Imperial College London

COURSEWORK 1

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

495 - Advanced Statistical Machine Learning and Pattern Recognition

Author:
Jiahao Lin (CID: 00837321)

Date: February 6, 2017

Figure 1: This is a figure.

Table 1: Notation

Scalars	x
Vectors	$\boldsymbol{x}$
Matrices	$\boldsymbol{X}$
Transpose	$^\top$
Inverse	$^{-1}$
Real numbers	$\mathbb{R}$
Expected values	$\mathbb{E}$

1 Introduction

This is a template for coursework submission. Many macros and definitions can be found in `notation.tex`. This document is not an introduction to LaTeX. General advice if get stuck: Use your favorite search engine. A great source is also `https://en.wikibooks.org/wiki/LaTeX`.

2 Basics

2.1 Figures

A figure can be included as follows: Fig. 1 shows the Imperial College logo.
Some guidelines:

- Always use vector graphics (scale free)
- In graphs, label the axes
- Make sure the font size (labels, axes) is sufficiently large
- When using colors, avoid red and green together (color blindness)
- Use different line styles (solid, dashed, dotted etc.) and different markers to make it easier to distinguish between lines

2.2 Notation

Table 1 lists some notation with some useful shortcuts (see latex source code).

2.2.1 Equations

Here are a few guidelines regarding equations

- Please use the `align` environment for equations (`eqnarray` is buggy)
- Please number all equations: It will make things easier when we need to refer to equation numbers. If you always use the `align` environment, equations are numbered by default.
- Vectors are by default column vectors, and we write

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \tag{1}$$

- Note that the same macro (`\colvec`) can produce vectors of variable lengths, as

$$\boldsymbol{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \tag{2}$$

- Matrices can be created with the same command. The & switches to the next column:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \end{bmatrix} \tag{3}$$

- Determinants. We provide a simple macro (`\matdet`) whose argument is just a matrix array:

$$\begin{vmatrix} 1 & 2 & 3 \\ 3 & 4 & 5 \\ 2 & 2 & 2 \end{vmatrix} \tag{4}$$

- If you do longer manipulations, please explain what you are doing: Try to avoid sequences of equations without text breaking up. Here is an example: We consider

$$U_1 = [\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}] \subset \mathbb{R}^4, \quad U_2 = [\begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}] \subset \mathbb{R}^4. \tag{5}$$

  To find a basis of $U_1 \cap U_2$, we need to find all $\boldsymbol{x} \in V$ that can be represented as linear combinations of the basis vectors of U_1 and U_2, i.e.,

$$\sum_{i=1}^{3} \lambda_i \boldsymbol{b}_i = \boldsymbol{x} = \sum_{j=1}^{2} \psi_j \boldsymbol{c}_j, \tag{6}$$

where $\boldsymbol{b}_i$ and $\boldsymbol{c}_j$ are the basis vectors of U_1 and U_2, respectively. The matrix $\boldsymbol{A} = [\boldsymbol{b}_1|\boldsymbol{b}_2|\boldsymbol{b}_3|-\boldsymbol{c}_1|-\boldsymbol{c}_2]$ is given as

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & -1 & -1 \\ 0 & 1 & 1 & -2 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}. \tag{7}$$

By using Gaussian elimination, we determine the corresponding reduced row echelon form

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{8}$$

We keep in mind that we are interested in finding $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ and/or $\psi_1, \psi_2 \in \mathbb{R}$ with

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \psi_1 \\ \psi_2 \end{bmatrix} = \mathbf{0}. \tag{9}$$

From here, we can immediately see that $\psi_2 = 0$ and $\psi_1 \in \mathbb{R}$ is a free variable since it corresponds to a non-pivot column, and our solution is

$$U_1 \cap U_2 = \psi_1 \boldsymbol{c}_1 = [\begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix}], \quad \psi_1 \in \mathbb{R}. \tag{10}$$

3 Exercise I

3.a

fdss

3.b

(ii) fsds

4 Exercise II

In Gaussian Mixture Models, assume given a set of N unlablled data, there K gaussian distributed clustered with different centers $\mu_1 ... \mu_k$ and same covariance matrix

Σ. The probability of a data belong in cluster l is $p(k = l) = \pi_l$. So we have our parameters:

$$\Sigma \tag{11}$$
$$\mu_i ... \mu_k \tag{12}$$
$$p(k = 1)...p(k = K) \tag{13}$$

And hidden variables:

$$\boldsymbol{Z}_n = \begin{bmatrix} Z_{n1} \\ Z_{n2} \\ ... \\ Z_{nk} \end{bmatrix} = [\begin{bmatrix} 1 \\ 0 \\ 0 \\ ... \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ ... \\ 0 \end{bmatrix}, \begin{bmatrix} ... \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ ... \\ 1 \end{bmatrix}] \tag{14}$$

where the probability of a sample in clusters is

$$p(\boldsymbol{Z}_n) = \prod_{k=1}^{K} \pi_k^{Z_{nk}} \tag{15}$$

and the probability of a sample x_n given a cluster k is

$$p(x_n | Z_{nk} = 1, \theta) = N(X_n, \mu_k, \Sigma) \tag{16}$$

and the probability of a sample x_n is

$$p(x_n|\theta) = \sum_{k=1}^{K} p(Z_{nk} = 1) p(X_n | Z_{nk} = 1, \theta) = \sum_{k=1}^{K} \pi_k N(X_n | \mu_k, \Sigma) \tag{17}$$

