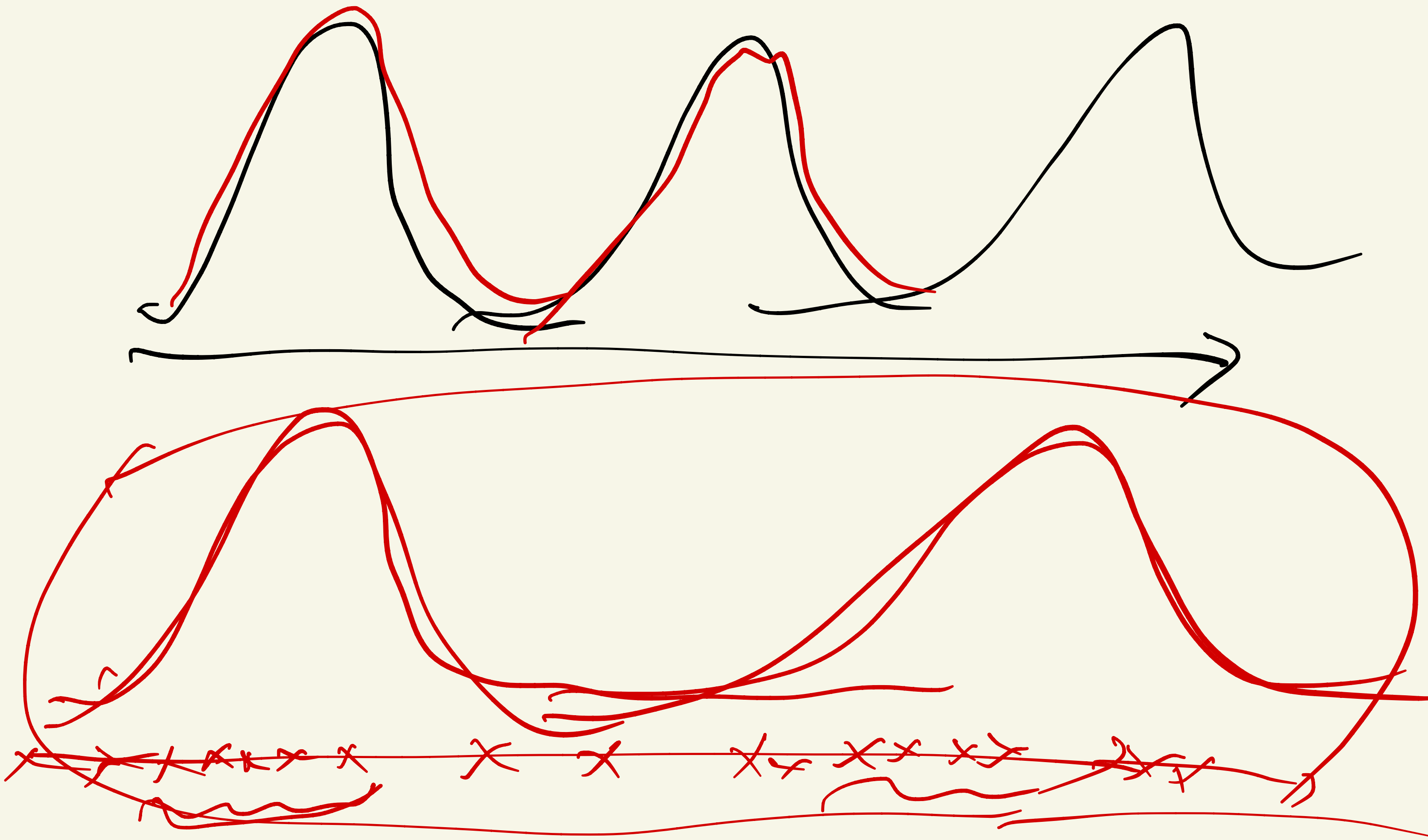
Lecture 12 EM

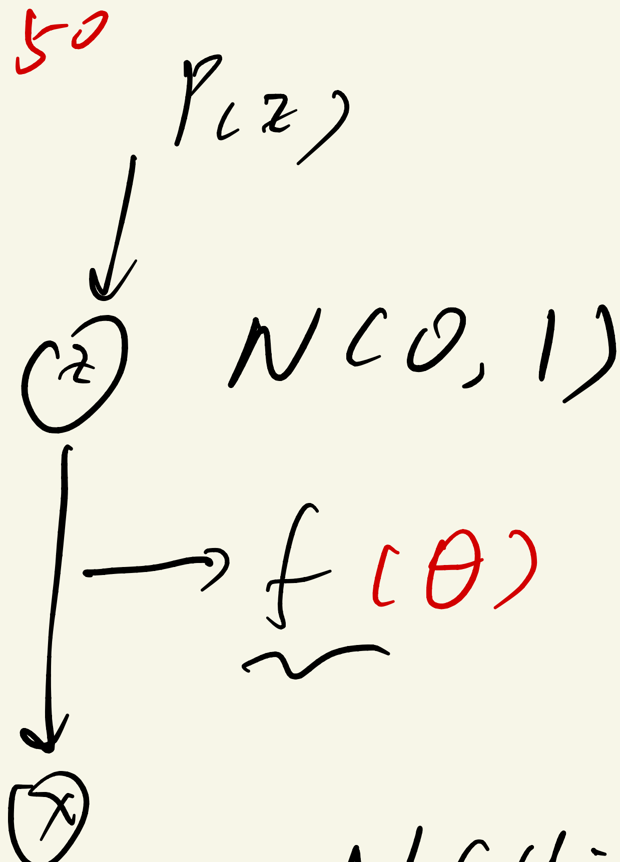
discriminative

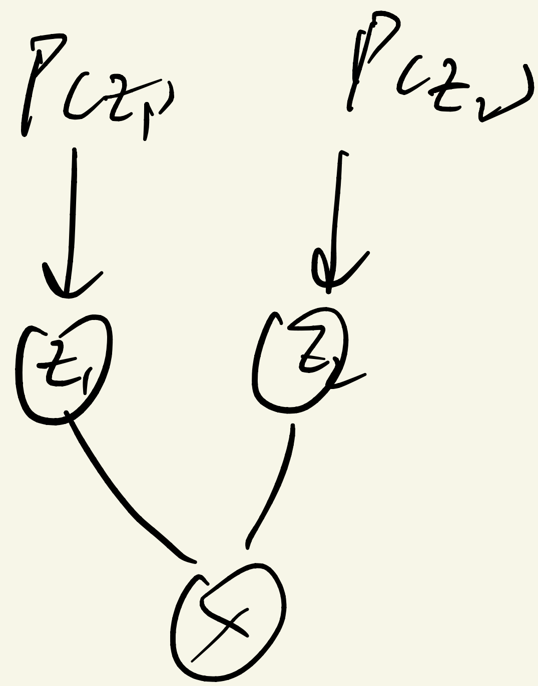
$P(x, y)$

$P(y|x)$



$K = 10$





$$x \sim N(\mu = f(z), 1)$$

$$P(x) = \int_z P(z)\, P(x|z)$$

$P(x)$

$$\log P(x) = \log \sum_z \underbrace{P(x,z)}$$

$$= \log \underbrace{\sum_z}_{} \underbrace{P(z)} \underbrace{P(x|z)}$$

2

$$\log E_{z\sim P(z)} P(x|z)$$

MC sampling

$$P(z|x) \propto P(z)P(x|z)$$

$$P(z=1|x)$$

$$P(z=2|x)$$

$$\mathbb{1}(z=j) \longrightarrow P(z=j|x)$$

$$P(x, \underset{\sim}{z}; \theta)$$

$l(\theta) \geq$ lower bound

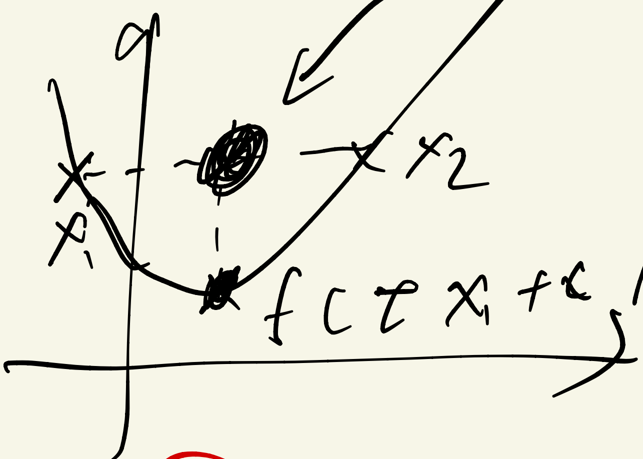
$$\sum_{z} Q(z) = 1$$

$$\log P(x;\theta) = \log \sum_z P(x,z;\theta)$$

$$= \log \sum_z Q_z \frac{P(x,z;\theta)}{Q(z)}$$

concave

$$f(t x_1 + (1-t)x_2) \le t f(x_1) + (1-t) f(x_2)$$



t

