Audit Access Expires Dec 22, 2020
You lose all access to this course, including your progress, on Dec 22, 2020.

7. Implementing EM for matrix completion

Bookmark this page

Project due Nov 25, 2020 18:59 EST

We need to update our EM algorithm a bit to deal with the fact that the observations are no longer complete vectors. We use Bayes' rule to find an updated expression for the posterior probability $p(j|u) = P(y = j|x_{C_u}^{(u)})$:

$$p(j \mid u) = \frac{p(u|j) \cdot p(j)}{p(u)} = \frac{p(u|j) \cdot p(j)}{\sum_{j=1}^{K} p(u|j) \cdot p(j)} = \frac{\pi_j N(x_{C_u}^{(u)}; \mu_{C_u}^{(j)}, \sigma_j^2 I_{C_u \times C_u})}{\sum_{j=1}^{K} \pi_j N(x_{C_u}^{(u)}; \mu_{C_u}^{(j)}, \sigma_j^2 I_{C_u \times C_u})}$$

This is the soft assignment of cluster j to data point u.

To minimize numerical instability, you will be re-implementing the E-step in the log-domain, so you should calculate the values for the log of the posterior probability, $\ell(j, u) = \log(p(j|u))$ (though the actual output of your E-step should include the non-log posterior).

Let $f(u, i) = \log(\pi_i) + \log\left(N(x_{C_u}^{(u)}; \mu_{C_u}^{(i)}, \sigma_i^2 I_{C_u \times C_u})\right)$. Then, in terms of f, the log posterior is:

$$\begin{aligned}
\ell(j|u) &= \log(p(j \mid u)) = \log\left(\frac{\pi_j N(x_{C_u}^{(u)}; \mu_{C_u}^{(j)}, \sigma_j^2 I_{C_u \times C_u})}{\sum_{j=1}^{K} \pi_j N(x_{C_u}^{(u)}; \mu_{C_u}^{(j)}, \sigma_j^2 I_{C_u \times C_u})}\right) \\
&= \log\left(\pi_j N(x_{C_u}^{(u)}; \mu_{C_u}^{(j)}, \sigma_j^2 I_{C_u \times C_u})\right) - \log\left(\sum_{j=1}^{K} \pi_j N(x_{C_u}^{(u)}; \mu_{C_u}^{(j)}, \sigma_j^2 I_{C_u \times C_u})\right) \\
&= \log(\pi_j) + \log\left(N(x_{C_u}^{(u)}; \mu_{C_u}^{(j)}, \sigma_j^2 I_{C_u \times C_u})\right) - \log\left(\sum_{j=1}^{K} \exp(\log(\pi_j N(x_{C_u}^{(u)}; \mu_{C_u}^{(j)}, \sigma_j^2 I_{C_u \times C_u})))\right) \\
&= f(u, j) - \log\left(\sum_{j=1}^{K} \exp(f(u, j))\right)
\end{aligned}$$

Once we have evaluated $p(j|u)$ in the E-step, we can proceed to the M-step. We wish to find the parameters π, μ, and σ that maximize $\ell(X; \theta)$,

First, note that, by decomposing the multivariate spherical Gaussians into univariate spherical Gaussians as before, we can write, if $l \in C_u$:

$$\begin{aligned}
\frac{\partial}{\partial \mu_l^{(k)}} N(x_{C_u}^{(u)} | \mu_{C_u}^{(k)}, \sigma_k^2 I_{|C_u| \times |C_u|}) &= N(\ldots) \frac{\frac{\partial}{\partial \mu_l^{(k)}}\left(\frac{1}{\sqrt{2\pi}\sigma_{l,(k)}} \exp(-\frac{1}{2\sigma_{l,(k)}^2}(x_l^{(u)} - \mu_l^{(k)})^2)\right)}{\left(\frac{1}{\sqrt{2\pi}\sigma_{l,(k)}} \exp(-\frac{1}{2\sigma_{l,(k)}^2}(x_l^{(u)} - \mu_l^{(k)})^2)\right)} \\
&= N(\ldots) \frac{x_l^{(u)} - \mu_l^{(k)}}{\sigma_{l,(k)}^2}
\end{aligned}$$

where $N(\ldots) = N(x_{C_u}^{(u)} | \mu_{C_u}^{(k)}, \sigma_k^2 I_{|C_u| \times |C_u|})$.

If $l \notin C_u$, that derivative is 0. To cover both cases, we can write:

$$\frac{\partial}{\partial \mu_l^{(k)}} N(x_{C_u}^{(u)} | \mu_{C_u}^{(k)}, \sigma_k^2 I_{|C_u| \times |C_u|}) = N(x_{C_u}^{(u)} | \mu_{C_u}^{(k)}, \sigma_k^2 I_{|C_u| \times |C_u|}) \, \delta(l, C_u) \frac{x_l^{(u)} - \mu_l^{(k)}}{\sigma_{l,(k)}^2}$$

where $\delta(i, C_u)$ is an indicator function: 1 if $i \in C_u$ and zero otherwise.

