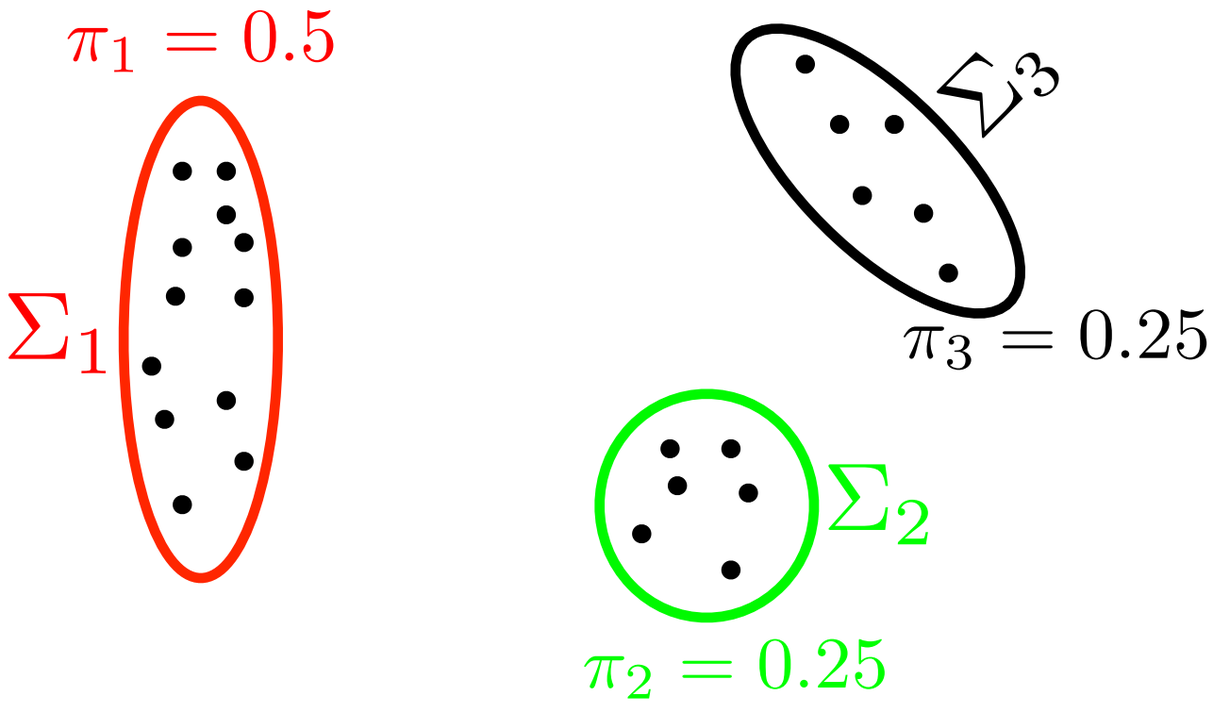
Machine Learning for Data Science (CS4786)
Lecture 17

MLE for GMM

Say we knew model parameters, how do we assign clusters?

Given probability of each point belonging to each of the clusters, how do we compute model parameters?