Assume all data samples are independent, we first formulate the joint likelihood

$$p(X, Z|\theta) = p(x_1, x_2, ..., x_n, z_1, z_2, ..., z_n|\theta) \tag{18}$$
$$= \prod_{n=1}^{N} p(x_n|z_n, \theta_x) \prod_{n=1}^{N} p(z_n|\theta_z) \tag{19}$$
$$= \prod_{n=1}^{N} \prod_{k=1}^{K} N(x_n|\mu_k, \Sigma)^{z_{nk}} \prod_{n=1}^{N} \prod_{k=1}^{K} \pi_k^{z_{nk}} \tag{20}$$

where the parameters are

$$\theta_x = \Sigma, \mu_1, \mu_2, ..., \mu_k \tag{21}$$
$$\theta_z = \pi_1, \pi_2, ..., \pi_k \tag{22}$$

In the Expectation Step, we take the log of likelihood

$$\ln p(X, Z|\theta) = \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \{\ln N(X_n|\mu_k, \Sigma) + \ln \pi_k\} \tag{23}$$

And then apply the operator $E_{p(Z|X,\theta)}$ which is expectation of the posterior:

$$E_{p(Z|X,\theta)}[\ln p(X,Z|\theta)] = \sum_{n=1}^{N}\sum_{k=1}^{K} E_{p(Z|X,\theta)}[Z_{nk}]\{\ln N(X_n|\mu_k,\Sigma) + \ln \pi_k\} \tag{24}$$

Now we need to compute $E_{p(Z|X,\theta)}[Z_{nk}]$:

$$E_{p(Z|X,\theta)}[Z_{nk}] = E_{p(Z|X,\theta)}[Z_{nk}] \tag{25}$$

$$= \sum_{z_1} ... \sum_{z_n} p(Z|X,\theta_{old}) = \sum_{z_n} z_{nk} p(z_n|x_n,\theta^{old}) \tag{26}$$

where

$$p(z_n|x_n,\theta^{old}) = \frac{p(x_n,z_n|\theta^{old})}{p(x_n|\theta^{old})} \tag{27}$$

$$= \frac{p(x_n|z_n,\theta^{old})p(z_n|\theta^{old})}{p(x_n|\theta^{old})} \tag{28}$$

$$= \frac{\prod_{k=1}^{K} N(x_n|\mu_k,\Sigma)^{z_{nk}} pi_k{}^{z_{nk}}}{\sum_{z_n}\prod_{k=1}^{K} N(x_n|\mu_k,\Sigma)^{z_{nk}} \pi_k^{z_{nk}}} \tag{29}$$

which gives the expectation

$$E_{p(Z|X,\theta)}[Z_{nk}] = \frac{\sum_{z_n k} z_{nk}\prod_{k=1}^{K} N(x_n|\mu_k,\Sigma)^{z_{nk}} pi_k{}^{z_{nk}}}{\sum_{z_n}\prod_{k=1}^{K} N(x_n|\mu_k,\Sigma)^{z_{nk}} \pi_k^{z_{nk}}} \tag{30}$$

$$= \frac{\pi_k N(x_n|\mu_k,\Sigma)}{\sum_{l=1}^{K} \pi_l N(x_n|\mu_l,\Sigma)} \tag{31}$$

now we see that the expectation is the posterior

$$E_{p(Z|X,\theta)}[Z_{nk}] = p(z_{nk}|x_n) \tag{32}$$

now in the maximization step, we can define

$$G(\theta) = E_{p(Z|X,\theta)}[\ln p(X,Z|\theta)] = \tag{33}$$

$$\sum_{n=1}^{N}\sum_{k=1}^{K} E_{p(Z|X,\theta)}[Z_{nk}]\{\ln N(X_n|\mu_k,\Sigma) + \ln \pi_k\} = \tag{34}$$

$$\sum_{n=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk})\left\{-\frac{1}{2}(x_n-\mu_k)^T\Sigma^{-1}(x_n-\mu_k) - \frac{1}{2}(F\ln 2\pi + \ln|\Sigma|) + \ln \pi_k\right\} \tag{35}$$

take the derivative of our cost function with respect to our parameters, and set them to zero:

$$\frac{dG(\theta)}{d\mu_k} = \sum_{n=1}^{N} \gamma(z_{nk})\Sigma^{-1}(x_n-\mu_k) = 0 \tag{36}$$

$$\mu_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) x_n}{\sum_{n=1}^{N} \gamma(z_{nk})} \tag{37}$$

$$\frac{dG(\theta)}{d\Sigma} = \sum_{n=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) \Big\{ (x_n - \mu_k)^T (x_n - \mu_k) - \Sigma \Big\} = 0 \tag{38}$$

$$\Sigma = \frac{\sum_{n=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk})(x_n - \mu_k)^T (x_n - \mu_k)}{\sum_{n=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk})} \tag{39}$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk})(x_n - \mu_k)^T (x_n - \mu_k) \tag{40}$$

Since we have the constraint that

$$\sum_{k=1}^{K} \pi_k = 1 \tag{41}$$

We can define the laplacian function:

$$L(\theta) = G(\theta) - \lambda(\sum_{k=1}^{K} \pi_k - 1) \tag{42}$$

$$\frac{dL(\theta)}{d\pi_k} = \sum_{n=1}^{N} \frac{\gamma(z_{nk})}{\pi_k} - \lambda = 0 \tag{43}$$

$$\pi_k = \frac{\sum_{n-1}^{N} \gamma(z_{nk})}{\lambda} \tag{44}$$

$$\tag{45}$$

Using the constraint, we can derive the value of λ

$$\sum_{k=1}^{K} \pi_k = 1 = \frac{\sum_{n=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk})}{\lambda} \tag{46}$$

$$\lambda = \sum_{n=1}^{N}\sum_{k=1}^{K} \gamma(z_{nk}) = N \tag{47}$$

$$\pi_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})}{N} \tag{48}$$