EM algorithm in one slide

X — observed
Z — latent
θ — parameters

$$\log p(X \mid \theta) \to \max_{\theta}$$
$$\vee\!|$$
$$\mathcal{L}(q, \theta) = \mathbb{E}_{q(Z)} \log p(X, Z \mid \theta) - \mathbb{E}_{q(Z)} \log q(Z) \to \max_{q,\theta}$$

EM algorithm in one slide

X — observed
Z — latent
θ — parameters

$$\log p(X \mid \theta) \to \max_\theta$$
$$\vee\!\mid$$
$$\mathcal{L}(q, \theta) = \mathbb{E}_{q(Z)} \log p(X, Z \mid \theta) - \mathbb{E}_{q(Z)} \log q(Z) \to \max_{q,\theta}$$

Iterations:

E-step: $$q(Z) = \arg\max_q \mathcal{L}(q, \theta) = \arg\min_q KL(q||p) = p(Z \mid X, \theta)$$

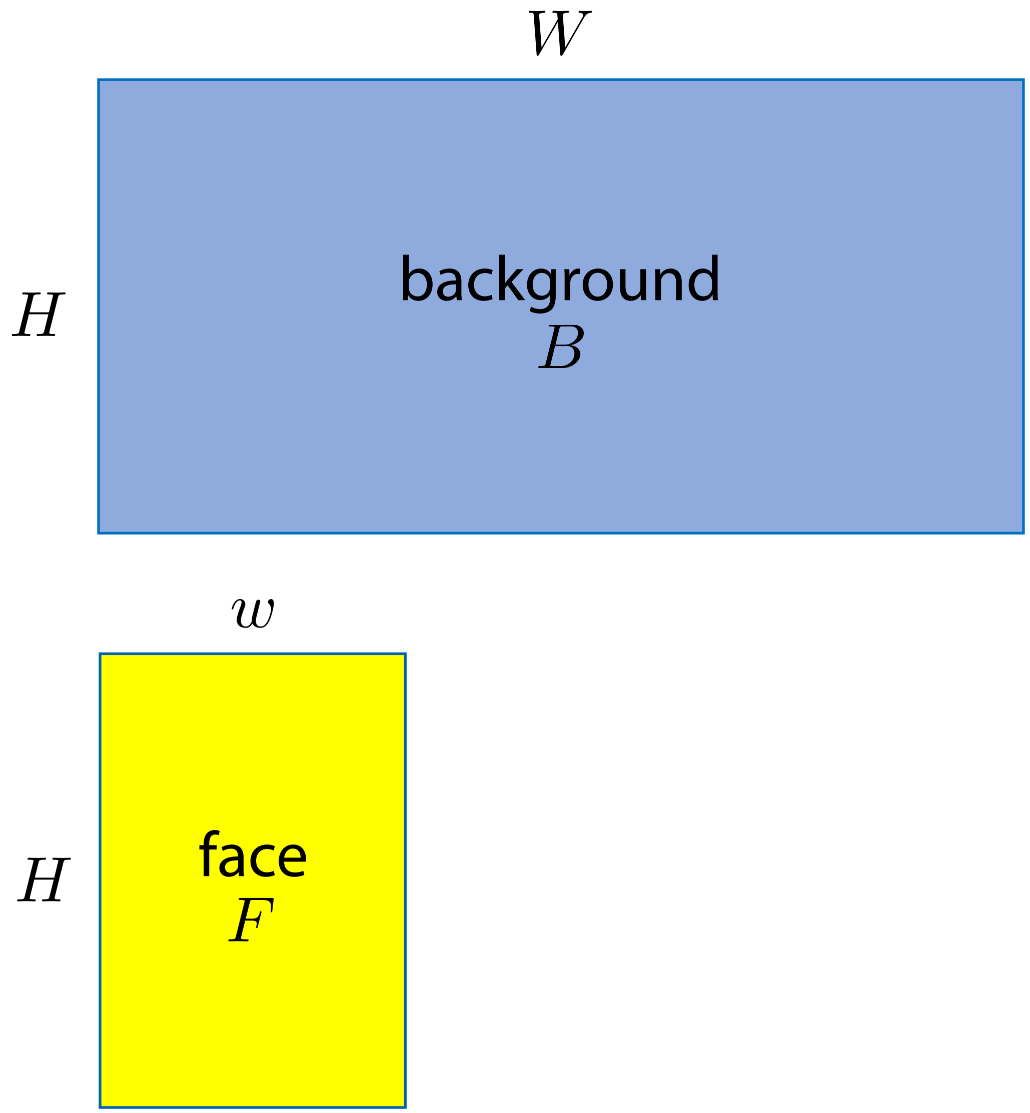
M-step: $$\theta = \arg\max_\theta \mathcal{L}(q, \theta) = \arg\max_\theta \mathbb{E}_{q(Z)} \log p(X, Z \mid \theta)$$

Stopping criteria: $$\mathcal{L}(q^{(t+1)}, \theta^{(t+1)}) - \mathcal{L}(q^{(t)}, \theta^{(t)}) < tol$$

