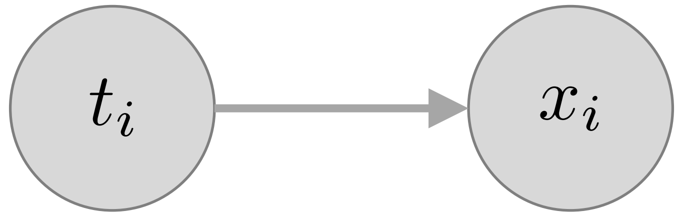
General form of Expectation Maximization

This is for a specific sample i.

t_i here is a latent variable.

x_i observed.



We marginalize by the t latent variable here:

LIKELIHOOD

$$p(x_i \mid \theta) = \sum_{c=1}^{3} p(x_i \mid t_i = c, \theta)\, p(t_i = c \mid \theta)$$

General form of Expectation Maximization

Take logarithm of our likelihood. This will turn products into summations and make calculations easier.

Assume there are N samples. They are all independent, so we can take the products.

$$\max_{\theta}.\, \log p(X \mid \theta) = \log \prod_{i=1}^{N} p(x_i \mid \theta)$$

$$= \sum_{i=1}^{N} \log p(x_i \mid \theta)$$

TURN PRODUCTS into summation, as by property of logarithm. This makes it easier!

General form of Expectation Maximization

$$\log p(X \mid \theta) = \sum_{i=1}^{N} \log p(x_i \mid \theta)$$

$$= \sum_{i=1}^{N} \log \sum_{c=1}^{3} p(x_i, t_i = c \mid \theta) \geq \mathcal{L}(\theta)$$

Jensen's inequality

We could do Stochastic gradient descent. But, there are reasons not to do this.

Idea:
build a lower bound L(theta). We use jensen's inequality.
This makes it easier (see following slides for why).

In case of maximising this marginal likelihood, we can maximise the lower bound by itself.

General form of Expectation Maximization

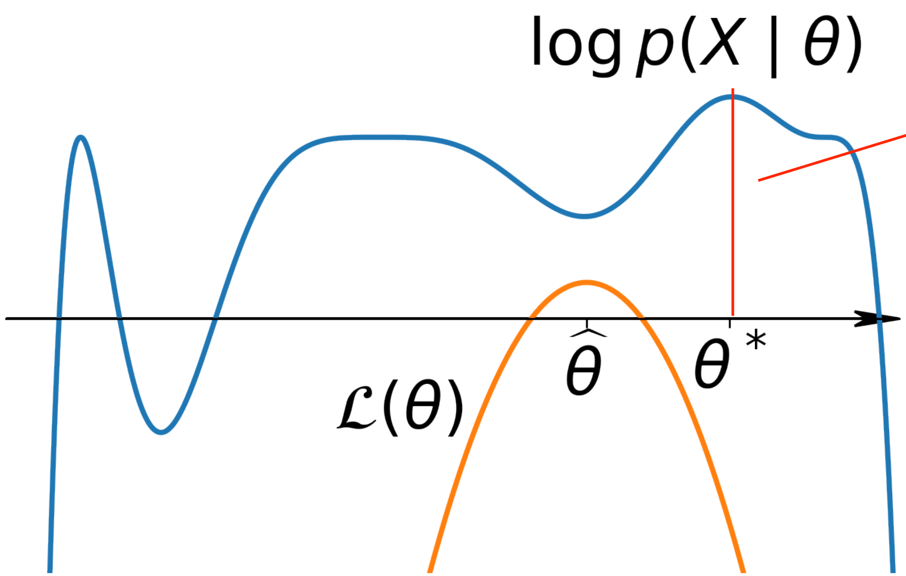
$$\log p(X \mid \theta) = \sum_{i=1}^{N} \log p(x_i \mid \theta)$$

$$= \sum_{i=1}^{N} \log \sum_{c=1}^{3} p(x_i, t_i = c \mid \theta) \geq \mathcal{L}(\theta)$$

'optimal' target here.

assume log likelihood is parameterized by theta.