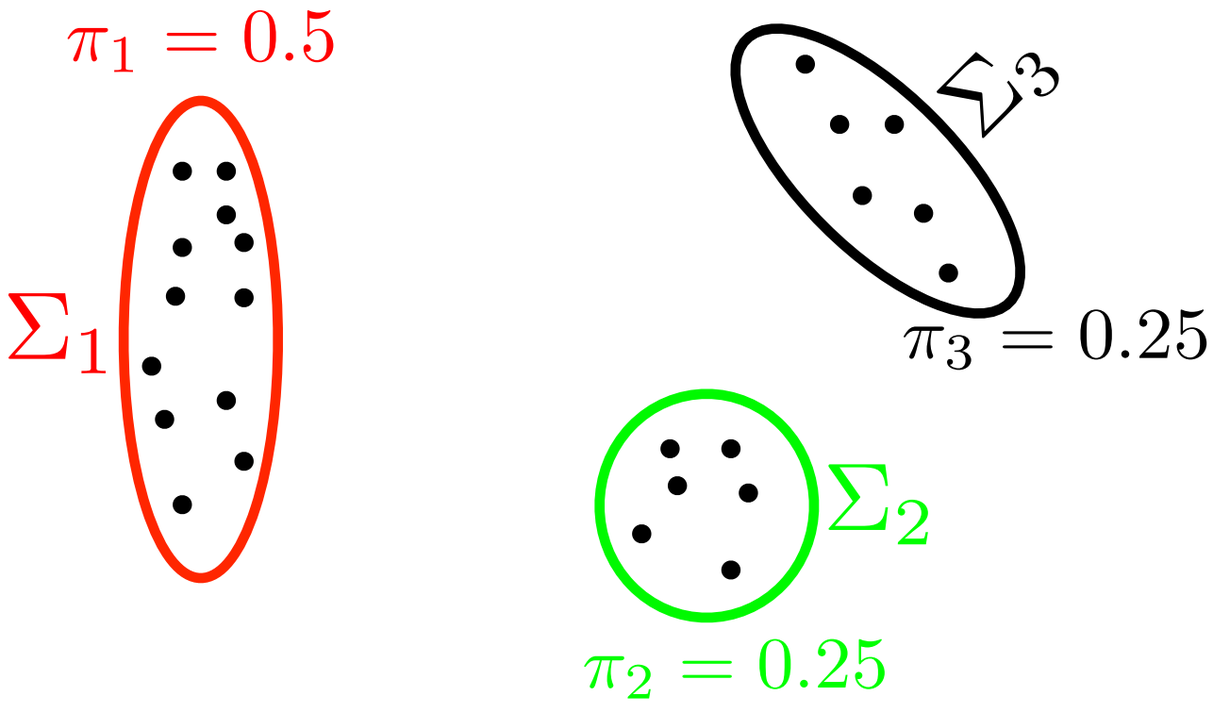


MLE for GMM

Say we knew model parameters, ~~how do we assign clusters?~~

what are the probabilities of points falling in each of the clusters?

Given probability of each point belonging to each of the clusters, how do we compute model parameters?



(Soft) Gaussian Mixture Model

- For all $j \in [K]$, initialize cluster centroids $\hat{\mathbf{r}}_j^0$ and ellipsoids $\hat{\Sigma}_j^0$ randomly and set $m = 1$
- Repeat until convergence (or until patience runs out)
  1. For each $t \in \{1, \ldots, n\}$, set cluster identity of the point

$$Q_t^m(j) = p(\mathbf{x}_t, \hat{\mathbf{r}}_j^{m-1}, \hat{\Sigma}^{m-1}) \times \pi^m(j)$$

  2. For each $j \in [K]$, set new representative as

$$\hat{\mathbf{r}}_j^m = \frac{\sum_{t=1}^n Q_t(j)\mathbf{x}_t}{\sum_{t=1}^n Q_t(j)} \qquad \hat{\Sigma}^m = \frac{\sum_{t=1}^n Q_t(j)(\mathbf{x}_t - \hat{\mathbf{r}}_j^m)(\mathbf{x}_t - \hat{\mathbf{r}}_j^m)^\top}{\sum_{t=1}^n Q_t(j)}$$

$$\pi_j^m = \frac{\sum_{t=1}^n Q_t(j)}{n}$$

  3. $m \leftarrow m + 1$

Expectation Maximization Algorithm

- For demonstration we shall consider the problem of finding MLE (MAP version is very similar)
- Initialize $\theta^{(0)}$ arbitrarily, repeat unit convergence:

(E step) For every t, define distribution Q_t over the latent variable c_t as:

$$Q_t^{(i)}(c_t) = P(c_t|x_t, \theta^{(i-1)})$$

(M step)

$$\theta^{(i)} = \text{argmax}_{\theta \in \Theta} \sum_{t=1}^{n} \sum_{c_t} Q_t^{(i)}(c_t) \log P(x_t, c_t|\theta)$$