Data and Notation

$B \in \mathbb{R}^{H\times W}$ — clean background image

$F \in \mathbb{R}^{H\times w}$ — clean face image



Data and Notation

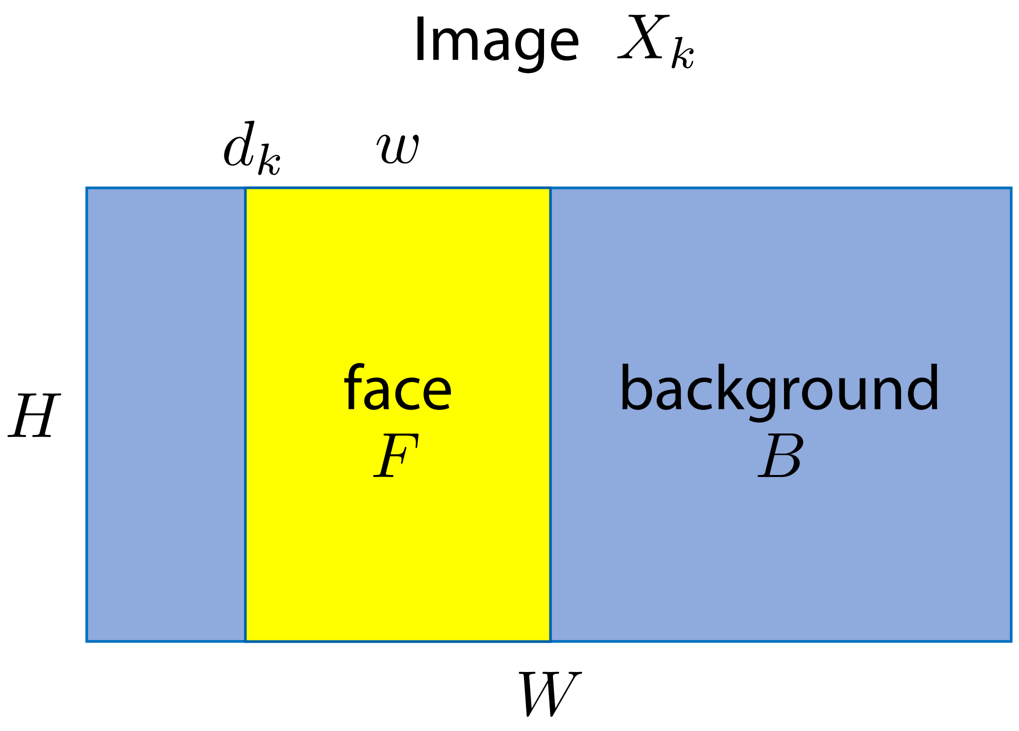
$B \in \mathbb{R}^{H \times W}$ — clean background image

$F \in \mathbb{R}^{H \times w}$ — clean face image

X_k — k-th image from the dataset

d_k — coordinate of the upper-left corner of the face on the k-th image

All images contain the whole face!



Data and Notation

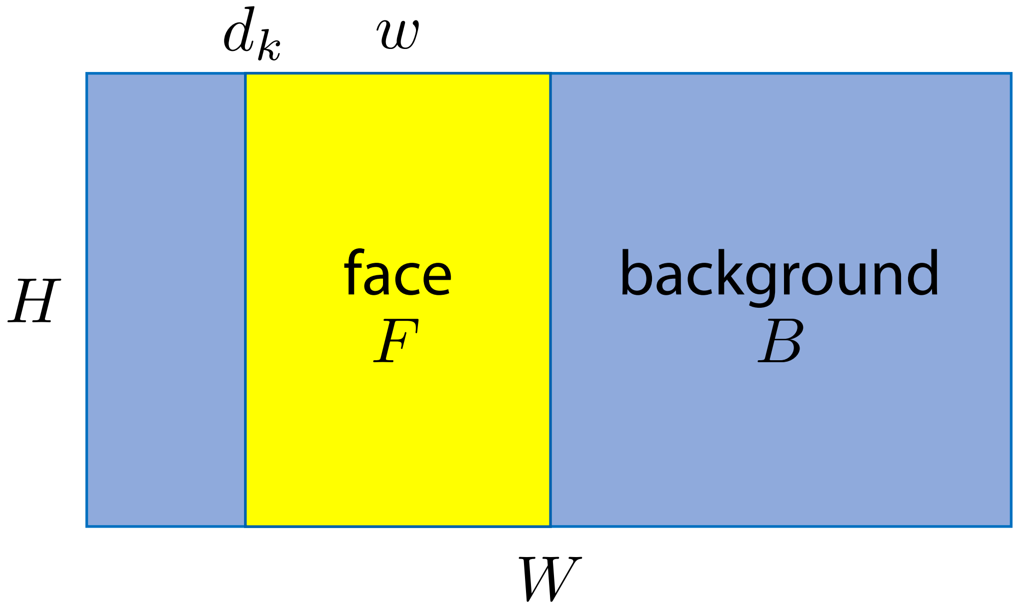
$B \in \mathbb{R}^{H \times W}$ — clean background image

