Matrix Factorization and Mixture Models

Joachim M. Buhmann, Patrick Pletscher

March 30, 2010

Here we consider a matrix factorization of the form

$$\min_{\mathbf{U},\mathbf{Z}} \|\mathbf{X} - \mathbf{U}\mathbf{Z}\|_2^2 \tag{1}$$

with the constraint $\sum_k z_{k,n} = 1$ and $z_{k,n} \geq 0$. We reformulate it as a problem that consists of two terms: The first term is similar to a Gaussian Mixture Model and the second term is a correction term that is independent of the data $\mathbf{X}$.

$$\begin{aligned}
\|\mathbf{X} - \mathbf{U}\mathbf{Z}\|_2^2 &= \sum_{n=1}^{N} \left\| \mathbf{x}_n - \sum_{k=1}^{K} \mathbf{u}_k z_{k,n} \right\|_2^2 && (2)\\
&= \sum_{n=1}^{N} \left(\|\mathbf{x}_n\|_2^2 - 2 \sum_{k=1}^{K} \mathbf{x}_n^T \mathbf{u}_k z_{k,n} + \sum_{k=1}^{K} \sum_{k'=1}^{K} \mathbf{u}_k^T \mathbf{u}_{k'} z_{k,n} z_{k',n} \right) && (3)\\
&= \sum_{n=1}^{N} \left(\sum_{k=1}^{K} z_{k,n} \|\mathbf{x}_n\|_2^2 - 2 \sum_{k=1}^{K} \mathbf{x}_n^T \mathbf{u}_k z_{k,n} + \sum_{k=1}^{K} z_{k,n} \|\mathbf{u}_k\|_2^2 \right. && (4)\\
&\quad \left. + \sum_{k=1}^{K} \sum_{k'=1}^{K} \mathbf{u}_k^T \mathbf{u}_{k'} z_{k,n} z_{k',n} - \sum_{k=1}^{K} z_{k,n} \|\mathbf{u}_k\|_2^2 \right) && (5)\\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} z_{k,n} \left(\|\mathbf{x}_n\|_2^2 - 2\mathbf{x}_n^T \mathbf{u}_k + \|\mathbf{u}_k\|_2^2 \right) && (6)\\
&\quad + \sum_{n=1}^{N} \left(\left\| \sum_{k=1}^{K} \mathbf{u}_k z_{k,n} \right\|_2^2 - \sum_{k=1}^{K} z_{k,n} \|\mathbf{u}_k\|_2^2 \right) && (7)\\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} z_{k,n} \|\mathbf{x}_n - \mathbf{u}_k\|_2^2 - \underbrace{\sum_{n=1}^{N} \sum_{k=1}^{K} z_{k,n} \left(\|\mathbf{u}_k\|_2^2 - \left\| \sum_{k'=1}^{K} z_{k',n} \mathbf{u}_{k'} \right\|_2^2 \right)}_{=:a_n} && (8)\\
&= \sum_{n=1}^{N} \sum_{k=1}^{K} z_{k,n} \|\mathbf{x}_n - \mathbf{u}_k\|_2^2 - \sum_{n=1}^{N} \sum_{k=1}^{K} z_{k,n} \left\| \mathbf{u}_k - \sum_{k'=1}^{K} z_{k',n} \mathbf{u}_{k'} \right\|_2^2 && (9)
\end{aligned}$$

In the last step from (8) to (9) we make use of the following equality for a given n:

$$
\begin{aligned}
a_n &= \sum_{k=1}^{K} z_{k,n} \left(\|\mathbf{u}_k\|_2^2 - 2 \left\| \sum_{k'=1}^{K} z_{k',n} \mathbf{u}_{k'} \right\|_2^2 + \left\| \sum_{k'=1}^{K} z_{k',n} \mathbf{u}_{k'} \right\|_2^2 \right) \\
&= \sum_{k=1}^{K} z_{k,n} \left\| \mathbf{u}_k - \sum_{k'=1}^{K} z_{k',n} \mathbf{u}_{k'} \right\|^2 .
\end{aligned}
$$

A similar trick that is used to show the relationship $\mathrm{Var}[X] = \mathbb{E}[X - \mathbb{E}[X]]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

We observe that the second term in (9) is always negative. Thus, if we consider the minimization problem

$$
\min_{\mathbf{U},\mathbf{Z}} \sum_{n=1}^{N} \sum_{k=1}^{K} z_{k,n} \|\mathbf{x}_n - \mathbf{u}_k\|_2^2, \tag{10}
$$

we minimize an upper bound of the original problem in (1). The problem in (10) is related to Gaussian Mixture Models as we show below.

In mixture models we assume the $z_{k,n} \in \{0, 1\}$. Consider the full data log-likelihood of the Gaussian Mixture Model, i.e., we assume that both, $\mathbf{X}$ and $\mathbf{Z}$, are given

$$
\ln p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) = \sum_{n=1}^{N} \sum_{k=1}^{K} z_{k,n} \left(\ln \pi_k + \ln \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right) .
$$

As we do not know the latent variables $\mathbf{Z}$, we consider the expected value of the complete data log-likelihood. It can be written as

$$
\mathbb{E}_{\mathbf{Z}}[\ln p(\mathbf{X}, \mathbf{Z} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})] = \sum_{n=1}^{N} \sum_{k=1}^{K} \mathbb{E}[z_{k,n}] \left(\ln \pi_k + \ln \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right) ,
$$

due to the linearity of the expectation. $\mathbb{E}[z_{k,n}]$ is the same as the responsibility $\gamma(z_{k,n})$ in the EM algorithm. We now assume that $\boldsymbol{\Sigma}_k = \mathbf{I}$. Rearranging terms and going from a maximization of the log-likelihood to a minimization of the negative log-likelihood we arrive at

$$
\min_{\mathbf{U},\mathbf{Z}} \frac{1}{2} \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{k,n}) \|\mathbf{x}_n - \boldsymbol{\mu}_k\|_2^2 + c(\mathbf{Z}, \boldsymbol{\pi}). \tag{11}
$$

Here $c(\mathbf{Z}, \boldsymbol{\pi})$ are (not necessarily only negative) terms that only depend on $\mathbf{Z}$ and $\boldsymbol{\pi}$. Remembering that the $z_{k,n}$ in (10) can be interpreted as probabilities we see the similarities between the minimization problems in (11) and (10).