Cogs 118 B HW3 solutions

1 EM Updates

$$\Theta = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \sigma_1^2 \\ \sigma_2^2 \end{pmatrix}$$

$$\Theta^0 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

$$\mathcal{D} = \left[\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 3 \\ * \end{pmatrix} \right]$$

$$Q(\theta; \theta^i) = E_{D_b}[\ln p(D_g|D_b; \theta)|D_g; \theta]$$

$$\begin{aligned}
&= E_{42}\left[\ln\ p(x_g|x_b;\theta)|\theta^0\right] \\
&= \int_{-\infty}^{+\infty} \left[\sum_{k=1}^{3} \ln\ p(x_k|\theta) + \ln\ p(x_4|\theta)\right] p(x_{42}|\theta^0; x_{41} = 3) d_{x_{42}} \\
&= \sum_{k=1}^{3} \ln\ p(x_k|\theta) + \int_{-\infty}^{\infty} \ln\ p\left(\begin{pmatrix} x_{41} \\ x_{42} \end{pmatrix} |\theta\right) p(x_{42}|\theta^0; x_{41} = 3) d_{x_{42}}
\end{aligned}$$

Collect the constant terms as

$$\beta = \sum_{k=1}^{3} \ln p(x_k|\theta)$$

and we get the last part by bayes rule

$$p(x_{42}|\theta^0; x_{41} = 3)d_{x_{42}} = \frac{p\begin{pmatrix} 3 \\ x_{42} \end{pmatrix} |\theta^0)}{\int \begin{pmatrix} 3 \\ x'_{42} \end{pmatrix} |\theta^0) \, dx'_{42}}$$

the x' is to denote a different x on the top and bottom of the fraction. We set K as the constant value of the integral

$$K = \int \begin{pmatrix} 3 \\ x'_{42} \end{pmatrix} |\theta^0) \, dx'_{42}$$

and K is not 1 because we haven't integrated over the entire probability distribution (a few peopel had this confusion in office hours). We've only marginalized out x_{42} : if you think of the distribution as a two-dimensional shape, the integral here is the integral along the line $x_{41} = 3$, and since the integral of the entire 2d surface is 1, then an integral along a line must necessarily be less than 1. Plugging back into the problem we get:

$$= \beta + \frac{1}{K} \int \ln\ p(\frac{3}{x_{42}}|\theta) p\left(\begin{array}{c} 3 \\ x_{42} \end{array} |\theta^0 \right) dx_{42}$$

$$= \beta + \frac{1}{K} \int \ln\ p(\frac{3}{x_{42}}|\theta) \left(\frac{1}{2\pi \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}} \exp^{\frac{\left(-3^2 - x_{42}^2\right)}{2}} \right) dx_{42}$$

$$= \beta + \frac{1}{K} \int \left[\ln\ \left(\frac{1}{2\pi\sigma_1\sigma_2} \right) - \frac{(3-\mu_1)^2}{2\sigma_1^2} - \frac{\left(x_{42} - \mu_2^2\right)^2}{2\sigma_2^2} \right] \left(\frac{1}{2\pi \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}} \exp^{\frac{\left(-3^2 - x_{42}^2\right)}{2}} \right) dx_{42}$$

$$= \beta + \frac{1}{K} \frac{1}{\sqrt{2\pi}} \exp^{\frac{-3^2}{2}} \int \left[\ln \left(\frac{1}{2\pi\sigma_1\sigma_2} \right) - \frac{(3-\mu_1)^2}{2\sigma_1^2} - \frac{(x_{42}-\mu_2)^2}{2\sigma_2^2} \right] \frac{1}{\sqrt{2\pi}} \exp^{\frac{-x_{42}^2}{2}} dx_{42}$$