$F \in \mathbb{R}^{H \times w}$ — clean face image

X_k — k-th image from the dataset

d_k — coordinate of the upper-left corner of the face on the k-th image

Image X_k



+ noise from $\mathcal{N}(0, s^2)$

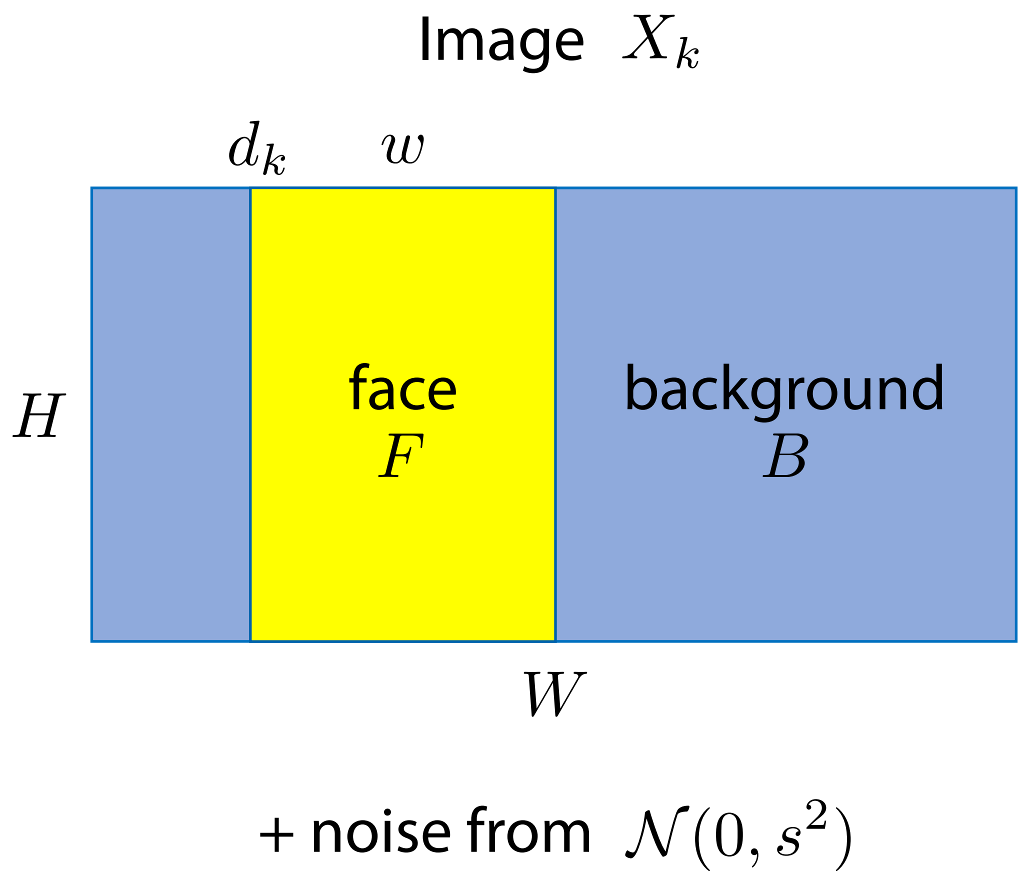
All images contain the whole face!