The orange graph is our lower bound.
It needs to move to the optimal!



General form of Expectation Maximization

$$\log p(X \mid \theta) = \sum_{i=1}^{N} \log p(x_i \mid \theta)$$

$$= \sum_{i=1}^{N} \log \sum_{c=1}^{3} p(x_i, t_i = c \mid \theta)$$

General form of Expectation Maximization

$$\log p(X \mid \theta) = \sum_{i=1}^{N} \log p(x_i \mid \theta)$$

$$= \sum_{i=1}^{N} \log \sum_{c=1}^{3} \boxed{\frac{q(t_i = c)}{q(t_i = c)}} p(x_i, t_i = c \mid \theta)$$

The idea is to introduce some weight that we can use to change our lower bound. Introduce a variable q for this.
This doesn't change the formula of course, as it becomes 1.

But see the next slide for the magic...

General form of Expectation Maximization

USE JENSEN's INEQUALITY!

$$\log p(X \mid \theta) = \sum_{i=1}^{N} \log p(x_i \mid \theta)$$

$$= \sum_{i=1}^{N} \log \sum_{c=1}^{3} \frac{q(t_i = c)}{q(t_i = c)} p(x_i, t_i = c \mid \theta)$$

$$\geq \sum_{i=1}^{N} \sum_{c=1}^{3} q(t_i = c) \log \frac{p(x_i, t_i = c \mid \theta)}{q(t_i = c)}$$

Jensen's inequality

$$\boxed{\log \left(\sum_c \alpha_c v_c \right) \geq \sum_c \alpha_c \log(v_c)}$$

General form of Expectation Maximization

$$\log p(X \mid \theta) = \sum_{i=1}^{N} \log p(x_i \mid \theta)$$

$$= \sum_{i=1}^{N} \log \sum_{c=1}^{3} \frac{q(t_i = c)}{q(t_i = c)} p(x_i, t_i = c \mid \theta)$$

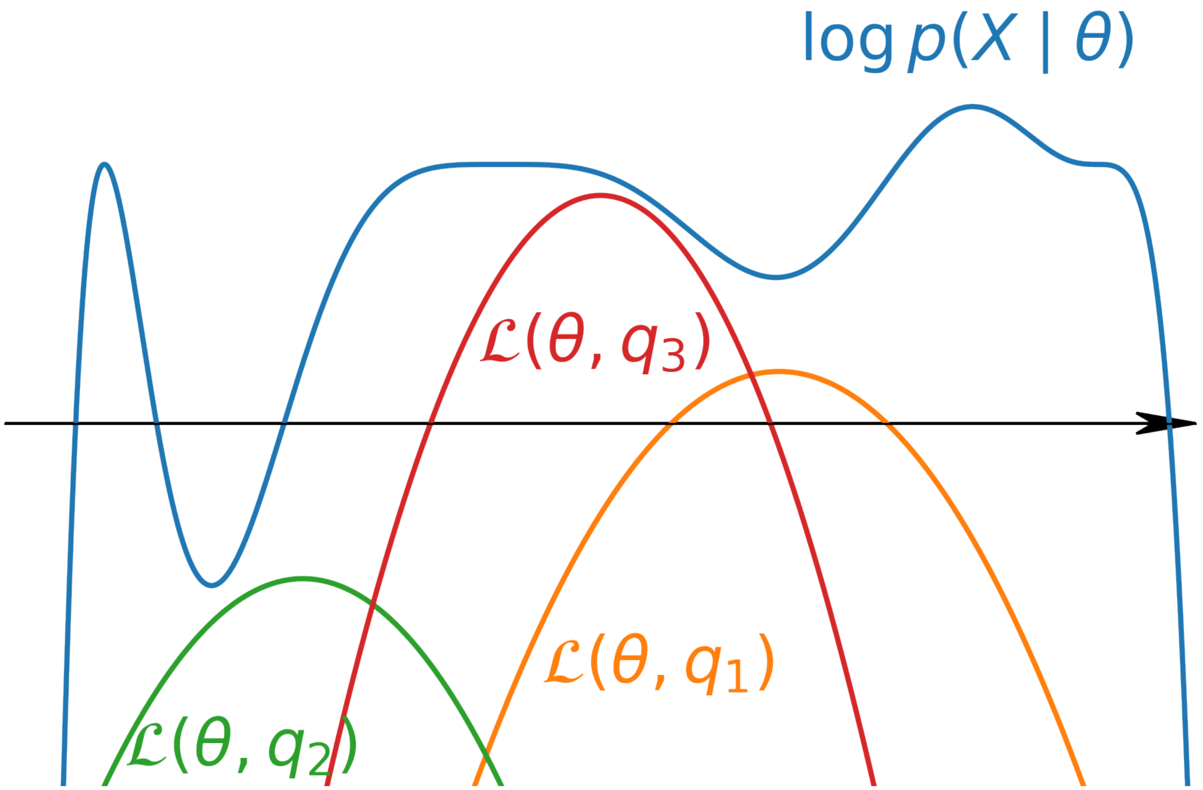
$$\geq \boxed{\sum_{i=1}^{N} \sum_{c=1}^{3} q(t_i = c) \log \frac{p(x_i, t_i = c \mid \theta)}{q(t_i = c)}}$$