Expectation Maximization Algorithm

- For demonstration we shall consider the problem of finding MLE (MAP version is very similar)
- Initialize $\theta^{(0)}$ arbitrarily, repeat unit convergence:

(E step) For every t, define distribution Q_t over the latent variable c_t as:

$$Q_t^{(i)}(c_t) = P(c_t|x_t, \theta^{(i-1)})$$

(M step)

$$\theta^{(i)} = \text{argmax}_{\theta \in \Theta} \sum_{t=1}^{n} \sum_{c_t} Q_t^{(i)}(c_t) \log P(x_t, c_t|\theta)$$

$$= \underset{\theta}{\text{argmax}} \sum_{t=1}^{n} \sum_{k=1}^{K} Q_t^{(i)}(k) \left(\log P(x_t|c_t = k, \theta) + \log P(c_t = k|\theta)\right)$$

Example: EM for GMM

- E step: For every $k \in [K]$,

$$Q_t^{(i)}(c_t = k) = P\left(c_t = k | x_t, \theta^{(i-1)}\right) = P\left(x_t | c_t = k, \theta^{(i-1)}\right) \times P\left(c_t = k | \theta^{(i-1)}\right)$$
$$\propto \underbrace{\phi\left(x_t; \mu_k^{(i-1)}, \Sigma_k^{(i-1)}\right)}_{\text{gaussian p.d.f.}} \times \pi_k^{(i-1)}$$

Example: EM for GMM

- E step: For every $k \in [K]$,

$$Q_t^{(i)}(c_t = k) = P\left(c_t = k | x_t, \theta^{(i-1)}\right) = P\left(x_t | c_t = k, \theta^{(i-1)}\right) \times P\left(c_t = k | \theta^{(i-1)}\right)$$
$$\propto \underbrace{\phi\left(x_t; \mu_k^{(i-1)}, \Sigma_k^{(i-1)}\right)}_{\text{gaussian p.d.f.}} \times \pi_k^{(i-1)}$$

- M step: Given $Q_1, \ldots, Q_n$, we need to find

$$\theta^{(i)} = \operatorname*{argmax}_{\theta \in \Theta} \sum_{t=1}^{n} \sum_{k=1}^{K} Q_t^{(i)}(k) \log P(x_t, c_t = k | \theta)$$
$$= \operatorname*{argmax}_{\theta} \sum_{t=1}^{n} \sum_{k=1}^{K} Q_t^{(i)}(k) \left(\log P(x_t | c_t = k, \theta) + \log P(c_t = k | \theta)\right)$$
$$= \operatorname*{argmax}_{\pi, \mu_{1,\ldots,K}, \Sigma_{1,\ldots,K}} \sum_{t=1}^{n} \sum_{c_t=1}^{K} Q_t^{(i)}(k) \left(\log \phi\left(x_t; \mu_k, \Sigma_k\right) + \log \pi_k\right)$$

Example: EM for GMM

For every $k \in [K]$, the maximization step yields,

$$\mu_k^{(i)} = \frac{\sum_{t=1}^n Q_t^{(i)}(k) x_t}{\sum_{t=1}^n Q_t(k)}, \quad \Sigma_k^{(i)} = \frac{\sum_{t=1}^n Q_t^{(i)}(k) \left(x_t - \mu_k^{(i)}\right)\left(x_t - \mu_k^{(i)}\right)^\top}{\sum_{t=1}^n Q_t(k)}$$

$$\pi_k^{(i)} = \frac{\sum_{t=1}^n Q_t^{(i)}(k)}{n}$$

Let us derive this!

Why should EM work?

A very high level view:

- Performing E-step will never decrease log-likelihood (or log a posteriori)

Why should EM work?

A very high level view:

- Performing E-step will never decrease log-likelihood (or log a posteriori)
- Performing M-step will never decrease log-likelihood (or log a posteriori)

Why should EM work?