$$= \beta + \int \left[\ln \left(\frac{1}{2\pi\sigma_1\sigma_2} \right) \frac{1}{\sqrt{2\pi}} \exp^{\frac{-x_{42}^2}{2}} - \frac{(3-\mu_1)^2}{2\sigma_1^2} \frac{1}{\sqrt{2\pi}} \exp^{\frac{-x_{42}^2}{2}} - \frac{(x_{42}-\mu_2)^2}{2\sigma_2^2} \frac{1}{\sqrt{2\pi}} \exp^{\frac{-x_{42}^2}{2}} \right] dx_{42}$$

$$= \beta + \left[\ln \left(\frac{1}{2\pi\sigma_1\sigma_2} \right) - \frac{(3-\mu_1)^2}{2\sigma_1^2} - \int \frac{(x_{42}-\mu_2)^2}{2\sigma_2^2} \frac{1}{\sqrt{2\pi}} \exp^{\frac{-x_{42}^2}{2}} dx_{42} \right]$$

$$= \beta + \left[\ln \left(\frac{1}{2\pi\sigma_1\sigma_2} \right) - \frac{(3-\mu_1)^2}{2\sigma_1^2} - \frac{\mu_2^2}{2\sigma_2^2} - \frac{1}{2\sigma_2^2} \right]$$

$$= 3 \ln \left(\frac{1}{2\pi\sigma_1\sigma_2} \right) - \frac{\mu_1^2}{2\sigma_1^2} - \frac{(1-\mu_2)^2}{2\sigma_2^2} - \frac{(2-\mu_1)^2}{2\sigma_1^2} - \frac{(1-\mu_1)^2}{2\sigma_1^2} - \frac{(1-\mu_2)^2}{2\sigma_2^2} - \frac{\mu_2^2}{2\sigma_2^2} + \left[\ln \left(\frac{1}{2\pi\sigma_1\sigma_2} \right) - \frac{(3-}{2}\right.$$

$$= 4 \ln \left(\frac{1}{2\pi\sigma_1\sigma_2} \right) - \frac{\mu_1^2}{2\sigma_1^2} - \frac{(1-\mu_2)^2}{2\sigma_2^2} - \frac{(2-\mu_1)^2}{2\sigma_1^2} - \frac{(1-\mu_1)^2}{2\sigma_1^2} - \frac{(3-\mu_1)^2}{2\sigma_1^2} - \frac{\mu_2^2}{2\sigma_2^2} - \frac{1}{2\sigma_2^2}$$

This gives us the new value of Q.

$$\frac{\partial Q}{\partial \mu_2} = 2 - 2\mu_2 - 2\mu_2$$
$$0 = 2 - 2\mu_2 - 2\mu_2$$
$$\mu_2 = \frac{1}{2}$$

2 2

$$\sum_{k=1}^{K} \pi_k = 1$$

and

$$p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z^k}$$

and

$$p(\mathbf{x}|\mathbf{z}) = \prod_{k=1}^{K} \mathcal{N}(x|\mu_k, \mathbf{\Sigma}_k)^{z^k}$$

First, we marginalize

$$\sum_{\mathbf{z}} p(\mathbf{x}|\mathbf{z})p(\mathbf{z})$$

then we can plug in the formulas from Bishop

$$\sum_{z} \prod_{k=1}^{K} \pi_k^{z^k} \mathcal{N}(x|\mu_k, \mathbf{\Sigma}_k)^{z^k}$$
$$= \sum_{z} \prod_{k=1}^{K} (\pi_k \mathcal{N}(x|\mu_k, \mathbf{\Sigma}_k))^{z^k}$$

and comparing to our target (Bishop 9.7) we have to eliminate the sum and product, and be left with a sum over k. We can think of this as a loop. For a fixed value of zk, every element in the product will be pi_k times Normal . This is a product over all our k's. For the outer loop, z will iterate, but for z's that aren't k, the entire product term will turn into all ones.

$$= \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)$$

3 Problem Three

(Mixture distribution mean and covariance) Consider a mixture distribution of the form

$$p(x) = \sum_{k=1}^{K} \pi_k p(x|k)$$

and

$$p(x|k) = \mathcal{N}(\mu_k, \mathbf{\Sigma}_k)$$

and also in addition

Show that the mean and covariance of the mixture distribution are given by Bishop equations (9.49) and (9.50) respectively.

