1 / 1
points

**1.**

Recall the derivation of EM algorithm. In our notation, $X$ is observable data, $Z$ is latent variable and $\theta$ 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 $\log p(X \mid \theta)$:

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

**Correct**

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

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

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

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

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

**Correct**

Z is integrated out:

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

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

**Correct**

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

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

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

**Correct**

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

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

---

1 / 1
points

**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 $\theta$ (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 $q(Z)$

**Correct**

Revise E-step details video

- ○ Because we cannot maximize lower bound w.r.t. $q(Z)$
- ○ Because posterior becomes tractable
- ○ Because of Jensen's inequality

---

1 / 1
points

### 3.

Select correct statements about EM algorithm:

- ☐ E-step can always be performed analytically

**Un-selected is correct**

- ☐ M-step can always be performed analytically

**Un-selected is correct**

- ☐ EM algorithm always converges

**Correct**
Revise M-step details video

- ☐ Complete likelihood is always a convex function as a function of parameters

**Un-selected is correct**

- ☐ EM algorithm always converges to a global optimum

**Un-selected is correct**

---

1 / 1
points

### 4.

Consider $p(x) = \mathcal{N}(\mu, \sigma_1)$ and $q(x) = \mathcal{N}(\mu, \sigma_2)$. Calculate KL divergence between these two gaussians KL$(p||q)$ (hint: note that KL divergence is an expectation):

- ○ $\log \frac{\sigma_2}{\sigma_1} - \frac{1}{2} + \frac{\sigma_1^2}{2\sigma_2^2}$

**Correct**

$$KL(p||q) = \mathbb{E}_p \log \frac{\mathcal{N}(x|\mu,\sigma_1^2)}{\mathcal{N}(x|\mu,\sigma_2^2)} = \mathbb{E}_p \log \frac{\left(\sqrt{2\pi\sigma_1^2}\right)^{-1} \exp\left(-\frac{(x-\mu)^2}{2\sigma_1^2}\right)}{\left(\sqrt{2\pi\sigma_2^2}\right)^{-1} \exp\left(-\frac{(x-\mu)^2}{2\sigma_2^2}\right)} =$$

$$= \mathbb{E}_p \left[ \log \frac{\sigma_2}{\sigma_1} + \log \frac{\exp\left(-\frac{(x-\mu)^2}{2\sigma_1^2}\right)}{\exp\left(-\frac{(x-\mu)^2}{2\sigma_2^2}\right)} \right] = \mathbb{E}_p \left[ \log \frac{\sigma_2}{\sigma_1} - \frac{(x-\mu)^2}{2\sigma_1^2} + \frac{(x-\mu)^2}{2\sigma_2^2} \right] =$$

$= \log \frac{\sigma_2}{\sigma_1} - \frac{\mathbb{E}_p(x-\mu)^2}{2\sigma_1^2} + \frac{\mathbb{E}_p(x-\mu)^2}{2\sigma_2^2} = \log \frac{\sigma_2}{\sigma_1} - \frac{\sigma_1^2}{2\sigma_1^2} + \frac{\sigma_1^2}{2\sigma_2^2} = \log \frac{\sigma_2}{\sigma_1} - \frac{1}{2} + \frac{\sigma_1^2}{2\sigma_2^2}$

- ○ $\log \frac{\sigma_1}{\sigma_2} + \frac{\sigma_1^2-\sigma_2^2}{\sigma_1^2+\sigma_2^2}$
- ○ $\log \frac{\sigma_1}{\sigma_2} + \frac{\sigma_2^2}{\sigma_1^2}$
- ○ $\log \frac{\sigma_2^2}{\sigma_1^2} - \frac{\sigma_1^2}{2\sigma_2^2}$