CS 228 Problem Set 5

Hugh Zhang

March 17, 2017

Problem 1

Note that from the hint, we can see that the Dirichlet distribution is a distributed beta, so all the lecture theorems apply.

$$
\begin{aligned}
&P(X[M+1] = x^i \mid D) \\
=\ & \int_\theta P(X[M+1] = x^i \mid \theta) * P(\theta \mid D) d\theta \\
=\ & \int_\theta \theta_i * P(\theta \mid D) d\theta \\
=\ & E_{\theta\mid D}[\theta_i] = \frac{\alpha_i + M[i]}{M + \alpha}
\end{aligned}
$$

with the last inequality taken straight from lecture.

1.2

$$
\begin{aligned}
&P(X[M+1] = x^i, X[M+2] = x^j \mid D) \\
=\ & P(X[M+1] = x^i \mid D) * P(X[M+2] = x^i \mid X[M+1] = x^j, D)
\end{aligned}
$$

But notice, the second part, is just augmenting the data set by one more example. Thus, our probabilities are just

$$
\frac{\alpha_i + M[i]}{M + \alpha} * \frac{\alpha_j + M[j] + 1(i = j)}{M + \alpha + 1}
$$

1.3

$$P(X[M+1] = x^i, X[M+2] = x^j \mid D)$$
$$/P(X[M+1] = x^i \mid D) * P(X[M+2] = x^i \mid D)$$
$$= \; (\frac{\alpha_i + M[i]}{M + \alpha} * \frac{\alpha_j + M[j] + 1(i = j)}{M + \alpha + 1})/(\frac{\alpha_i + M[i]}{M + \alpha} * \frac{\alpha_j + M[j]}{M + \alpha})$$
$$= \; \frac{(\alpha_j + M[j] + 1(i = j))}{(\alpha_j + M[j])} * \frac{(M + \alpha + 1))}{(M + \alpha)}$$
$$=$$

The difference here is if you update your Dirichlet counts after you sample the M+1th sample when you do exact inference and don't if you do approximate. Thus, if your data set is large, then not updating doesn't matter, but if it is small it might make a large difference.

With the function above, both factors of the product clearly have the +1 become irrelevant as M goes to infinity and dwarfs it.

Problem 2

Let T be the total number of examples.

$$\pi = \sum_{ij} \frac{1[Z_{ij} = 1]}{T} = 0.57$$

$$\mu_0 = \sum_{ij} \frac{1[Z_{ij} = 0] * X_{ij}}{1[Z = 0]} = [-0.99437209, -1.11730233]$$

$$\mu_1 = \sum_{ij} \frac{1[Z_{ij} = 1] * X_{ij}}{1[Z = 1]} = [1.04922807, 0.98085965]$$

$$\sigma_0 = \sum_{ij} \frac{1[Z_{ij} = 0] * (X_{ij} - \mu_0)(X_{ij} - \mu_0)^T}{1[Z = 0]} = [[0.30811884, 0.28553768][0.28553768, 0.81346635]]$$

$$\sigma_1 = \sum_{ij} \frac{1[Z_{ij} = 1] * (X_{ij} - \mu_1)(X_{ij} - \mu_1)^T}{1[Z = 1]} = [[0.77827888, 0.19683566][0.19683566, 0.24996938]]$$
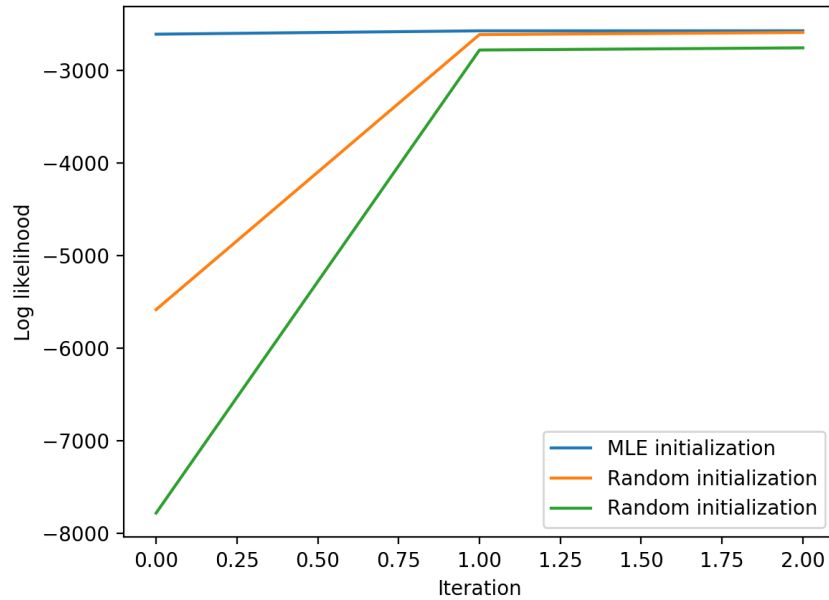
2.A2

Figure 1: Log likelihood functions

MLE start:

```
{'pi': 0.58616150703175707, 'sigma_1': matrix([[ 0.72112418,  0.14499114],
[ 0.14499114,  0.30825171]]), 'sigma_0': matrix([[ 0.36212646,  0.31082931],
[ 0.31082931,  0.75836101]]),
'mu_1': matrix([[ 0.98729587,  0.99618266]]),
'mu_0': matrix([[-1.04406391, -1.02551825]])}
```

Random 1 start:

```
{'pi': 0.51535490062201894, 'sigma_1': matrix([[ 0.5953022 ,  0.54367732],
[ 0.54367732,  1.01903159]]), 'sigma_0': matrix([[ 0.6061935 ,  0.05204946],
[ 0.05204946,  0.24784284]]), \\ 'mu_1': matrix([[-0.80990003, -0.72964472]]),
'mu_0': matrix([[ 1.16379378,  1.10503744]])}
```

Random 2 start:

```
{'pi': 0.18461282486659261, 'sigma_1': matrix([[ 1.66246823,  1.30203827],
[ 1.30203827,  1.3425169 ]]), 'sigma_0': matrix([[ 1.31417998,  1.17278467],
[ 1.17278467,  1.51739247]]), \\ 'mu_1': matrix([[ 1.07480467,  0.22227733]]),
'mu_0': matrix([[-0.06350569,  0.14531717]])}
```

One of the random starts seems to have found approximately the same maxima
that MLE found, but the other one ended up in a slightly worse local minimum.

2.B1

See part 2.A1 for other formulas

$$\phi = \sum_i \frac{1[Y_i = 1]}{N} = 0.6$$

$$\lambda = \sum_{ij} \frac{1[Z_{ij} = Y_i]}{M * N} = .93$$

2.B2

$$P(Y_i = 1 \mid X_{i,1...M}) = P(Y_i = 1) \sum_{z_i,1...M} \prod_{j=1}^{M} p(x_{ij} \mid z_{ij}$$