$f(E(x)) \le E[f(x)]$

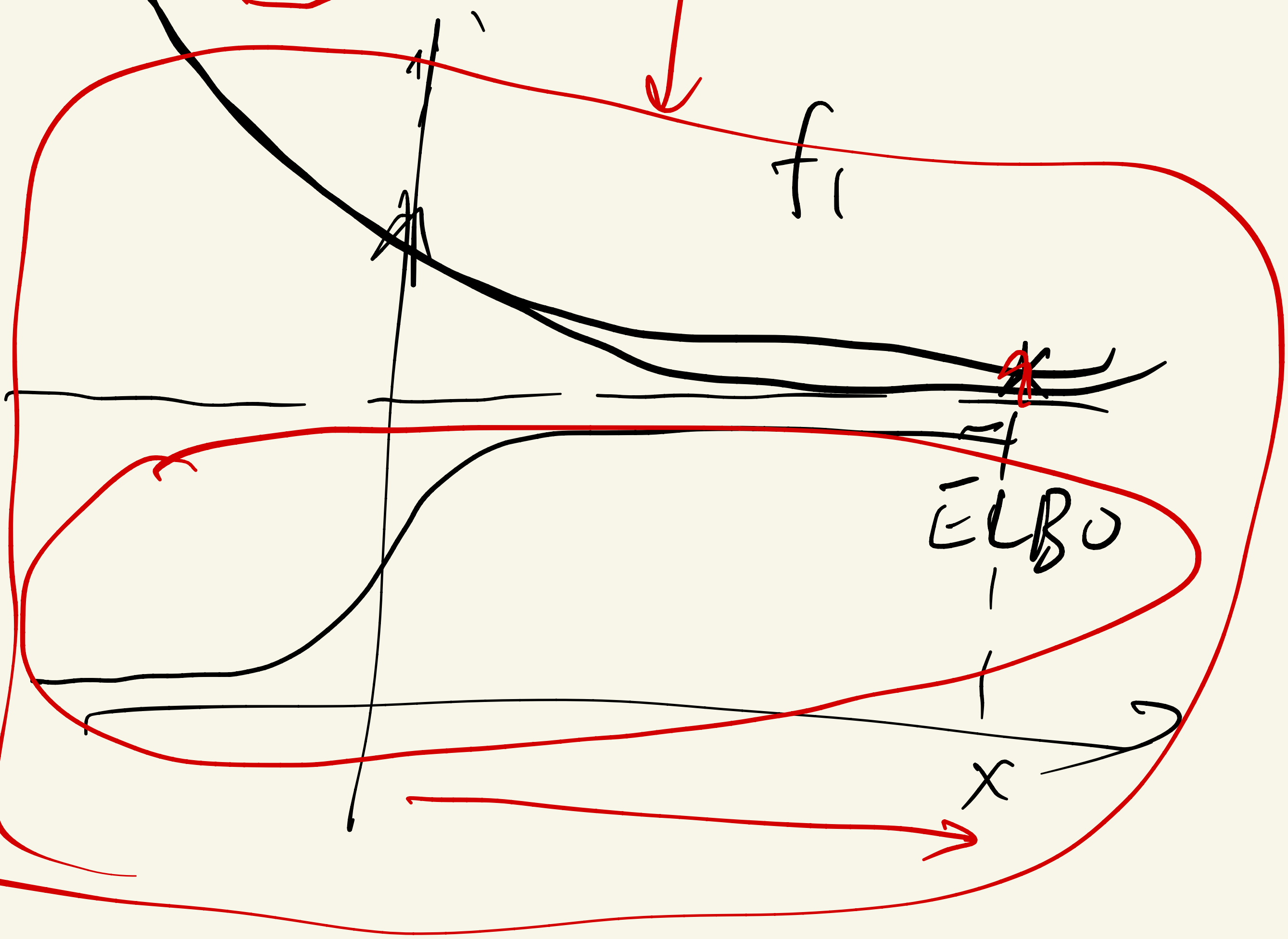
$$0 \le t \le 1$$

$$f(t_1 x_1 + t_2 x_2 + t_3 x_3 + \cdots t_n x_n) \le$$

$$t_1 f(x_1) + t_2 f(x_2) + \cdots t_n f(x_n)$$

$$\underbrace{(t_1 + t_2 \cdots t_N) = 1}_{x = c}$$

$f_1 \geq \text{ELBO}$



$$\boxed{\frac{P(x, z; \theta)}{Q(z)} = c}$$ not related to z

$$Q(z) = \frac{P(x, z; \theta)}{c}$$

$$\overbrace{\sum_z Q(z) = 1}$$

$$= \frac{P(x, z; \theta)}{\underbrace{\sum_z P(x, z; \theta)}_{P(x)}}$$

$$= \boxed{\frac{P(x, z; \theta)}{P(x)}}$$

$$= P(z \mid x; \theta)$$

$$ELBO = \sum_z Q(z) \log \frac{P(x, z; \theta)}{Q(z)}$$

$$= \sum_z P(z|x) \log \frac{P(x, z; \theta)}{P(z|x)}$$

$$= \sum_z P(z|x) \log P(x)$$

$$= \log P(x) \underbrace{\sum_z P(z|x)}_{1}$$

$$= \log P(x)$$

$Q(z) = P(z \mid x; \theta)$

$\log P(x;\theta) \geq \underbrace{\text{ELBO}(x, Q(z), \theta)}$