Probabilistic Model

Observed: **?**

Latent: **?**

Parameters: **?**



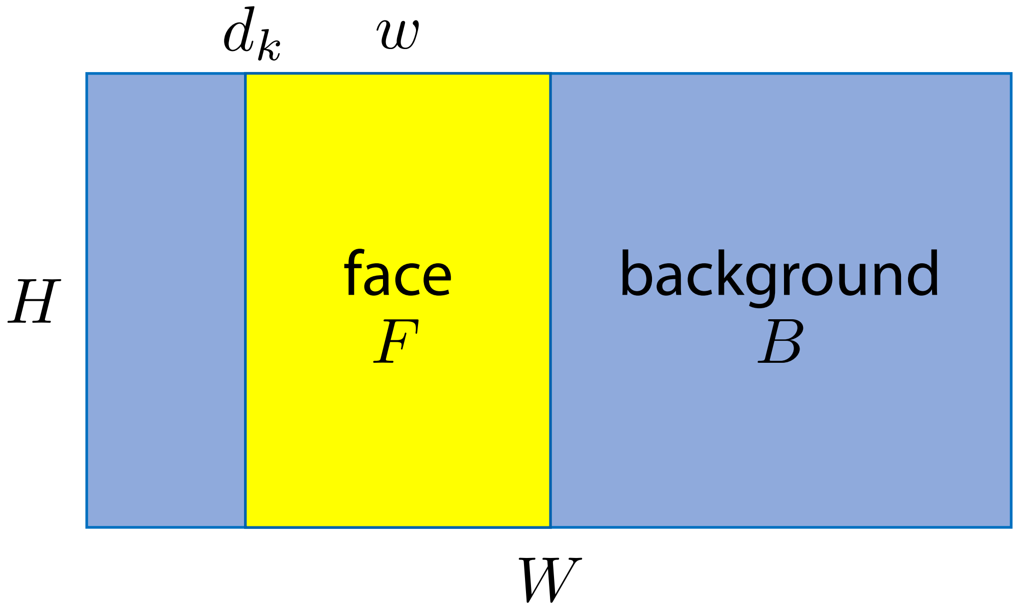
Probabilistic Model

Observed: $X = \{X_1, \ldots, X_K\}$

Latent: **?**

Parameters: **?**

Image X_k



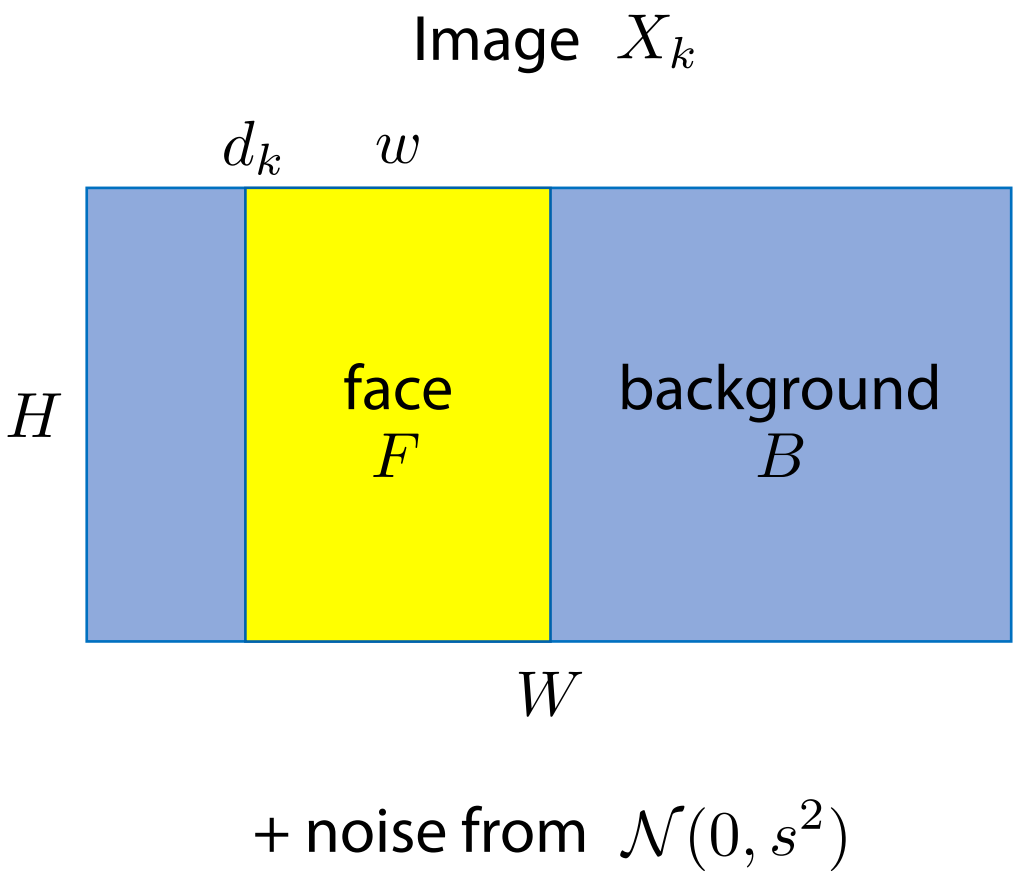
+ noise from $\mathcal{N}(0, s^2)$

Probabilistic Model

Observed: $X = \{X_1, \ldots, X_K\}$

Latent: $d = \{d_1, \ldots, d_K\}$

Parameters: **?**

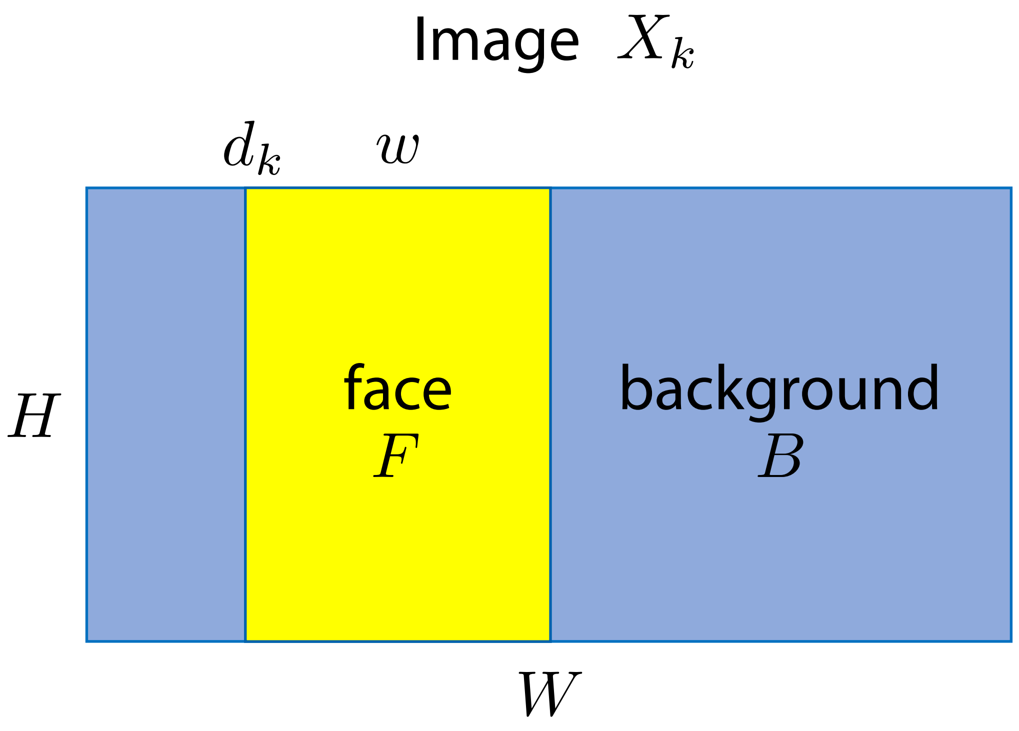


Probabilistic Model

Observed: $X = \{X_1, \ldots, X_K\}$

Latent: $d = \{d_1, \ldots, d_K\}$

Parameters: $\theta = \{B, F, s^2\}$



+ noise from $\mathcal{N}(0, s^2)$

Probabilistic Model

Generation of one image:

$$p(X_k \mid d_k, \theta) = \ \textbf{?}$$

Probabilistic Model

Generation of one image:

$$p(X_k \mid d_k, \theta) = \prod_{ij} \begin{cases} \mathcal{N}(X_k[i,j] \mid F[i, j - d_k], s^2), & \text{if } [i,j] \in faceArea(d_k) \\ \mathcal{N}(X_k[i,j] \mid B[i,j], s^2), & \text{otherwise} \end{cases}$$

What else do we need?

Probabilistic Model

Generation of one image:

$$p(X_k \mid d_k, \theta) = \prod_{ij} \begin{cases} \mathcal{N}(X_k[i,j] \mid F[i, j - d_k], s^2), & \text{if } [i,j] \in faceArea(d_k) \\ \mathcal{N}(X_k[i,j] \mid B[i,j], s^2), & \text{otherwise} \end{cases}$$

Prior on face positions:

$$p(d_k \mid a) = a[d_k], \qquad \sum_j a[j] = 1, \qquad a \in \mathbb{R}^{W-w+1}$$

Probabilistic Model

Generation of one image:

$$p(X_k \mid d_k, \theta) = \prod_{ij} \begin{cases} \mathcal{N}(X_k[i,j] \mid F[i, j-d_k], s^2), & \text{if } [i,j] \in faceArea(d_k) \\ \mathcal{N}(X_k[i,j] \mid B[i,j], s^2), & \text{otherwise} \end{cases}$$

Prior on face positions:

$$p(d_k \mid a) = a[d_k], \qquad \sum_j a[j] = 1, \qquad a \in \mathbb{R}^{W-w+1}$$

