EM test

Sunday, June 24, 2018 1:03 PM

1. Recall the derivation of EM algorithm. In our notation, X is observable data, Z is latent variable and θ is a vector of model parameters. We introduced $q(Z)$ — an arbitrary distribution over the latent variable. Choose the correct expressions for the marginal log-likelihood $\log p(X \mid \theta)$:

- [ ] $\int q(Z) \log \frac{p(X,Z|\theta)}{q(Z)} dZ + \int q(Z) \log \frac{q(Z)}{p(Z|X,\theta)} dZ$?

  This should be selected

- [x] $\log \int p(X,Z|\theta) dZ$

  Correct
  Z is integrated out:

  $\log \int p(X,Z|\theta) dZ = \log p(X|\theta)$

- [ ] $\mathbb{E}_{q(Z)} \log p(X,Z|\theta) - \mathbb{E}_{q(Z)} \log p(Z|X,\theta)$?

  This should be selected

- [ ] $\int q(Z) \log p(X|\theta) dZ$?

  This should be selected

2. In EM algorithm, we maximize variational lower bound $\mathcal{L}(q,\theta) = \log p(X|\theta) - \mathrm{KL}(q||p)$ with respect to q (E-step) and θ (M-step) iteratively. Why is the maximization of lower bound on E-step equivalent to minimization of KL divergence?

- ( ) Because uncomplete likelihood does not depend on $q(Z)$?
- ( ) Because we cannot maximize lower bound w.r.t. $q(Z)$
- ( ) Because posterior becomes tractable
- (•) Because of Jensen's inequality

  This should be selected
  Revise E-step details video

✔ 6. Imagine that you want to pat your friend's cat Becky. Cats are really random creatures.

1 / 1 points

Becky might get grumpy and scratch you with probability p or curl up and start purring (with prob. $1-p$). You don't know Becky well yet, so you estimate prior on p to be distributed as $Beta(2,2)$. Within one evening, Becky has scratched you 6 times and only 2 times she purred. What will be the parameters for posterior distribution over p? What is the MAP-estimate for p?

Enter your answers separated by comma: e.g. if you think that correct answer is $Beta(1,0.2)$ and MAP is 3, you should enter 1,0.2,3. Express real numbers as decimals with dot as delimiter.

8,4,0.7

Correct Response

1. Recall the derivation of EM algorithm. In our notation, X is observable data, Z is latent variable and θ is a vector of model parameters. We introduced $q(Z)$ — an arbitrary distribution over the latent variable. Choose the correct expressions for the marginal log-likelihood $\log p(X \mid \theta)$:

- [x] $\int q(Z) \log \frac{p(X,Z|\theta)}{q(Z)} dZ + \int q(Z) \log \frac{q(Z)}{p(Z|X,\theta)} dZ$

  Correct
  $\int q(Z) \log \frac{p(X,Z|\theta)}{q(Z)} dZ + \int q(Z) \log \frac{q(Z)}{p(Z|X,\theta)} dZ =$

  $\int q(Z) \log p(X,Z|\theta) dZ - \cancel{\int q(Z) \log q(Z) dZ} +$

  $+ \cancel{\int q(Z) \log q(Z) dZ} - \int q(Z) \log p(Z|X,\theta) dZ =$

  $= \int q(Z) \log \frac{p(X,Z|\theta)}{p(Z|X,\theta)} dZ = \int q(Z) \log p(X|\theta) dZ = \log p(X|\theta)$

  " $p(x,z|\theta) / p(x|\theta)$

- [x] $\log \int p(X,Z|\theta) dZ$

  Correct
  Z is integrated out:

  $\log \int p(X,Z|\theta) dZ = \log p(X|\theta)$

- [x] $\mathbb{E}_{q(Z)} \log p(X,Z|\theta) - \mathbb{E}_{q(Z)} \log p(Z|X,\theta)$

  Correct
  $\mathbb{E}_{q(Z)} \log p(X,Z|\theta) - \mathbb{E}_{q(Z)} \log p(Z|X,\theta) =$

  $= \mathbb{E}_{q(Z)} \log \frac{p(X,Z|\theta)}{p(Z|X,\theta)} = \mathbb{E}_{q(Z)} \log p(X|\theta) = \log p(X|\theta)$

  $= p(Z,\cancel{X}|\theta) / p(X|\theta)$

- [x] $\int q(Z) \log p(X|\theta) dZ$

  Correct
  $\log p(X|\theta)$ does not depend on Z.

  $\int q(Z) \log p(X|\theta) dZ = \log p(X|\theta)$

  $\int q(z) dz \cdot \log(x|\theta)$

$$\Rightarrow \mathcal{L}(y|\theta) = \prod_{i=1}^{8} \theta^{y_i} (1-\theta)^{1-y_i} = \theta^{\sum y_i} (1-\theta)^{8-\sum y_i} = \theta^6 (1-\theta)^2$$

$$p(y|\theta) = \theta^y (1-\theta)^{1-y}$$

$$p(\theta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

$$p(\theta|y) = \frac{p(y|\theta)\, p(\theta)}{\int p(y|\theta)\, p(\theta)\, d\theta}$$

$$\propto p(y|\theta)\, p(\theta)$$

$$\propto \theta^{\alpha+6-1} (1-\theta)^{\beta+2-1}$$

$$= \theta^{2+6-1} (1-\theta)^{2+2-1}$$

$= \text{Beta}(8, 4)$

Calculate

Mode, not mean

Mode of Beta dist

$$\frac{\alpha - 1}{\alpha + \beta - 2} = \frac{8-1}{8-4-2} = \frac{7}{10}$$