EECS 545: Homework #5

Mingliang Duanmu
duanmuml@umich.edu

March 31, 2022

1 K-means for image compression

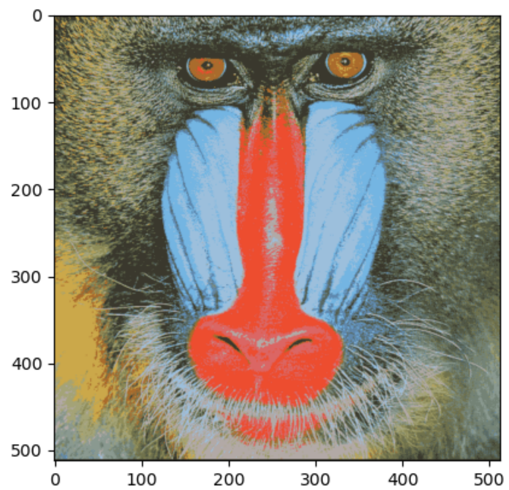
a

Refer to `q1.py`.

b

Refer to `q1.py`.

c



The error is 15.068.

d

0.83.

2 Gaussian mixtures for image compression

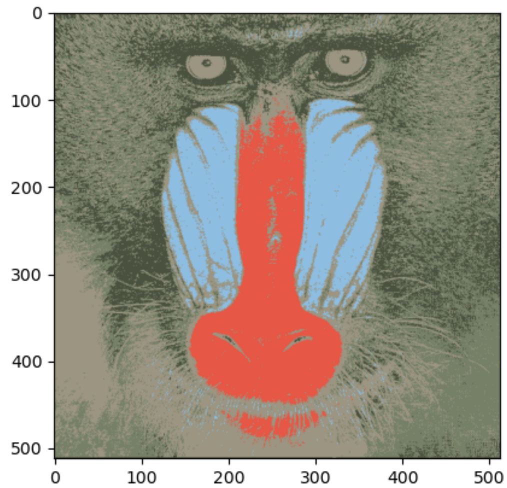
a

Refer to `q2.py`, log-likelihoods are displayed in output.

b

Refer to `q2.py`.

c



The error is 33.227. (μ_k, Σ_k) are displayed in output.

d

0.875.

3 PCA and eigenfaces

a

We have

$$\frac{1}{N}\sum_{n=1}^{N}\left\|x^{(n)}-UU^{T}x^{(n)}\right\|^{2}=\frac{1}{N}\sum_{n=1}^{N}\left[\left(x^{(n)}-UU^{T}x^{(n)}\right)^{T}\left(x^{(n)}-UU^{T}x^{(n)}\right)\right]$$
$$=\frac{1}{N}\sum_{n=1}^{N}\left[x^{(n)T}x^{n}-2x^{(n)T}UU^{T}x^{(n)}+x^{(n)T}UU^{T}U^{T}Ux^{(n)}\right]$$

Because $U_1, U_2, U_3, \cdots, U_K$ are orthonormal, $U^T U = I$, we can simplify the result as

$$\frac{1}{N}\sum_{n=1}^{N}\left\|x^{(n)}-UU^{\top}x^{(n)}\right\|^{2}=\frac{1}{N}\sum_{n=1}^{N}\left[x^{(n)T}x^{n}-x^{(n)T}UU^{T}x^{(n)}\right]$$
$$=\frac{1}{N}\sum_{n=1}^{N}\left[x^{(n)T}x^{n}\right]-\frac{1}{N}\sum_{n=1}^{N}\left[x^{(n)T}UU^{T}x^{(n)}\right]$$

Since $\sum_{n=1}^{N} x^{(n)T}x^{n}$ is constant, we can write the objective function as

$$\max\frac{1}{N}\sum_{n=1}^{N}\left[x^{(n)T}UU^{T}x^{(n)}\right]=\max\frac{1}{N}\sum_{n=1}^{N}\left\|U^{T}x^{(n)}\right\|^{2}$$
$$=\max\frac{1}{N}\sum_{i=1}^{K}\sum_{n=1}^{N}u_i^{T}x^{(n)}x^{(n)T}u_i$$

Since the data is zero-centered, we can simplify the covariance as

$$S=\frac{1}{N}\sum_{n=1}^{N}\left(x^{(n)}-\bar{x}\right)\left(x^{(n)}-\bar{x}\right)^{T}=\frac{1}{N}\sum_{n=1}^{N}\left(x^{(n)}\right)\left(x^{(n)}\right)^{T}$$

So the objective function becomes

$$\max\sum_{i=1}^{K}u_i^{T}Su_i=\max\sum_{i=1}^{K}u_i^{T}Su_i+\lambda_i(1-u_i^{T}u_i)$$

where $u_i^T u_i = 1$ Compute the gradient and set to zero, we find

$$Su_i=\lambda_i u_i$$

which satisfies the definition of eigenvalues, λ_i are the eigenvalues of the covariance matrix S and u_i are the corresponding eigenvectors. At the same time, we have

$$\max\sum_{i=1}^{K}u_i^{T}Su_i=\sum_{i=1}^{K}\lambda_i$$

Therefore, we have

$$\frac{1}{N}\sum_{n=1}^{N}\left[x^{(n)T}x^{n}\right]=\frac{1}{N}\sum_{n=1}^{N}\left\|x^{(n)}\right\|^{2}=\frac{1}{N}\sum_{n=1}^{N}\left\|V^{T}x^{(n)}\right\|^{2}$$