A very high level view:

- Performing E-step will never decrease log-likelihood (or log a posteriori)
- Performing M-step will never decrease log-likelihood (or log a posteriori)

Where did M-step even come from?!!

$$\theta^{(i)} = \text{argmax}_{\theta \in \Theta} \sum_{t=1}^{n} \sum_{c_t} Q_t^{(i)}(c_t) \log P(x_t, c_t | \theta)$$

Why should EM work?

Steps to show that $\log \mathrm{Lik}(\theta^{(i)}) \geq \log \mathrm{Lik}(\theta^{(i-1)})$:

$$\log P_{\theta^{(i)}}(x_1, \ldots, x_n)$$

Why should EM work?

Steps to show that $\log \mathrm{Lik}(\theta^{(i)}) \geq \log \mathrm{Lik}(\theta^{(i-1)})$:

$$\log P_{\theta^{(i)}}(x_1, \ldots, x_n) = \sum_{t=1}^{n} \log P_{\theta^{(i)}}(x_t)$$

Why should EM work?

Steps to show that $\log \mathrm{Lik}(\theta^{(i)}) \geq \log \mathrm{Lik}(\theta^{(i-1)})$:

$$\log P_{\theta^{(i)}}(x_1, \ldots, x_n) = \sum_{t=1}^{n} \log P_{\theta^{(i)}}(x_t)$$

$$= \sum_{t=1}^{n} \log \left(\sum_{c_t=1}^{K} P_{\theta^{(i)}}(x_t, c_t) \right)$$

Why should EM work?

Steps to show that $\log \mathrm{Lik}(\theta^{(i)}) \geq \log \mathrm{Lik}(\theta^{(i-1)})$:

$$\begin{aligned}
\log P_{\theta^{(i)}}(x_1, \ldots, x_n) &= \sum_{t=1}^{n} \log P_{\theta^{(i)}}(x_t) \\
&= \sum_{t=1}^{n} \log \left(\sum_{c_t=1}^{K} P_{\theta^{(i)}}(x_t, c_t) \right) \\
&= \sum_{t=1}^{n} \log \left(\sum_{c_t=1}^{K} Q^{(i)}(c_t) \left(\frac{P_{\theta^{(i)}}(x_t, c_t)}{Q^{(i)}(c_t)} \right) \right)
\end{aligned}$$

Why should EM work?

Steps to show that $\log \mathrm{Lik}(\theta^{(i)}) \geq \log \mathrm{Lik}(\theta^{(i-1)})$:

$$\begin{aligned}
\log P_{\theta^{(i)}}(x_1, \ldots, x_n) &= \sum_{t=1}^{n} \log P_{\theta^{(i)}}(x_t) \\
&= \sum_{t=1}^{n} \log \left(\sum_{c_t=1}^{K} P_{\theta^{(i)}}(x_t, c_t) \right) \\
&= \sum_{t=1}^{n} \log \left(\sum_{c_t=1}^{K} Q^{(i)}(c_t) \left(\frac{P_{\theta^{(i)}}(x_t, c_t)}{Q^{(i)}(c_t)} \right) \right) \\
&\geq \sum_{t=1}^{n} \sum_{c_t=1}^{K} Q^{(i)}(c_t) \log \left(\frac{P_{\theta^{(i)}}(x_t, c_t)}{Q^{(i)}(c_t)} \right)
\end{aligned}$$

Why should EM work?

Steps to show that $\log \mathrm{Lik}(\theta^{(i)}) \geq \log \mathrm{Lik}(\theta^{(i-1)})$:

$$\log P_{\theta^{(i)}}(x_1, \ldots, x_n) = \sum_{t=1}^{n} \log P_{\theta^{(i)}}(x_t)$$

$$= \sum_{t=1}^{n} \log \left(\sum_{c_t=1}^{K} P_{\theta^{(i)}}(x_t, c_t) \right)$$

$$= \sum_{t=1}^{n} \log \left(\sum_{c_t=1}^{K} Q^{(i)}(c_t) \left(\frac{P_{\theta^{(i)}}(x_t, c_t)}{Q^{(i)}(c_t)} \right) \right)$$

$$\geq \sum_{t=1}^{n} \sum_{c_t=1}^{K} Q^{(i)}(c_t) \log \left(\frac{P_{\theta^{(i)}}(x_t, c_t)}{Q^{(i)}(c_t)} \right)$$

Log(average) > average of Log

Why should EM work?