Joint probabilistic model:

$$p(X, d \mid \theta, a) = \prod_k p(X_k \mid d_k, \theta) p(d_k \mid a)$$

Task overview

X — observed
d — latent
θ, a — parameters

$$\log p(X \mid \theta, a) \to \max_{\theta, a}$$
$$\vee|$$
$$\mathcal{L}(q, \theta, a) = \mathbb{E}_{q(d)} \log p(X, d \mid \theta, a) - \mathbb{E}_{q(d)} \log q(d) \to \max_{q, \theta, a}$$

Iterations:

E-step: $q(d) = p(d \mid X, \theta, a)$

M-step: $\mathbb{E}_{q(d)} \log p(X, d \mid \theta, a) \to \max_{\theta, a}$

Stopping criteria: $\mathcal{L}(q^{(t+1)}, \theta^{(t+1)}, a^{(t+1)}) - \mathcal{L}(q^{(t)}, \theta^{(t)}, a^{(t)}) < tol$

E-step

PROBLEM

$$q(d) = p(d \mid X, \theta, a) = \quad \textbf{?}$$

E-step

$$q(d) = p(d \mid X, \theta, a) = \prod_k p(d_k \mid X_k, \theta, a)$$

E-step

SOLUTION

$$q(d) = p(d \mid X, \theta, a) = \prod_k p(d_k \mid X_k, \theta, a)$$

$$= \prod_k \frac{p(X_k, d_k \mid \theta, a)}{\sum_{d'_k} p(X_k, d'_k \mid \theta, a)}$$

E-step

SOLUTION

$$q(d) = p(d \mid X, \theta, a) = \prod_k p(d_k \mid X_k, \theta, a)$$

$$= \prod_k \frac{p(X_k, d_k \mid \theta, a)}{\sum_{d'_k} p(X_k, d'_k \mid \theta, a)}$$

$$= \prod_k \frac{p(X_k \mid d_k, \theta) p(d_k \mid a)}{\sum_{d'_k} p(X_k \mid d'_k, \theta) p(d'_k \mid a)}$$

$p(X_k \mid d_k, \theta), \quad p(d_k \mid a)$ — we know from the probabilistic model

E-step

SOLUTION

$$q(d) = p(d \mid X, \theta, a) = \prod_k p(d_k \mid X_k, \theta, a)$$

$$= \prod_k \frac{p(X_k, d_k \mid \theta, a)}{\sum_{d'_k} p(X_k, d'_k \mid \theta, a)}$$

$$= \prod_k \frac{p(X_k \mid d_k, \theta) p(d_k \mid a)}{\sum_{d'_k} p(X_k \mid d'_k, \theta) p(d'_k \mid a)}$$

$p(X_k \mid d_k, \theta), \quad p(d_k \mid a)$ — we know from the probabilistic model

In practice for each object k:

$$q(d_k) \propto p(X_k \mid d_k, \theta) p(d_k \mid a) \in \mathbb{R}^{W-w+1}, \qquad \sum_j q(d_k = j) = 1$$

M-step: function

PROBLEM

$$Q(\theta, a) = \mathbb{E}_{q(d)} \log p(X, d \mid \theta, a) \to \max_{\theta, a}$$

Let's first simplify Q and rewrite it as a function of individual parameters:

$$\mathbb{E}_{q(d)} \quad \to \quad \mathbb{E}_{q(d_k)} \qquad\qquad \theta, a \quad \to \quad F, B, s^2, a$$

M-step: function

$$Q(\theta, a) = \mathbb{E}_{q(d)} \log p(X, d \mid \theta, a) \to \max_{\theta, a}$$

Let's first simplify Q and rewrite it as a function of individual parameters:

$$\boxed{\mathbb{E}_{q(d)} \quad \to \quad \mathbb{E}_{q(d_k)}} \qquad\qquad \theta, a \quad \to \quad F, B, s^2, a$$

$$Q(\theta, a) = \mathbb{E}_{q(d)} \log \prod_k p(X_k \mid d_k, \theta) p(d_k \mid a) = \mathbb{E}_{q(d)} \sum_k \Big(\log p(X_k \mid d_k, \theta) + \log p(d_k \mid a) \Big)$$

$$= \sum_k \mathbb{E}_{q(d)} \Big(\log p(X_k \mid d_k, \theta) + \log p(d_k \mid a) \Big) = \sum_k \mathbb{E}_{q(d_k)} \Big(\log p(X_k \mid d_k, \theta) + \log p(d_k \mid a) \Big)$$

M-step: function

SOLUTION

$$Q(\theta, a) = \mathbb{E}_{q(d)} \log p(X, d \mid \theta, a) \to \max_{\theta, a}$$

Let's first simplify Q and rewrite it as a function of individual parameters:

$$\mathbb{E}_{q(d)} \quad \to \quad \mathbb{E}_{q(d_k)} \qquad\qquad \boxed{\theta, a \quad \to \quad F, B, s^2, a}$$

$$Q(\theta, a) = \sum_k \mathbb{E}_{q(d_k)} \Big(\log p(X_k \mid d_k, \theta) + \log p(d_k \mid a) \Big)$$

M-step: function

$$Q(\theta, a) = \mathbb{E}_{q(d)} \log p(X, d \mid \theta, a) \to \max_{\theta, a}$$

Let's first simplify Q and rewrite it as a function of individual parameters:

$$\mathbb{E}_{q(d)} \quad \to \quad \mathbb{E}_{q(d_k)} \qquad\qquad \boxed{\theta, a \quad \to \quad F, B, s^2, a}$$

$$Q(\theta, a) = \sum_k \mathbb{E}_{q(d_k)} \Big(\log p(X_k \mid d_k, \theta) + \log p(d_k \mid a) \Big)$$

$$\log p(X_k \mid d_k, \theta) = \sum_{i,j} \Big(-\log(\sqrt{2\pi} s) - \frac{1}{2s^2} \Big[(X_k[i,j] - B[i,j])^2 I^B_{ijd_k} + (X_k[i,j] - F[i, j - d_k])^2 I^F_{ijd_k} \Big] \Big)$$

$$I^F_{ijd_k} = \mathbb{I}\big([i,j] \in faceArea(d_k)\big), \qquad I^B_{ijd_k} = \mathbb{I}\big([i,j] \notin faceArea(d_k)\big) = 1 - I^F_{ijd_k}$$