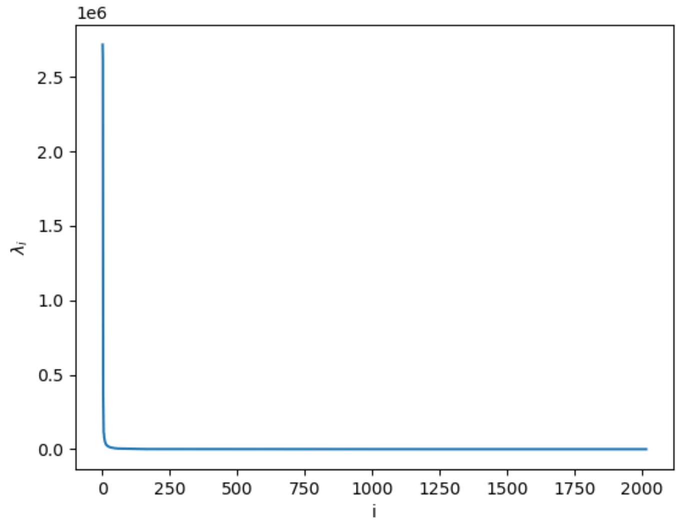
where V is orthogonal matrix.
Let V be the eigenvectors of the covariance matrix, we have

$$\frac{1}{N}\sum_{n=1}^{N}\left\|V^{T}x^{(n)}\right\|^{2}=\frac{1}{N}\sum_{i=1}^{d}\sum_{n=1}^{N}v_i^{T}x^{(n)}x^{(n)T}v_i$$
$$=\sum_{i=1}^{d}v_i^{T}Sv_i$$
$$=\sum_{i=1}^{d}\lambda_i$$

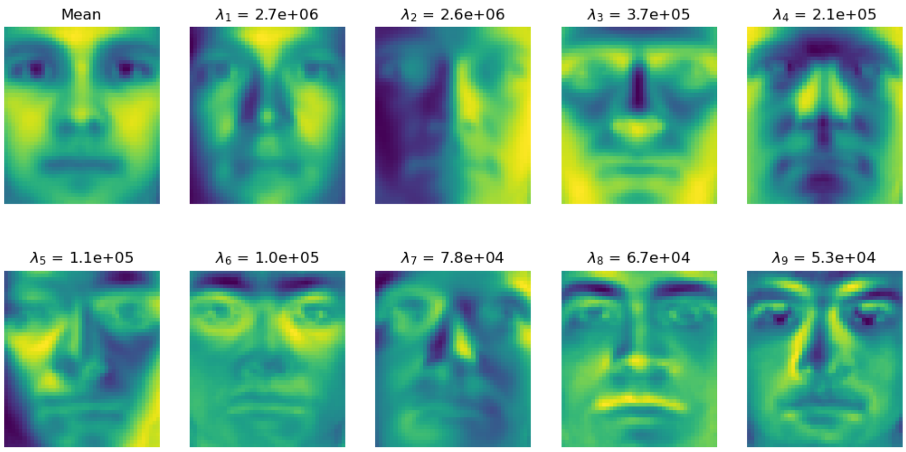
Finally we can conclude that

$$\min_{\mathbf{U}\in\mathcal{U}} \frac{1}{N}\sum_{n=1}^{N}\left\|\mathbf{x}^{(n)} - \mathbf{U}\mathbf{U}^T\mathbf{x}^{(n)}\right\|^2 = \min_{\mathbf{U}\in\mathcal{U}} \frac{1}{N}\sum_{n=1}^{N}\left\|\mathbf{x}^{(n)} - \sum_{i=1}^{K}\mathbf{u}_i\mathbf{u}_i^T\mathbf{x}^{(n)}\right\|^2 = \sum_{k=K+1}^{d}\lambda_k$$

b



c



The fourth principal component captures the light variations on sides of nose; the sixth principal component captures the light variations around the eyes; the eighth principal component captures the feature of lips.

d

43 principal components are needed to represent 95% of the total variance, 97.9% of reduction in dimension.
167 principal components are needed to represent 99% of the total variance, 91.7% of reduction in dimension.

