EM algorithm for the investigation

Deep|Bayes 2018

1 Story time

As you know, we plan a Game Night on Thursday. Unfortunately, it has been put at risk by one of the school organizers who decided to prank us and hid all games somewhere. We were able to obtain a set of K photographs of the suspect from a surveillance video camera, but all of them have been corrupted by a directed electromagnetic noise (an example is shown in figure 1)). All we know is that in each photo the suspect face is in a random position on the background and the background is the same for all photos. Let's investigate this incident and expose the prankster using the knowledge of Bayesian methods! Top three fastest detectives will receive a trophy =)

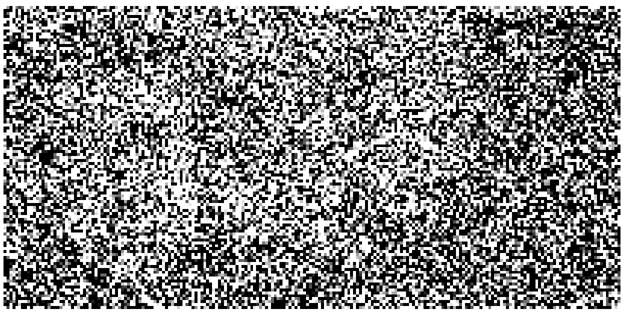

Figure 1: Example of the corrupted image.

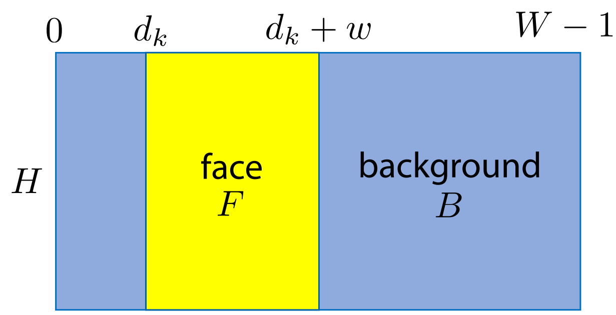

Figure 2: Layout of the image structure for X_k.

2 Model description

You are given samples $\mathbf{X} = \{\mathbf{X}_k\}_{k=1}^{K}$ of very noisy grayscale images of size $H \times W$ pixels. Each image contains the same background of size $H \times W$ and the face of the suspect of size $H \times w$ at an uncertain position. All images contain the whole face. A layout of the image structure is shown in figure 2.

Let us use the following notation:

- $B \in \mathbb{R}^{H \times W}$ – a noiseless background image,
- $F \in \mathbb{R}^{H \times w}$ – a noiseless face of the suspect,
- X_k – k-th image from the dataset, $X = \{X_1, \ldots, X_K\}$ – the whole dataset,
- for indexation of pixels of X_k, B or F we will use square brackets: $X_k[i,j], B[i,j], F[i,j]$,
- d_k – horizontal coordinate of the upper-left corner of the face on the k-th image, $d = \{d_1, \ldots, d_K\}$ – the vector of coordinates for all K images.

Images are corrupted by zero mean Gaussian noise with variance s^2. The noise is independent for all pixels. Thus for image X_k we have the following model:

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

where $\theta = \{B, F, s^2\}$, $faceArea(d_k) = \{[i,j] \mid d_k \leq j \leq d_k + w - 1\}$.

Prior distribution for coordinates of the face in images is determined by a learnable parameter vector $a \in \mathbb{R}^{W-w+1}$:

$$p(\boldsymbol{d}_k \mid a) = a[d_k], \qquad \sum_j a[j] = 1.$$

Finally the joint probabilistic model looks as follows:

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

3 Problem statement

The goal is to restore the face F using the observed images X, therefore you need to solve the following task:

$$p(X \mid \theta, a) \to \max_{\theta, a}.$$

In order to do that EM algorithm should be used. EM algorithm optimizes ELBO:

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

The optimization is done in a block-coordinate manner. We use an iterative procedure with two steps on each iteration: the E-step and the M-step. On the E-step ELBO is optimized w.r.t. the variational distribution $q(d)$:

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

and on the M-step ELBO is optimized w.r.t. parameters θ, a:

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

The iterative procedure converges when the value of ELBO value stops changing:

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

In the theoretical part of the task you need to derive formulas for the following quantities:

- variational distribution $q(d_k \mid X_k, \theta, a)$ for the E-step;
- point estimates for $a, \theta = \{F, B, s^2\}$ for the M-step;
- ELBO $\mathcal{L}(q, \theta, a)$ for stopping criteria.

In the practical part of the task you need to implement EM-algorithm and expose the prankster.

The task is based on the lab from the Pattern Recognition and Machine Learning course at Czech Technical University in Prague.