SOLUTION

M-step: function

SOLUTION

$$Q(\theta, a) = \mathbb{E}_{q(d)} \log p(X, d \mid \theta, a) \to \max_{\theta, a}$$

Let's first simplify Q and rewrite it as a function of individual parameters:

$$\mathbb{E}_{q(d)} \quad \to \quad \mathbb{E}_{q(d_k)} \qquad\qquad \boxed{\theta, a \quad \to \quad F, B, s^2, a}$$

$$Q(\theta, a) = \sum_k \mathbb{E}_{q(d_k)} \Big(\log p(X_k \mid d_k, \theta) + \log p(d_k \mid a) \Big)$$

$$\log p(X_k \mid d_k, \theta) = \sum_{i,j} \Big(-\log(\sqrt{2\pi} s) - \frac{1}{2s^2} \Big[(X_k[i,j] - B[i,j])^2 I^B_{ijd_k} + (X_k[i,j] - F[i, j - d_k])^2 I^F_{ijd_k} \Big] \Big)$$

$$I^F_{ijd_k} = \mathbb{I}\big([i,j] \in faceArea(d_k)\big), \qquad I^B_{ijd_k} = \mathbb{I}\big([i,j] \notin faceArea(d_k)\big) = 1 - I^F_{ijd_k}$$

$$\log p(d_k \mid a) = \log a[d_k]$$

M-step: maximisation w.r.t. a

PROBLEM

$$\begin{cases} Q(F,B,s^2,a) = \sum_k \mathbb{E}_{q(d_k)}\Big[\log a[d_k] + \sum_{i,j}\Big(-\log(\sqrt{2\pi}s) - \\ \qquad\qquad - \frac{1}{2s^2}\Big[(X_k[i,j]-B[i,j])^2 I^B_{ijd_k} + (X_k[i,j]-F[i,j-d_k])^2 I^F_{ijd_k}\Big]\Big)\Big] \to \max_a \\ \sum_j a[j] = 1 \end{cases}$$

M-step: maximisation w.r.t. a

SOLUTION

$$\begin{cases} Q(F,B,s^2,a) = \sum_k \mathbb{E}_{q(d_k)} \Big[\boxed{\log a[d_k]} + \sum_{i,j} \Big(-\log(\sqrt{2\pi}s) - \\ \qquad - \frac{1}{2s^2} \Big[(X_k[i,j] - B[i,j])^2 I^B_{ijd_k} + (X_k[i,j] - F[i,j-d_k])^2 I^F_{ijd_k} \Big] \Big) \Big] \to \max_a \\ \sum_j a[j] = 1 \end{cases}$$

M-step: maximisation w.r.t. a

$$\begin{cases} Q(a) = \sum_k \mathbb{E}_{q(d_k)} \log a[d_k] \to \max_a \\ \sum_j a[j] = 1 \end{cases}$$

M-step: maximisation w.r.t. a

$$\begin{cases} Q(a) = \sum_k \mathbb{E}_{q(d_k)} \log a[d_k] \to \max_a \\ \sum_j a[j] = 1 \end{cases}$$

SOLUTION

The Lagrangian has form:

$$L(a, \lambda) = Q(a) - \lambda(\sum_j a[j] - 1)$$

M-step: maximisation w.r.t. a

SOLUTION

$$\begin{cases} Q(a) = \sum_k \mathbb{E}_{q(d_k)} \log a[d_k] \to \max_a \\ \sum_j a[j] = 1 \end{cases}$$

The Lagrangian has form:

$$L(a, \lambda) = Q(a) - \lambda(\sum_j a[j] - 1)$$

$$0 = \frac{\partial L(a, \lambda)}{\partial a[j]} = \sum_k \frac{q(d_k = j)}{a[j]} - \lambda \quad \Rightarrow \quad a[j] = \frac{\sum_k q(d_k = j)}{\lambda}$$

$$0 = \frac{\partial L(a, \lambda)}{\partial \lambda} = \sum_j a[j] - 1 \quad \Rightarrow \quad \lambda = \sum_{j,k} q(d_k = j)$$

M-step: maximisation w.r.t. a

SOLUTION

$$\begin{cases} Q(a) = \sum_k \mathbb{E}_{q(d_k)} \log a[d_k] \to \max_a \\ \sum_j a[j] = 1 \end{cases}$$

The Lagrangian has form:

$$L(a, \lambda) = Q(a) - \lambda(\sum_j a[j] - 1)$$

$$0 = \frac{\partial L(a,\lambda)}{\partial a[j]} = \sum_k \frac{q(d_k = j)}{a[j]} - \lambda \quad \Rightarrow \quad a[j] = \frac{\sum_k q(d_k = j)}{\lambda}$$

$$0 = \frac{\partial L(a,\lambda)}{\partial \lambda} = \sum_j a[j] - 1 \quad \Rightarrow \quad \lambda = \sum_{j,k} q(d_k = j)$$

In practice: $a[j] \propto \sum_k q(d_k = j), \qquad \sum_j a[j] = 1$

M-step: maximisation w.r.t. F

PROBLEM

$$Q(F, B, s^2, a) = \sum_k \mathbb{E}_{q(d_k)} \Big[\log a[d_k] + \sum_{i,j} \Big(-\log(\sqrt{2\pi} s) -$$

$$- \frac{1}{2s^2} \Big[(X_k[i,j] - B[i,j])^2 I^B_{ijd_k} + (X_k[i,j] - F[i, j - d_k])^2 I^F_{ijd_k} \Big] \Big) \Big] \rightarrow \max_F$$