Following the EM algorithm's approach of maximizing a proxy likelihood function $\hat{\ell}(X; \theta)$ during the M step, consider the following function:

$$\begin{aligned}
\hat{\ell}(X; \theta) = & \sum_{u=1}^{n} \sum_{j=1}^{K} p(j \mid u) \log\left(\frac{p\left(x^{(u)} \text{ generated by cluster } j; \theta\right)}{p(j \mid u)}\right) \\
= & \sum_{u=1}^{n} \sum_{j=1}^{K} p(j \mid u) \log\left(\frac{\pi_j \mathcal{N}(x_{C_u}^{(u)} \mid \mu_{C_u}^{(j)}, \sigma_j^2 I_{|C_u| \times |C_u|})}{p(j \mid u)}\right),
\end{aligned}$$

where $p\left(x^{(u)} \text{ generated by cluster } j; \theta\right)$ is the likelihood of $x^{(u)}$ generated by cluster j and the parameter set is θ. The values $p(j \mid u)$ are the ones as we computed in the E step and they are constants for the M step.

We now take the derivative of $\hat{\ell}(X; \theta)$ with respect to $\mu_l^{(k)}$ to find the optimal value of $\mu_l^{(k)}$ that maximizes $\hat{\ell}(X; \theta)$.

$$\begin{aligned}
\frac{\partial \hat{\ell}(X; \theta)}{\partial \mu_l^{(k)}} &= -\frac{\partial}{\partial \mu_l^{(k)}}\left[\sum_{u=1}^{n} \sum_{j=1}^{K} p(j \mid u) \cdot \frac{1}{2} \cdot \frac{\|x_{C_u}^{(u)} - \mu_{C_u}^{(j)}\|^2}{\sigma_j^2}\right] \\
&= \sum_{u=1}^{n} p(k \mid u) \, \delta(l, C_u) \frac{x_l^{(u)} - \mu_l^{(k)}}{\sigma_k^2},
\end{aligned}$$

where $\delta(i, C_u) = 1$ if $i \in C_u$ and $\delta(i, C_u) = 0$ if $i \notin C_u$.

Setting the partial derivative equal to zero, we obtain that

$$\widehat{\mu_l^{(k)}} = \frac{\sum_{u=1}^{n} p(k \mid u) \, \delta(l, C_u) \, x_l^{(u)}}{\sum_{u=1}^{n} p(k \mid u) \, \delta(l, C_u)}.$$

We leave it as an exercise to the reader to obtain the estimates of σ_k^2 and π_k for $k = 1, \ldots, K$. Verify that

$$\begin{aligned}
\widehat{\sigma_k^2} &= \frac{1}{\sum_{u=1}^{n} |C_u| p(k \mid u)} \sum_{u=1}^{n} p(k \mid u) \|x_{C_u}^{(u)} - \widehat{\mu_{C_u}^{(k)}}\|^2, \\
\widehat{\pi_k} &= \frac{1}{n} \sum_{u=1}^{n} p(k \mid u).
\end{aligned}$$

Implementation guidelines:

- You may find LogSumExp useful. But remember that your M-step should return the new $P = \hat{\pi}$, not the log of $\hat{\pi}$.
- The following will not affect the update equation above, but will affect your implementation: since we are dealing with incomplete data, we might have a case where most of the points in cluster j are missing the i-th coordinate. If we are not careful, the value of this coordinate in the mean will be determined by a small number of points, which leads to erratic results. Instead, we should only update the mean when $\sum_{u=1}^{n} p(j|u) \, \delta(i, C_u) \geq 1$. Since $p(j|u)$ is a soft probability assignment, this corresponds to the case when at least one full point supports the mean.
- To also avoid the variances of clusters going to zero due to a small number of points being assigned to them, in the M-step you will need to implement a minimum variance for your clusters. We recommend a value of 0.25, though you are free to experiment with it if you wish. Note that this issue, as well as the thresholded mean update in the point above, are better dealt with through regularization; however, to keep things simple, we do not do regularization here.
- To debug your EM implementation, you may use the data files `test_incomplete.txt` and `test_complete.txt`. Compare your results to ours from `test_solutions.txt`.

Implementing E-step (2)

0.0/1.0 point (graded)

In `em.py`, fill in the `estep` function so that it works with partially observed vectors where missing values are indicated with zeros, and perform the computations in the log domain to help with numerical stability.

Available Functions: You have access to the NumPy python library as `np`, to the `GaussianMixture` class and to typing annotation `typing.Tuple` as `Tuple`. You also have access to `scipy.special.logsumexp` as `logsumexp`

Hint: For this function, you will want to use `log(mixture.p[j] + 1e-16)` instead of `log(mixture.p[j])` to avoid numerical underflow

```
def estep(X: np.ndarray, mixture: GaussianMixture) -> Tuple[np.ndarray, float]:
    """E-step: Softly assigns each datapoint to a gaussian component

    Args:
        X: (n, d) array holding the data, with incomplete entries (set to 0)
        mixture: the current gaussian mixture

    Returns:
        np.ndarray: (n, K) array holding the soft counts
            for all components for all examples
        float: log-likelihood of the assignment

    """
    raise NotImplementedError

```

Press ESC then TAB or click outside of the code editor to exit

Unanswered

Save Reset

Submit You have used 0 of 25 attempts

Implementing M-step (2)

0.0/1.0 point (graded)

In `em.py`, fill in the `mstep` function so that it works with partially observed vectors where missing values are indicated with zeros, and perform the computations in the log domain to help with numerical stability.