$$= \mathcal{L}(\theta, q)$$

And turns out, this function is our lower bound!

General form of Expectation Maximization

$$\log p(X \mid \theta) \geq \mathcal{L}(\theta, q) \text{ for any q}$$

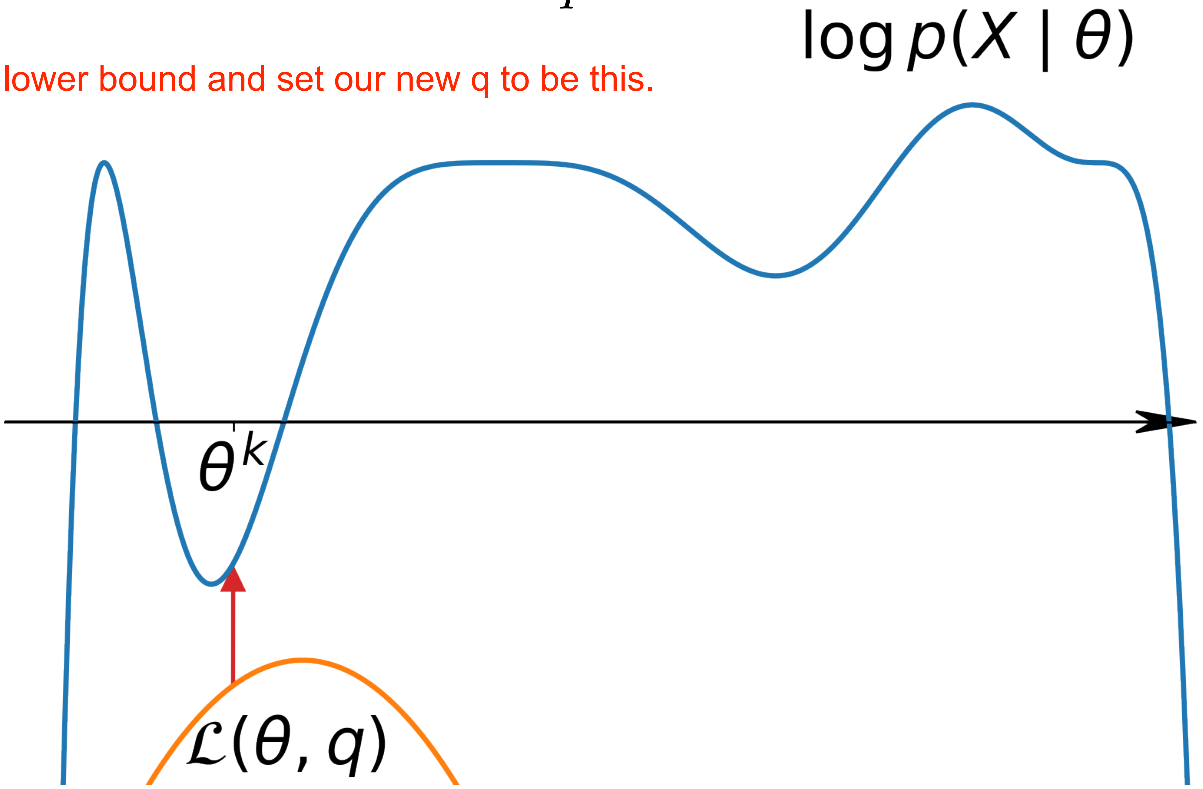


