Graphical Model (1)

$$x_0 \rightarrow x_1 \rightarrow x_2 \rightarrow x_3$$

part a:

1- you draw sample from x_0

2- draw conditional sample from x_1 given sample from x_0

⋮

part b

Here, we use message passing idea:

$$p(x_0, x_1, x_2, x_3) = p(x_0)\,p(x_1|x_0)\,p(x_2|x_1)\,p(x_3|x_2)$$

Marginal distribution for $p(x_0)$ is defined by:

$$p(x_0) = \sum_{x_1}\sum_{x_2}\sum_{x_3} p(x_0)\,p(x_1|x_0)\,p(x_2|x_1)\,p(x_3|x_2)$$

$$= \sum_{x_1}\sum_{x_2} p(x_0)\,p(x_1|x_0)\,p(x_2|x_1)\sum_{x_3} p(x_3|x_2)$$

$$= \sum_{x_1} p(x_0)\,p(x_1|x_0)\sum_{x_2} p(x_2|x_1)\sum_{x_3} p(x_3|x_2)$$

Slides 42–46 from the second lecture on graphical models.

$$= \underbrace{p(x_0)}_{\mu_\alpha(x_0)} \underbrace{\sum_{x_1} p(x_1|x_0)\sum_{x_2} p(x_2|x_1)\sum_{x_3} p(x_3|x_2)}_{\mu_\beta(x_0)}$$

We combine this two messages to calculate $p(x_0)$

We repeat the same procedure for x_1, x_2, x_3. Note that you calculate $\mu_\alpha(\cdot)$ and $\mu_\beta(\cdot)$ for different X.

part c

you can do two things: 1) use $x_2 = \text{'hot'}$ in the above model where

$$p(x_2 = \text{'cold'}|x_1) = 0$$
$$p(x_2 = \text{'mild'}|x_1 = 0$$
$$p(x_2 = \text{'hot'}|x_1) = 1$$

2) or, use the triplet idea give x_2 is known; then (x_0, x_1) and (x_3) conditioned on x_2 are independent. Now: you have

two graphs.

$$x_0 \to x_1 \to x_2$$

$$x_2 \to x_3$$

part d We have

$$p(x_0, x_1, x_2, x_3) = p(x_0)\,p(x_1|x_0)\,p(x_2|x_1)\,p(x_3|x_2)$$

and we want to maximize

$$\hat{x}_0, \hat{x}_1, \hat{x}_2, \hat{x}_3 = \underset{x_0, x_1, x_2, x_3}{\arg\max}\; p(x_0, x_1, x_2, x_3)$$

$$= \underset{x_0, x_1, x_2, x_3}{\arg\max}\; \log p(x_0) + \log p(x_1|x_0) + \log p(x_2|x_1) + \log p(x_3|x_2)$$

We can do: $\max_{x_0, x_1, x_2} \log p(x_0) + \log p(x_1|x_0) + \log p(x_2|x_1) + \max_{x_3} \log p(x_3|x_2)$

$$= \max_{x_0} \log p(x_0) + \max_{x_1} \log p(x_1|x_0) + \max_{x_2} \log p(x_2|x_1) + \max_{x_3} \log p(x_3|x_2)$$

$\therefore$ find x_3 that for every possible value of x_2 has the largest probability

$\therefore$ Now, find x_2 the same way you find x_3, but you maximimize $\log p(x_2|x_1) + \log p(\hat{x}_3|x_2)$

$\therefore$ Now, you do the same for x_1

$\therefore$ Finally, you find x_0 with the same rule.

Graphical Model (2) This is very similar to previous one; except we do not see xs directly; instead, we observe y which are function of x. The chain process in the top is

a Markov process and now we have hidden Markov process; which is briefly called HMM.

part a: It is a geratire process; you draw sample from x_s first. you then draw samples from y given x_s.
Note that in practice, you observe y and you try to estimate x.

part b: Now, you have x and you only draw samples of y.

part c: Our objective is to find $p(x_0, x_1, x_2, x_3 | y_0, y_1, y_2, y_3)$

The full likelihood is:

$$p(x_0, x_1, x_2, x_3, y_0, y_1, y_2, y_3)$$

$$= p(x_0)p(y_0|x_0)p(x_1|x_0)p(y_1|x_1)p(x_2|x_1)p(y_2|x_2)p(x_3|x_2)p(y_3|x_3)$$

What we want to find is:

$$p(x_0, x_1, x_2, x_3 | y_0, y_1, y_2, y_3) = \frac{p(x_0, x_1, x_2, x_3, y_0, y_1, y_2, y_3)}{p(y_0, y_1, y_2, y_3)}$$

here, the denominator is a fixed number; thus, we can calculate $p(x_0, x_1, x_2, x_3, y_0, y_1, y_2, y_3)$ and normalize it. Note that y_0, y_1, y_2, y_3 are observed and we replace them with their observed values.