Available Functions: You have access to the NumPy python library as `np`, to the `GaussianMixture` class and to typing annotation `typing.Tuple` as `Tuple`.

```
def mstep(X: np.ndarray, post: np.ndarray, mixture: GaussianMixture,
          min_variance: float = .25) -> GaussianMixture:
    """M-step: Updates the gaussian mixture by maximizing the log-likelihood
    of the weighted dataset

    Args:
        X: (n, d) array holding the data, with incomplete entries (set to 0)
        post: (n, K) array holding the soft counts
            for all components for all examples
        mixture: the current gaussian mixture
        min_variance: the minimum variance for each gaussian

    Returns:
        GaussianMixture: the new gaussian mixture
    """
    raise NotImplementedError
```

Press ESC then TAB or click outside of the code editor to exit

Unanswered

Save Reset

Submit You have used 0 of 25 attempts

Implementing run

0.0/1.0 point (graded)

In `em.py`, fill in the `run` function so that it runs the EM algorithm. As before, the convergence criteria that you should use is that the improvement in the log-likelihood is less than or equal to 10^{-6} multiplied by the absolute value of the new log-likelihood.

Available Functions: You have access to the NumPy python library as `np`, to the `GaussianMixture` class and to typing annotation `typing.Tuple` as `Tuple`. You also have access to the `estep` and `mstep` functions you have just implemented

```
def run(X: np.ndarray, mixture: GaussianMixture,
        post: np.ndarray) -> Tuple[GaussianMixture, np.ndarray, float]:
    """Runs the mixture model

    Args:
        X: (n, d) array holding the data
        post: (n, K) array holding the soft counts
            for all components for all examples

    Returns:
        GaussianMixture: the new gaussian mixture
        np.ndarray: (n, K) array holding the soft counts
            for all components for all examples
        float: log-likelihood of the current assignment
    """
    raise NotImplementedError
```

Press ESC then TAB or click outside of the code editor to exit

Unanswered

Save Reset

Submit You have used 0 of 25 attempts

Discussion

Hide Discussion

Topic: Unit 4 Unsupervised Learning (2 weeks) :Project 4: Collaborative Filtering via Gaussian Mixtures / 7. Implementing EM for matrix completion

Add a Post

Show all posts by recent activity

- Post your code solutions for this project here: — It's been a tough and challenging project and hope everyone was able to figure it out !! I propose that **AFTER** the due date (**Nov 25, 2020, 1... — Pinned, Community TA — 6
- For those stuck @estep & mstep — Many people have faced similar issues and discussed at a different thread : https://courses.edx.org/courses/course-v1:MITx+6.86x+3T2020/disc... — Pinned — 5
- Another approach - ESTEP — Hello guys, This project has been rough for me. So this is coming a bit late. I had to give up on the last tab to go back to Probability. Anyways, he... — Pinned, Community TA — 1
- fill_matrix() implementation — I am stuck at the implementation of the fill_matrix() function. In particular, given an X row X[u,:], to decide what cluster to use in order to fill in th... — 21
- Mean update condition — Hi all, I got the idea of why we apply a condition before we update the mean. If this condition is not satisfied we don't update. When we say we d... — 3
- running test_complete.txt on naive_em — I had a question, are folks able to run the test_complete.txt database with the naive_em code? I am seeing that my mixture.variance values are re... — 1
- @Makshi, just a massive thanks! — You have gone above and beyond with your help and hints. Thanks to you I have managed to code a solution based entirely upon broadcasting wi... — 10
- Implementing M-Step(2); Test: output3 — 12
- Getting the correct log-likelihood but incorrect posterior values — > "though the actual output of your E-step should include the non-log posterior" Clearly I'm missing something about moving the log-posterior v... — 2
- @STAFF - Grader issue with fill_matrix — I seem to be having an issue with the grader for the fill_matrix function on the last tab. I have tried submitting it a few times. The first time I just c... — 9
- How to update mu? — 8
- Exercise 8: overflow issue — With netflix_incomplete.txt file I get in troubles when I calculate the pdf because the term 1/(2*pi*var)**(cu_cnt/2) is equal to 0. That's because t... — 22

© All Rights Reserved

edX
About
Affiliates
edX for Business
Open edX
Careers
News

Legal
Terms of Service & Honor Code
Privacy Policy
Accessibility Policy
Trademark Policy
Sitemap

Connect
Blog
Contact Us
Help Center
Media Kit
Donate

GET IT ON Google play
Download on the App Store

© 2020 edX Inc. All rights reserved.
深圳市恒宇博科技有限公司 粤ICP备17044299号-2