$$\mathbb{E}[x] = \sum_{k=1}^{K} \pi_k \mu_k$$

$$\text{cov}[x] = \sum_{k=1}^{K} \pi_k \{\Sigma_k + \mu_k \mu_k^T\} - \mathbb{E}[x]\mathbb{E}[x]^T$$

First, we can write out the full distribution for x by filling out the marginalization

$$p(x) = \sum_{k=1}^{K} \pi_k p(x|k)$$

For the expectation,we just take

$$\begin{aligned}
\mathbb{E}(x) &= \sum_x x(p(x)) \\
&= \sum_x x \left(\sum^{K} \pi_k \mathrm{p(x|k)} \right) \\
&= \sum_x \sum_{k=1}^{K} \pi_k x \ \mathrm{p(x|k)} \\
&= \sum_{k=1}^{K} \sum_x \pi_k x \mathrm{p(x|k)} \\
&= \sum_{k=1}^{K} \pi_k \ \sum_x x \mathrm{p(x|k)} \\
&= \sum_{k=1}^{K} \pi_k \mu_k
\end{aligned}$$

For the covariance it is in the separate part b file

$$
\begin{aligned}
cov(x) &= \mathbb{E}_{x,y}\left[\{x - E\left[x\right]\}\left\{x^T - E\left[x^T\right]\right\}\right] \\
&=
\end{aligned}
$$

4 Problem Four

(Data marginal distribution for Bernoulli mixture model) Consider the joint distribution of latent and observed variables for the Bernoulli distribution obtained by forming the product of p(xjz;) (given by Bishop equation (9.52))

$$p(x|z,\mu) = \prod_{k=1}^{K} p(x|\mu_k)^{z^k}$$

and p(zj) (given by Bishop equation (9.53)).

$$p(z|\pi) = \prod_{k=1}^{K} \pi_k^{z^k}$$

Show that if we marginalize this joint distribution with respect to z,

$$
\begin{aligned}
p(x|\mu,\pi) &= \sum_z p(x|z,\mu)(p(z|\pi) \\
&= \sum_z \prod_{k=1}^{K} \pi_k^{z^k} p(x|\mu_k)^{z^k} \\
&= \sum_z \prod_{k=1}^{K} (\pi_k p(x|\mu_k))^{z^k}
\end{aligned}
$$

then we obtain Bishop equation (9.47).

$$p(x|\mu,\pi) = \sum_{k=1}^{K} \pi_k p(x|\mu_k)$$

5 Problem Five

(Bernoulli mixture model M step) Show that if we maximize the expected complete-data log likelihood function (Bishop equation (9.55))

$$E_z \left[\ln p(X, Z|\mu, \pi)\right] = \sum \sum \gamma(z_{nk}) \left\{ \ln \pi_k + \sum \left[x_{ni} \ln \mu_{ki} + (1 - x_{ni}) \ln(1 - \mu_{ki})\right] \right\}$$

for a mixture of Bernoulli distributions with respect to k, we obtain the M step equation shown in Bishop equation (9.59).

This question is done in the separate file

6 Eigenvectors

$$\begin{aligned}
AA^T v &= \lambda v \\
u &= A^T v \\
A^T A u &= A^T A (A^T v) \\
&= A^T (\lambda v) \\
&= \lambda (A^T v)
\end{aligned}$$

Note that $A^T v$ is a vector, so we have an eigenvector by definition.

Questions 7-9 are also in a separate file

A note about eigenvectors:

When you try to calculate eigenvectors of the identity matrix (try it!) you run into problems: this is a well known flaw of the 'power method' for calculating eigenvectors: it fails when you have repeated or closely spaced eigenvalues. For more details on more stable algorithms, consult a good numerical linear algebra text.