Now, we do the marginilazation one more time.

For $p(x_0)$, we have:

$$p(x_0) = \sum_{x_1}\sum_{x_2}\sum_{x_3} p(x_0)p(y_0|x_0)p(x_1|x_0)p(y_1|x_1)\cdots$$
$$p(x_2|x_1)p(y_2|x_2)p(x_3|x_2)p(y_3|x_3)$$

$$= p(x_0)p(y_0|x_0)\sum_{x_1} p(y_1|x_1)p(x_1|x_0)\sum_{x_2} p(y_2|x_2)p(x_2|x_1)\cdots$$

$$\sum_{x_3} p(y_3|x_3)p(x_3|x_2)$$

Note that; this is very similar to the above problem; except, at each node, we have a weight (scale) defined by $p(y_k|x_k)$.

$\therefore$ forward message is $p(x_0)p(y_0|x_0) = \mu_\alpha(x_0)$

$\therefore$ Backward message starts from $\sum_{x_3} p(y_3|x_3)p(x_3|x_2) = \mu_\beta(x_2)$

then $\mu_\beta(x_1) = \sum_{x_2} p(y_2|x_2)\, p(x_2|x_1)\mu_\beta(x_2)$

then $\mu_\beta(x_0) = \sum_{x_1} p(y_1|x_1)p(x|x_0)\mu_\beta(x_1)$

$$\therefore p(x_0) = \mu_\beta(x_0)\,\mu_\alpha(x_0)\,\frac{1}{Z}$$

$$Z = \sum_{x_0} \mu_\beta(x_0)\mu_\alpha(x_0)$$

Gaussian Mixture Model

This problem is straight forward; note that if you start from different starting point, you might get different estimate for the mixture model parameters.

There are functions in python and Matlab, which fit the GMM. I recommend to write your code and compare your results with those functions' results. You have every term in the slides. we

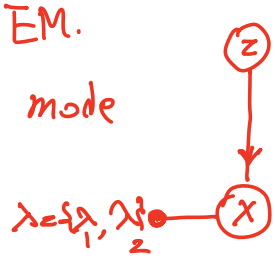
Covered in the class.

Poisson Mixture model.

part a Here, you plot the histogram of the data. Note that, the observed data points are discrete numbers. You can set your histogram bin to match discrete numbers.

part b Here, we find maximum likelihood estimate of λ.
Now using λ_{ML}, you can plot pmf of your fitted model and compare how close the fit is to your histogram (empirical) in part (a).

part c You repeat steps we discussed for EM.
You have the following graph mode



You take steps of GMM here: (we have $K=2$)

E-Step: $P(z_{ni} \mid x_n, \lambda_1^{old}, \lambda_2^{old}, \pi_1^{old}, \pi_2^{old})$ for every z_n $n = 1 \dots N$ we have N data points

$$Q(\theta, \theta^{old}) = E_{P(Z \mid X, \theta^{old})} \left[\text{Log of } \underbrace{\text{full likelihood over } X \text{ and } Z}_{\text{Note, here instead of Normal distribution; you have Poisson}} \right]$$

M-Step:

$$\theta^{new} = \underset{\theta}{\operatorname{argmax}} \; Q(\theta, \theta^{old})$$

$$\theta^{new} = \{ \pi_1^{new}, \pi_2^{new}, \lambda_1^{new}, \lambda_2^{new} \}$$

note, $\pi_2^{new} = 1 - \pi_1^{new}$

part d In part C, you derive EM update rule.

You run the EM for many iterations till some stopping criteria is met. For example, $\max\left(\frac{|\lambda_1^{new} - \lambda_1^{old}|}{\lambda_1^{old}}, \frac{|\lambda_2^{new} - \lambda_2^{old}|}{\lambda_2^{old}}\right) < \varepsilon$

$\varepsilon \simeq 0.01$

or

$\varepsilon \simeq 0.001$

At the end, you plug $\pi_1, \pi_2, \lambda_1, \lambda_2$ to you mixture model function, and that will be your pmf.

We expect this pmf will be close to empirical pmf/histogram.

part e

We know model evidence is defined by

$$P(M|D) = \int P(D|\theta)\, P(\theta|M)\, d\theta$$

now, θ is our model parameters.

$\Leftarrow$ you can ignore this part

We discussed BIC; number of freedom in model a is ? and number of freedom in model b is ?

put proper number

put proper number

We have ML of model parameters in part a, and b.