M-step: maximisation w.r.t. F

SOLUTION

$$Q(F, B, s^2, a) = \sum_k \mathbb{E}_{q(d_k)} \Big[\log a[d_k] + \sum_{i,j} \Big(-\log(\sqrt{2\pi} s) - \\ - \frac{1}{2s^2} \Big[(X_k[i,j] - B[i,j])^2 I^B_{ijd_k} + \boxed{(X_k[i,j] - F[i, j-d_k])^2 I^F_{ijd_k}} \Big] \Big) \Big] \to \max_F$$

M-step: maximisation w.r.t. F

$$Q(F) = \sum_k \mathbb{E}_{q(d_k)} \left[\sum_{i,j} \left(-\frac{1}{2s^2} (X_k[i,j] - F[i, j - d_k])^2 I^F_{ijd_k} \right) \right] \to \max_F$$

SOLUTION

M-step: maximisation w.r.t. F

$$Q(F) = \sum_k \mathbb{E}_{q(d_k)} \left[\sum_{i,j} \left(-\frac{1}{2s^2} (X_k[i,j] - F[i, j-d_k])^2 I^F_{ijd_k} \right) \right]$$

$$= \sum_k \mathbb{E}_{q(d_k)} \left[\sum_{i=0,m=0}^{H-1,w-1} \left(-\frac{1}{2s^2} (X_k[i, m+d_k] - F[i,m])^2 \right) \right] \to \max_F$$

M-step: maximisation w.r.t. F

SOLUTION

$$Q(F) = \sum_k \mathbb{E}_{q(d_k)} \left[\sum_{i,j} \left(-\frac{1}{2s^2} (X_k[i,j] - F[i,j-d_k])^2 I^F_{ijd_k} \right) \right]$$

$$= \sum_k \mathbb{E}_{q(d_k)} \left[\sum_{i=0,m=0}^{H-1,w-1} \left(-\frac{1}{2s^2} (X_k[i,m+d_k] - F[i,m])^2 \right) \right] \to \max_F$$

$$0 = \frac{\partial Q(F)}{\partial F[i,m]} = \sum_{k,d_k} \frac{q(d_k)}{s^2} \Big(X_k[i,m+d_k] - F[i,m] \Big)$$

$$F[i,m] = \frac{\sum_{k,d_k} q(d_k) X_k[i,m+d_k]}{\sum_{k,d_k} q(d_k)} = \frac{\sum_{k,d_k} q(d_k) X_k[i,m+d_k]}{K}$$

M-step: maximisation w.r.t. B

PROBLEM

$$Q(F, B, s^2, a) = \sum_k \mathbb{E}_{q(d_k)} \Big[\log a[d_k] + \sum_{i,j} \Big(-\log(\sqrt{2\pi} s) - \\ - \frac{1}{2s^2} \Big[(X_k[i,j] - B[i,j])^2 I^B_{ijd_k} + (X_k[i,j] - F[i, j - d_k])^2 I^F_{ijd_k} \Big] \Big) \Big] \to \max_B$$

M-step: maximisation w.r.t. B

SOLUTION

$$Q(F, B, s^2, a) = \sum_k \mathbb{E}_{q(d_k)} \Big[\log a[d_k] + \sum_{i,j} \Big(-\log(\sqrt{2\pi}s) - \\ - \frac{1}{2s^2} \Big[\boxed{(X_k[i,j] - B[i,j])^2 I^B_{ijd_k}} + (X_k[i,j] - F[i, j - d_k])^2 I^F_{ijd_k} \Big] \Big) \Big] \to \max_B$$

M-step: maximisation w.r.t. B

$$Q(B) = \sum_k \mathbb{E}_{q(d_k)} \left[\sum_{i,j} \left(-\frac{1}{2s^2} (X_k[i,j] - B[i,j])^2 I^B_{ij d_k} \right) \right] \to \max_B$$

M-step: maximisation w.r.t. B

SOLUTION

$$Q(B) = \sum_k \mathbb{E}_{q(d_k)} \left[\sum_{i,j} \left(-\frac{1}{2s^2} (X_k[i,j] - B[i,j])^2 I^B_{ijd_k} \right) \right] \to \max_B$$

$$0 = \frac{\partial Q(B)}{\partial B[i,j]} = \sum_{k,d_k} \frac{q(d_k) I^B_{ijd_k}}{s^2} \Big(X_k[i,j] - B[i,j] \Big)$$

$$B[i,j] = \frac{\sum_{k,d_k} q(d_k) I^B_{ijd_k} X_k[i,j]}{\sum_{k,d_k} q(d_k) I^B_{ijd_k}}$$

M-step: maximisation w.r.t. s^2

PROBLEM

$$Q(F, B, s^2, a) = \sum_k \mathbb{E}_{q(d_k)} \Big[\log a[d_k] + \sum_{i,j} \Big(-\log(\sqrt{2\pi} s) - \\ - \frac{1}{2s^2} \Big[(X_k[i,j] - B[i,j])^2 I^B_{ijd_k} + (X_k[i,j] - F[i, j - d_k])^2 I^F_{ijd_k} \Big] \Big) \Big] \to \max_{s^2}$$

M-step: maximisation w.r.t. s^2

SOLUTION

$$Q(F, B, s^2, a) = \sum_k \mathbb{E}_{q(d_k)} \left[\log a[d_k] + \sum_{i,j} \left(\boxed{-\log(\sqrt{2\pi} s) -} \right. \right.$$

$$\left. \left. - \boxed{\frac{1}{2s^2} \left[(X_k[i,j] - B[i,j])^2 I^B_{ijd_k} + (X_k[i,j] - F[i, j - d_k])^2 I^F_{ijd_k} \right]} \right) \right] \to \max_{s^2}$$

