1/2, \theta/4)$. $\mathbf{y}$ is the augmented or complete data. Now define by $\mathbf{y}_{obs} = (y_1, y_2, y_3 + y_4)^T$. This is an incomplete data problem because $y_3 + y_4$ is observed, not y_3 or y_4.

      i. Derive the E and M steps.

      ii. Plot the estimated θ_t values vs the number of iterations t. Does it converge to the MLE you calculated earlier?

4. **Linear Discriminant Analysis**

   (a) Consider Fisher's discriminant analysis which finds $w^* := \arg\max_w \frac{(w^T(\mu_1-\mu_2))^2}{w^T(\Sigma_1+\Sigma_2)w}$. Now consider data generated from two Gaussians with parameters μ_i, Σ_i, $\pi = 1/2$, for $i \in \{1, 2\}$. Show that the direction of the Fisher Discriminant Analysis is exactly the direction found by a Linear Bayes classifier. For this exercise you can assume that μ_i, Σ_i, π are known and $\Sigma = (\Sigma_1 + \Sigma_2)/2$.

(b) Using the vowel data available at `http://web.stanford.edu/~hastie/ElemStatLearn/data.html`, reproduce the figures 4.8, and 4.11 in the HTF book. You do not have to generate the linear class boundaries in 4.11, just the scatter plot.

(c) (Extra credit) Also reproduce Figure 4.10.