θ is fixed

doesn't change

$\log P(x;\theta)$

$\underline{Q(z)} = P(z \mid x; \theta_{current})$

ELBO

$\max \int Q(z) \log \frac{P(x,z)}{Q(z)}$ by gradient descent

max ELBO by EM

$\log P(x) \geq \text{ELBO}$

$\int \log P(x)$

$\mathbb{E}_{data} \log P_{data}(x)$

$\int \mathbb{E}_{data} \log P(x)$

direct max:

$$E_{z\sim p(z)}\, p(x|z;\theta)$$

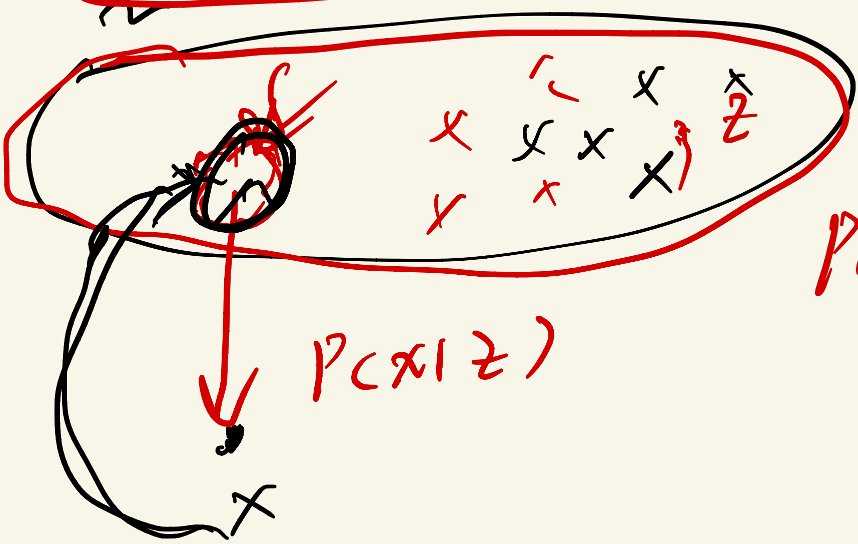
x is given

$$= \sum_z p(z)\, p(x|z)$$

$p(z)$

$$= \sum_z p(z,x)$$

z space



$$E_{z\sim p(z|x)} \log p(x,z;\theta)$$

$$E_{z\sim p(z)}$$

$$ELBO = E_{z\sim Q}\left[\log \frac{P(x,z)}{Q(z)}\right]$$

$$= E_{z\sim Q}\left[\log P(x,z) - \log Q(z)\right]$$

$$= E_{z\sim Q}\left[\log P(x|z) + \log P(z) - \log Q(z)\right]$$

$$= \underbrace{E_{z\sim Q}\log P(x|z)}\ -\ \underbrace{E_{z\sim Q}\log \frac{Q(z)}{P(z)}}$$

KL divergen