M-step: maximisation w.r.t. s^2

SOLUTION

$$Q(s^2) = \sum_k \mathbb{E}_{q(d_k)} \Big[\sum_{i,j} \Big(-\frac{1}{2}\log(s^2) - \\ - \frac{1}{2s^2}\Big[(X_k[i,j] - B[i,j])^2 I^B_{ijd_k} + (X_k[i,j] - F[i,j-d_k])^2 I^F_{ijd_k} \Big] \Big) \Big] \to \max_{s^2}$$

M-step: maximisation w.r.t. s^2

SOLUTION

$$Q(s^2) = \sum_k \mathbb{E}_{q(d_k)}\Big[\sum_{i,j}\Big(-\frac{1}{2}\log(s^2) - \frac{1}{2s^2}\Big[(X_k[i,j]-B[i,j])^2 I^B_{ijd_k} + (X_k[i,j]-F[i,j-d_k])^2 I^F_{ijd_k}\Big]\Big)\Big] \to \max_{s^2}$$

$$0 = \frac{\partial Q(s^2)}{\partial s^2} = \sum_{k,d_k,i,j} q(d_k)\Big(-\frac{1}{2s^2} + \frac{1}{2s^4}\Big[(X_k[i,j]-B[i,j])^2 I^B_{ijd_k} + (X_k[i,j]-F[i,j-d_k])^2 I^F_{ijd_k}\Big]\Big)$$

$$s^2 = \frac{1}{\sum_{k,d_k,i,j} q(d_k)} \sum_{k,d_k,i,j} q(d_k)\Big[(X_k[i,j]-B[i,j])^2 I^B_{ijd_k} + (X_k[i,j]-F[i,j-d_k])^2 I^F_{ijd_k}\Big]$$

$$= \frac{1}{NWH} \sum_{k,d_k,i,j} q(d_k)\Big[(X_k[i,j]-B[i,j])^2 I^B_{ijd_k} + (X_k[i,j]-F[i,j-d_k])^2 I^F_{ijd_k}\Big]$$

Stopping criteria

PROBLEM

$$\mathcal{L}(q^{(t+1)}, \theta^{(t+1)}, a^{(t+1)}) - \mathcal{L}(q^{(t)}, \theta^{(t)}, a^{(t)}) < tol$$

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

Stopping criteria

SOLUTION

$$\mathcal{L}(q^{(t+1)}, \theta^{(t+1)}, a^{(t+1)}) - \mathcal{L}(q^{(t)}, \theta^{(t)}, a^{(t)}) < tol$$

$$\mathcal{L}(q, \theta, a) = \mathbb{E}_{q(d)}\Big[\log p(X, d \mid \theta, a) - \log q(d)\Big]$$

Stopping criteria

SOLUTION

$$\mathcal{L}(q^{(t+1)}, \theta^{(t+1)}, a^{(t+1)}) - \mathcal{L}(q^{(t)}, \theta^{(t)}, a^{(t)}) < tol$$

$$\begin{aligned}
\mathcal{L}(q, \theta, a) &= \mathbb{E}_{q(d)}\Big[\log p(X, d \mid \theta, a) - \log q(d)\Big] \\
&= \mathbb{E}_{q(d)}\Big[\log \prod_k p(X_k \mid d_k, \theta) p(d_k \mid a) - \log \prod_k q(d_k)\Big] \\
&= \mathbb{E}_{q(d)} \sum_k \Big[\log p(X_k \mid d_k, \theta) + \log p(d_k \mid a) - \log q(d_k)\Big]
\end{aligned}$$

Stopping criteria

SOLUTION

$$\mathcal{L}(q^{(t+1)}, \theta^{(t+1)}, a^{(t+1)}) - \mathcal{L}(q^{(t)}, \theta^{(t)}, a^{(t)}) < tol$$

$$\begin{aligned}
\mathcal{L}(q, \theta, a) &= \mathbb{E}_{q(d)}\Big[\log p(X, d \mid \theta, a) - \log q(d)\Big] \\
&= \mathbb{E}_{q(d)}\Big[\log \prod_k p(X_k \mid d_k, \theta) p(d_k \mid a) - \log \prod_k q(d_k)\Big] \\
&= \mathbb{E}_{q(d)} \sum_k \Big[\log p(X_k \mid d_k, \theta) + \log p(d_k \mid a) - \log q(d_k)\Big] \\
&= \sum_k \mathbb{E}_{q(d)}\Big[\log p(X_k \mid d_k, \theta) + \log p(d_k \mid a) - \log q(d_k)\Big] \\
&= \sum_k \mathbb{E}_{q(d_k)}\Big[\log p(X_k \mid d_k, \theta) + \log p(d_k \mid a) - \log q(d_k)\Big]
\end{aligned}$$

Task overview

E-step:

$$q(d_k) \propto p(X_k \mid d_k, \theta) p(d_k \mid a), \qquad \sum_j q(d_k = j) = 1$$

M-step:

$$a[j] \propto \sum_k q(d_k = j), \qquad \sum_j a[j] = 1$$

$$F[i,m] = \frac{\sum_{k,d_k} q(d_k) X_k[i, m + d_k]}{K} \qquad B[i,j] = \frac{\sum_{k,d_k} q(d_k) I^B_{ijd_k} X_k[i,j]}{\sum_{k,d_k} q(d_k) I^B_{ijd_k}}$$

$$s^2 = \frac{1}{NWH} \sum_{k,d_k,i,j} q(d_k) \left[(X_k[i,j] - B[i,j])^2 I^B_{ijd_k} + (X_k[i,j] - F[i, j - d_k])^2 I^F_{ijd_k} \right]$$

Stopping criteria:

$$\mathcal{L}(q^{(t+1)}, \theta^{(t+1)}, a^{(t+1)}) - \mathcal{L}(q^{(t)}, \theta^{(t)}, a^{(t)}) < tol$$

$$\mathcal{L}(q, \theta, a) = \sum_k \mathbb{E}_{q(d_k)} \Big[\log p(X_k \mid d_k, \theta) + \log p(d_k \mid a) - \log q(d_k) \Big]$$