CSE 291B. Assignment 1

Out: *Thu Jan 29*
Due: *Thu Feb 12*

1.1 Gaussian integrals

It is important to be able to work easily (and fearlessly) with multivariate Gaussian distributions. The following problems will give you some useful practice.

(a) **Kullback-Leibler divergence**

For continuous distributions $P(x)$ and $Q(x)$, the Kullback-Leibler divergence is defined as:

$$\mathrm{KL}(P,Q) = \int\! dx\; P(x) \log\left[\frac{P(x)}{Q(x)}\right].$$

Compute the Kullback-Leibler divergence between two multivariate Gaussian distributions with means μ_1 and μ_2 and covariance matrices Σ_1 and Σ_2.

(b) **Hellinger distance**

For continuous distributions $P(x)$ and $Q(x)$, the squared Hellinger distance is defined as:

$$H(P,Q) = \frac{1}{2}\int\! dx\, \left(\sqrt{P(x)} - \sqrt{Q(x)}\right)^2.$$

Compute the squared Hellinger distance between two multivariate Gaussian distributions with means μ_1 and μ_2 and covariance matrices Σ_1 and Σ_2.

1.2 Relation between EM algorithm and k-means clustering

Consider a Gaussian mixture model (GMM) with hidden variable $z \in \{1, 2, \ldots, k\}$ and observed variable $x \in \Re^d$. The mixture component distributions of the GMM are given by:

$$P(x|z\!=\!i) \;=\; \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} \exp\left\{-\frac{1}{2}(x-\mu_i)^\top \Sigma_i^{-1}(x-\mu_i)\right\}.$$

Show that if $\Sigma_i = \sigma^2 I$ for all i, where σ^2 is a scalar and I is the identity matrix, then:

$$\lim_{\sigma^2 \to 0} P(z\!=\!i|x) \;=\; \begin{cases} 1 & \text{if } i = \arg\min_j \|x - \mu_j\| \\ 0 & \text{otherwise} \end{cases}$$

1.3 Matrix lemmas

Let $\Psi \in \Re^{D\times D}$ denote a diagonal square matrix and $\Lambda \in \Re^{D\times d}$ a tall rectangular matrix with $d \leq D$. Prove the matrix inverse and matrix determinant lemmas stated in class:

$$\begin{aligned}
(\Psi + \Lambda\Lambda^\top)^{-1} &= \Psi^{-1} - \Psi^{-1}\Lambda(I + \Lambda^\top\Psi^{-1}\Lambda)^{-1}\Lambda^\top\Psi^{-1} \\
\det(\Psi + \Lambda\Lambda^\top) &= \det(\Psi)\ \det(I + \Lambda^\top\Psi^{-1}\Lambda)
\end{aligned}$$

Your proofs may appeal to standard results from linear algebra (e.g., that the determinant of a matrix is equal to the product of its eigenvalues).

1.4 Factor analysis

In factor analysis of zero mean data, the latent and observed variables are assumed to have the multivariate Gaussian distributions:

$$\begin{aligned}
P(z) &= \frac{1}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2}z^\top z\right\}, \\
P(x|z) &= \frac{1}{\sqrt{(2\pi)^d|\Psi|}} \exp\left\{-\frac{1}{2}(x - \Lambda z)^\top\Psi^{-1}(x - \Lambda z)\right\}.
\end{aligned}$$

Starting from the above, derive the form of the marginal distribution $P(x)$. In particular, show that $P(x)$ is a multivariate Gaussian distribution with $\mathrm{E}[x] = 0$ and $\mathrm{E}[xx^\top] = \Psi + \Lambda\Lambda^\top$.

1.5 Matrix factorizations

Download the data set for this problem from the course web site. The data set consists of grayscale images of handwritten digit TWOS. Some of these images are shown below.

The images are stored in MATLAB format as a matrix X with D rows and N columns, where $D = 784$ is the number of pixels per image and $N = 5958$ is the number of images. The i^{th} image in the data set can be displayed using the command:

$$\texttt{imagesc(reshape(X(:,i),28,28));}$$

In this problem, you will compute various low rank factorizations of the matrix $X \approx VY$, where V is a $D \times k$ matrix and Y is a $k \times N$ matrix, with $k = 25$. You will also explore the representations that these factorizations discover. Turn in your source code along with the results requested below.

(a) **Vector quantization**

Minimize the approximation error $\|X - VY\|$ subject to the constraints that $Y_{\alpha n} \in \{0, 1\}$ and $\sum_\alpha Y_{\alpha n} = 1$. Initialize the k columns of V using the first k columns of X. From your final solution, display the columns of V as images, and turn in a print-out of these images.

(b) **Principal component analysis**

Subtract out the mean image from each column of X, and call the resulting matrix $\bar{X}$. Minimize the approximation error $\|\bar{X} - VY\|$ subject to the constraint that the columns of V are orthonormal. From your final solution, display the columns of V as images, and turn in a print-out of these images. Also display the mean image.

(c) **Nonnegative matrix factorization**

Minimize the approximation error $\|X - VY\|$ subject to the constraint that matrices V and Y are nonnegative. Initialize the k columns of V using the first k columns of X, and initialize the matrix Y by setting every element equal to $1/k$. From your final solution, display the columns of V as images, and turn in a print-out of these images.