x is given

small

Q(z) so that we can generate observed x

$Q(z) \rightarrow P(z)$

$$\log P(x) \geq ELBO$$

$$\log P(x) = ELBO + \underbrace{f}_{\geq 0}$$

$$f = \log P(x) - ELBO$$

$$= \log P(x) - E_{z\sim Q(z)} \log \frac{P(x,z)}{Q(z)}$$

$$= \log \cancel{P(x)} - E_{z\sim Q(z)} \log P(z|x)$$

$$\underbrace{- E_{z\sim Q(z)} \log P(x)}_{\log \cancel{P(x)}} + E_{z\sim Q(z)} \log Q(z)$$

$$= E_{z\sim Q(z)} \log \frac{Q(z)}{P(z|x)} = KL(Q(z) \,||\, P(z|x))$$

$$\underbrace{ELBO}_{} = \underbrace{\log P(x)} - \underbrace{KL[Q(z) \| P(z|x)]}_{\geq 0}$$

$$\underset{\underbrace{Q(z)}}{\arg\max} \ \underbrace{ELBO}$$

$$Q(z) = P(z|x)$$

equivalent to

$$\underset{Q(z)}{\min} \ KL(Q(z) \| P(z|x))$$

$$Q(z) \longrightarrow P(z|x)$$

$$E_{z \sim Q(z)} \log \frac{Q(z)}{P(z|x)} \geq 0$$

E-step

$$Q(z) = P(z|x)$$

$$p(z|x) \propto p(z)\,p(x|z)$$