4 Expectation Maximization

a

The marginal distribution of x looks like a normal distribution but with higher peak and longer tails.

b

First we get the joint probability

$$
\begin{aligned}
P(x \mid \theta) &= \prod_{i=1}^{N} p(z) N\left(x \mid \mu, z, \sigma^2\right) \\
&= \phi^{N_0} \pi \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) (1-\phi)^{N_1} \pi \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(\lambda-1-\mu)^2\right) \\
&= \phi^{N_0}(1-\phi)^{N_1} \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(x^n - z^n - \mu\right)^2\right)
\end{aligned}
$$

Converting to log likelihood, we have

$$
\log P(x \mid \theta) = N_0 \log\phi + N_1 \log(1-\phi) + \sum_{i=1}^{N}\left(\log\frac{1}{\sqrt{2\pi}} - \log\sigma - \frac{1}{2\sigma^2}\left(x^n - z^n - \mu\right)^2\right)
$$

To maximize the likelihood, we derive the gradient as

$$
\frac{\partial \log p}{\partial \phi} = \frac{N_0}{\phi} - \frac{N_1}{1-\phi} = 0
$$

where N_0 corresponds to number when $z = 0$, same applies for N_1.
Therefore,

$$
\phi = \frac{N_\Omega}{N_0 + N_1}
$$

Similarly, we have

$$
\frac{\partial \log P}{\partial \mu} = \sum_{i=1}^{N} \frac{1}{\sigma^2}\left(x^n - z^n - \mu\right) = 0
$$

$$
\frac{\partial \log P}{\partial \sigma} = \sum_{i=1}^{N}\left(-\frac{1}{\sigma} + \frac{1}{\sigma^3}\left(x^n - z^n - \mu\right)^2\right) = 0
$$

Therefore, we get

$$
\mu = \frac{\sum_{i=1}^{N} \varepsilon_i}{N} \quad \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(\varepsilon^n - \mu)^2
$$

c

According to EM, we have

$$
\log P(\boldsymbol{x} \mid \theta) = \sum_{\boldsymbol{z}} q(\boldsymbol{z}) \log\frac{p(\boldsymbol{z}, \boldsymbol{x} \mid \theta)}{q(\boldsymbol{z})} + KL(q(\boldsymbol{z}) \| p(\boldsymbol{z} \mid \boldsymbol{x}, \theta))
$$

For the E-step,

$$
q^n\left(z_k\right) = P\left(z_k \mid x^n\right) = \frac{P\left(z_k, x^n\right)}{P\left(x^n\right)} = \frac{\pi_k \phi_k N\left(x^n \mid \mu_k, \sigma_k\right)}{\sum_{k=1}^{K} \pi_k \phi_k N\left(x^n \mid \mu_k, \sigma_k\right)} = \gamma\left(z_{nk}\right)
$$

For the M-step,

$$
\begin{aligned}
&\operatorname{argmax} \sum_{\theta} \sum_{z} \gamma\left(z_{nk}\right) \log\left(z_1^n \mid \theta\right) \\
&= \sum_{n=1}^{N} \sum_{k=1}^{k} \gamma\left(z_{nk}\right) \log\left(z_k, x^n \mid \phi_k, \mu_k, \sigma_k\right) \\
&= \sum_{n=1}^{N} \sum_{k=1}^{k} \gamma\left(z_{nk}\right)\left(\log\pi_k + \log\phi_k - \frac{1}{2}\log(2\pi) - \log\sigma_k - \frac{1}{2}\left(\frac{x^n - \mu_k}{\sigma_k}\right)^2\right)
\end{aligned}
$$

We need to derive the gradient of the parameters.
For π_k, the case is the same as in the lecture slides,

$$\pi_k = \frac{\sum_{n=1}^{N} \gamma\left(z_{nk}\right)}{N}$$

Since $\phi_0 + \phi_1 = 1$,

$$\frac{\partial J}{\partial \phi} = \sum_{n=1}^{N} \gamma\left(z_{n0}\right) \frac{1}{\phi_0} - \gamma\left(z_{n1}\right) \frac{1}{1-\phi_0} = 0$$

Therefore,

$$\phi_0 = \frac{\sum_{n=1}^{N} \gamma\left(z_{n0}\right)}{\sum_{n=1}^{N} \gamma\left(z_{n0}\right) + \gamma\left(z_{n1}\right)}$$

Similarly, we have

$$\frac{\partial J}{\mu_k} = \sum_{n=1}^{N} \gamma\left(z_{nk}\right) \frac{x^n - \mu_k}{\sigma_k} = 0$$

$$\frac{\partial J}{\partial \sigma_k} = \sum_{n=1}^{N} \gamma\left(z_{nk}\right) \left[-\frac{1}{\sigma_k} + \left(\frac{x^n - \mu_k}{\sigma_k}\right)\left(\frac{x^n - \mu_k}{\sigma_k^2}\right)\right] = 0$$

Therefore, we get

$$\mu_k = \frac{\sum_{n=1}^{N} \gamma\left(z_{nk}\right) x^n}{\sum_{n=1}^{N} \gamma\left(z_{nk}\right)} \quad \sigma_k^2 = \frac{\sum_{n=1}^{N} \gamma\left(z_{nk}\right)\left(x^n - \mu_k\right)^2}{\sum_{n=1}^{N} \gamma\left(z_{nk}\right)}$$

d

The addition of λ makes x possible for an arbitrary linear combination of two distributions, so the modified model is a generalized one that can be used to deal with more complicated real-world problems.

5 Independent Component Analysis

a

Refer to `q5.py` and the generated audio files.