General form of Expectation Maximization

$$\log p(X \mid \theta) \geq \mathcal{L}(\theta, q) \text{ for any q}$$

$$q^{k+1} = \arg\max_q \mathcal{L}(\theta^k, q)$$

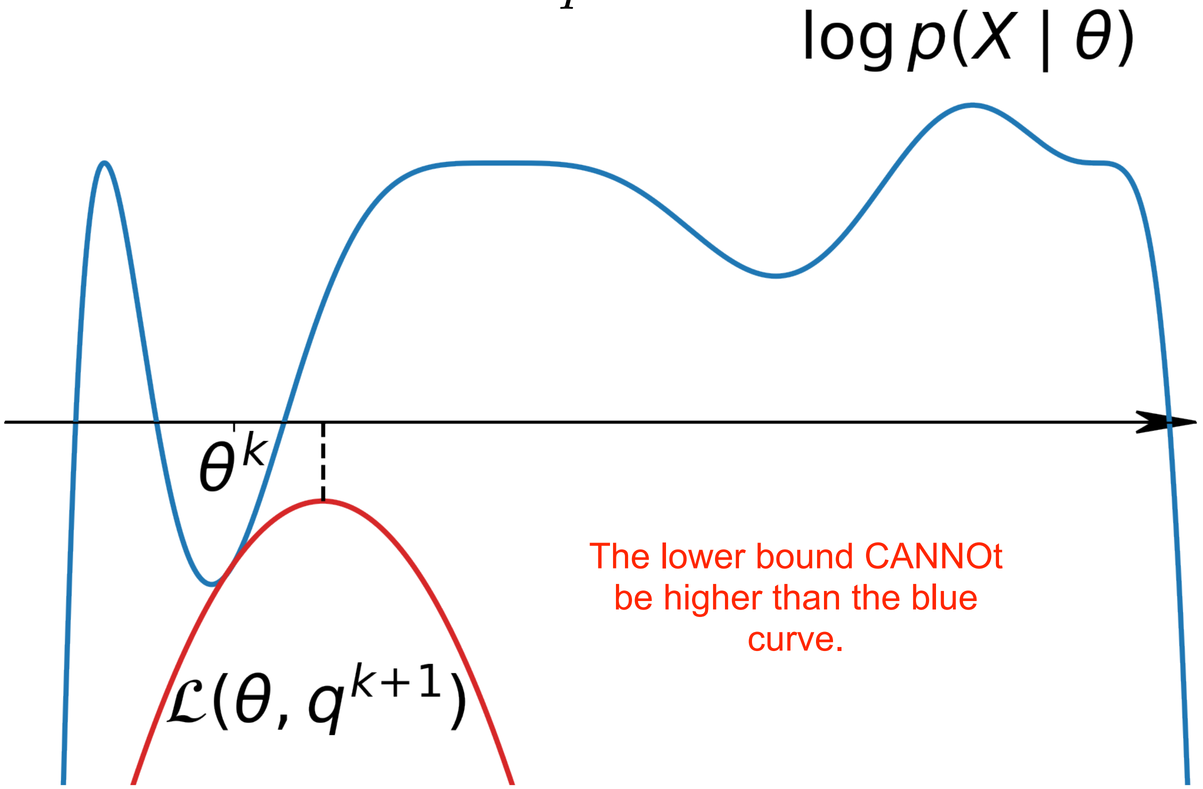
Firrst, maximise lower bound and set our new q to be this.



