# 6. Mixture models for matrix completion

We can now extend our Gaussian mixture model to predict actual movie ratings. Let $X$
data matrix. The rows of this matrix correspond to users and columns specify movies s
rating value of user $u$ for movie $i$ (if available). Both $n$ and $d$ are typically quite large. T
one to five stars and are mapped to integers $\{1, 2, 3, 4, 5\}$. We will set `X[u,i]=0` whe
missing.

In a realistic setting, most of the entries of $X$ are missing. For this reason, we define $C$
(column indexes) that user $u$ has rated and $H_u$ as its complement (the set of remainin
movies we wish to predict ratings for). We use $|C_u|$ to denote the number of observed
$u$. From the point of view of our mixture model, each user $u$ is an example $x^{(u)}$ =`X[u,`
the coordinates of $x^{(u)}$ are missing, we need to focus the model during training on just
To this end, we use $x^{(u)}_{C_u} = \{x^{(u)}_i : i \in C_u\}$ as the vector of only observed ratings. If
$\{0, \ldots, d-1\}$, then a user $u$ with a rating vector $x^{(u)} = (5, 4, 0, 0, 2)$, where zeros
has $C_u = \{0, 1, 4\}$, $H_u = \{2, 3\}$, and $x^{(u)}_{C_u} = (5, 4, 2)$.

In this part, we will extend our mixture model in two key ways.

- First, we are going to estimate a mixture model based on partially observed ratings. S
- Second, since we will be dealing with a large, high-dimensional data set, we will nee
numerical underflow issues. To this end, you should perform most of your computati
Remember, $\log(a \cdot b) = \log(a) + \log(b)$. This can be useful to remember when $a$
these cases, addition should result in fewer numerical underflow issues than multiplic

  An additional numerical optimization trick that you will find useful is the LogSumExp t
wish to evaluate $y = \log(\exp(x_1) + \ldots \exp(x_n))$. We define $x^* = \max\{x_1 \ldots x_n$
$y = x^* + \log(\exp(x_1 - x^*) + \ldots \exp(x_n - x^*))$. This is just another trick to he
stability.

---

**Marginalizing over unobserved coordinates**

If $x^{(u)}$ were a complete rating vector, the mixture model from Part 1 would simply say t
$P(x^{(u)}|\theta) = \sum_{j=1}^{K} \pi_j N(x^{(u)}; \mu^{(j)}, \sigma_j^2 I)$. In the presence of missing values, we mus
probability $P(x^{(u)}_{C_u}|\theta)$ that is over only the observed values. This marginal corresponds
mixture density $P(x^{(u)}|\theta)$ over all the unobserved coordinate values. In our case, this
computed as follows.

The mixture model for a complete rating vector is written as:

< Previous Next >

## edX

[About](#)
[Affiliates](#)
[edX for Business](#)
[Open edX](#)
[Careers](#)
[News](#)

## Legal

[Terms of Service & Honor Code](#)
[Privacy Policy](#)
[Accessibility Policy](#)
[Trademark Policy](#)
[Sitemap](#)
[Cookie Policy](#)
Do Not Sell My Personal Information

## Connect

[Blog](#)
[Contact Us](#)
[Help Center](#)
[Security](#)
[Media Kit](#)

GET IT ON Google play
Download on the App Store