Statistical Machine Learning, - Problem set 1

Jacob Goldberger

1. **Mixture of Exponential Distributions** Suppose that the time from when a machine is manufactured to when it fails is exponentially distributed. However, suppose that some machines have a manufacturing defect that causes them to be more likely to fail early than machines that don't have the defect. Denote the probability that a machine has the defect by p, the mean time to failure for machine with and without defect by $\frac{1}{\lambda_1}$ and $\frac{1}{\lambda_2}$. The probability density for the time to fail is the following mixture density:

$$f(x) = p\lambda_1 \exp(-\lambda_1 x) + (1-p)\lambda_2 \exp(-\lambda_2 x)$$

Suppose you have n independent observations of times to fail $x_1, ..., x_n$. Apply the EM algorithm to learn the model parameters $\{p, \lambda_1, \lambda_2\}$.

2. Consider the following two-step experiment:

   (a) A uniform r.v. y is sampled, $p(y=1) = p(y=2) = p(y=3) = \frac{1}{3}$.

   (b) A Gaussian random variable x is sampled according to $x \sim N(y\mu, 1)$.

   Given that μ is an unknown parameter, explain how can we utilize the EM algorithm to estimate it from n independent samples $x_1, ..., x_n$.

3. Consider the following two-step discriminative model:

   (a) Given a feature vector x, we sample a binary variable $p(y=1|x) = \sigma(wx+b)$ such that σ is the sigmoid function.

   (b) We next sample a normal r.v. z with mean $y \cdot \mu$ and variance 1, i.e. $z|(y=0) \sim N(0,1)$ and $z|(y=1) \sim N(\mu, 1)$.

   The conditional density of z given x, therefore, is:

$$f(z|x) = \sum_{i=0,1} p(y=i|x) f(z|y=i)$$

   - Compute the conditional probability: $p(y=1|x,z)$.
   - Explain how can we find the maximum-likelihood estimation of the model parameters w, b and μ from n independent samples $(x_1, y_1, z_1), ..., (x_n, y_n, z_n)$.
   - Apply the EM algorithm to estimate the model parameters from n independent samples $(x_1, z_1), ..., (x_n, z_n)$.

4. **Simulation** Let $f(x) = 0.2 \cdot N(-1, 2) + 0.3 \cdot N(4, 3) + 0.5 \cdot N(9, 1)$

   (a) Use the Mixture of Gaussians defined by f to Generate 20 samples.

   (b) Use EM to estimate the parameters of the model starting with the true model. Plot the log-likelihood of the data as a function of iteration. How do your estimated parameters compare to the true ones? What happens when you increase the number of samples to 200 ?

   (c) Repeat parameter estimation with 200 samples starting with 10 random starting points. Do you always obtain, after the EM converges, the same solution (or the same likelihood)? What can you say about EM as a learning algorithm?

Robust computation Let $x_1, ..., x_n$ be n real numbers.

Denote $m = \max x_i$, $s = \sum_{\{i|x_i - m > -10\}} e^{(x_i - m)}$

$$\log\left(\sum e^{x_i}\right) \approx m + \log(s) \qquad\qquad \frac{e^{x_k}}{\sum_i e^{x_i}} \approx \frac{e^{x_k - m}}{s}$$