General form of Expectation Maximization

$$\log p(X \mid \theta) \geq \mathcal{L}(\theta, q) \text{ for any q}$$

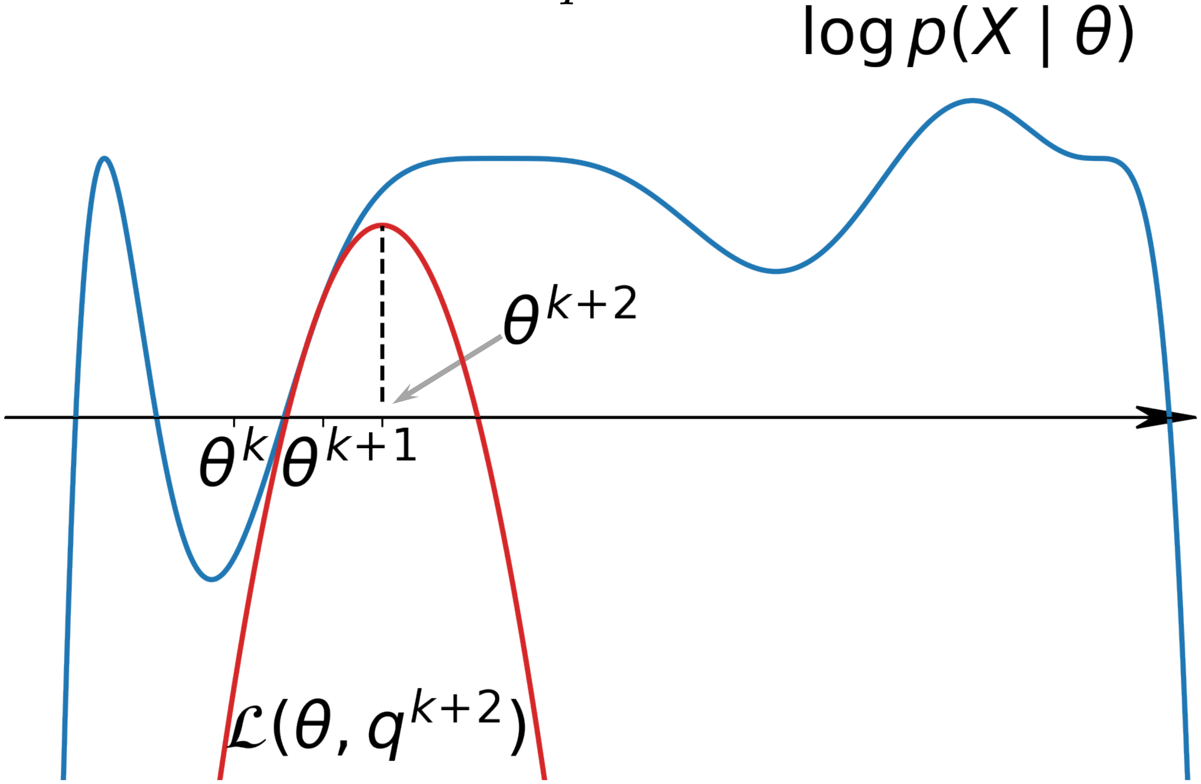
$$q^{k+1} = \arg\max_{q} \mathcal{L}(\theta^k, q)$$



General form of Expectation Maximization

$$\log p(X \mid \theta) \geq \mathcal{L}(\theta, q) \text{ for any q}$$

$$q^{k+1} = \arg\max_q \mathcal{L}(\theta^k, q)$$



Summary of Expectation Maximization

$$\log p(X \mid \theta) \geq \mathcal{L}(\theta, q) \text{ for any q}$$

Variational lower bound

I seeee

E-step

$$q^{k+1} = \arg\max_{q} \mathcal{L}(\theta^k, q)$$

FIX theta k. Then we maximise on q (the weights distribution of the lower bound.)

M-step

$$\theta^{k+1} = \arg\max_{\theta} \mathcal{L}(\theta, q^{k+1})$$

Fix q. Now, we change the theta parameters.