Steps to show that $\log \mathrm{Lik}(\theta^{(i)}) \geq \log \mathrm{Lik}(\theta^{(i-1)})$:

$$\log P_{\theta^{(i)}}(x_1, \ldots, x_n) \geq \sum_{t=1}^{n} \sum_{c_t=1}^{K} Q^{(i)}(c_t) \log \left(\frac{P_{\theta^{(i)}}(x_t, c_t)}{Q^{(i)}(c_t)} \right)$$

Why should EM work?

Steps to show that $\log \text{Lik}(\theta^{(i)}) \geq \log \text{Lik}(\theta^{(i-1)})$:

$$\log P_{\theta^{(i)}}(x_1, \ldots, x_n) \geq \sum_{t=1}^{n} \sum_{c_t=1}^{K} Q^{(i)}(c_t) \log \left(\frac{P_{\theta^{(i)}}(x_t, c_t)}{Q^{(i)}(c_t)} \right)$$

$$\geq \sum_{t=1}^{n} \sum_{c_t=1}^{K} Q^{(i)}(c_t) \log \left(\frac{P_{\theta^{(i-1)}}(x_t, c_t)}{Q^{(i)}(c_t)} \right) \qquad \textbf{M-step}$$

Why should EM work?

Steps to show that $\log \mathrm{Lik}(\theta^{(i)}) \geq \log \mathrm{Lik}(\theta^{(i-1)})$:

$$\log P_{\theta^{(i)}}(x_1, \ldots, x_n) \geq \sum_{t=1}^{n} \sum_{c_t=1}^{K} Q^{(i)}(c_t) \log \left(\frac{P_{\theta^{(i)}}(x_t, c_t)}{Q^{(i)}(c_t)} \right)$$

$$\geq \sum_{t=1}^{n} \sum_{c_t=1}^{K} Q^{(i)}(c_t) \log \left(\frac{P_{\theta^{(i-1)}}(x_t, c_t)}{Q^{(i)}(c_t)} \right) \qquad \textbf{M-step}$$

$$= \sum_{t=1}^{n} \sum_{c_t=1}^{K} Q^{(i)}(c_t) \log \left(\frac{P_{\theta^{(i-1)}}(x_t, c_t)}{P_{\theta^{(i-1)}}(c_t | x_t)} \right) \qquad \textbf{E-step}$$

Why should EM work?

Steps to show that $\log \mathrm{Lik}(\theta^{(i)}) \geq \log \mathrm{Lik}(\theta^{(i-1)})$:

$$\log P_{\theta^{(i)}}(x_1, \ldots, x_n) \geq \sum_{t=1}^{n} \sum_{c_t=1}^{K} Q^{(i)}(c_t) \log \left(\frac{P_{\theta^{(i)}}(x_t, c_t)}{Q^{(i)}(c_t)} \right)$$

$$\geq \sum_{t=1}^{n} \sum_{c_t=1}^{K} Q^{(i)}(c_t) \log \left(\frac{P_{\theta^{(i-1)}}(x_t, c_t)}{Q^{(i)}(c_t)} \right) \qquad \textbf{M-step}$$

$$= \sum_{t=1}^{n} \sum_{c_t=1}^{K} Q^{(i)}(c_t) \log \left(\frac{P_{\theta^{(i-1)}}(x_t, c_t)}{P_{\theta^{(i-1)}}(c_t | x_t)} \right) \qquad \textbf{E-step}$$

$$= \sum_{t=1}^{n} \sum_{c_t=1}^{K} Q^{(i)}(c_t) \log P_{\theta^{(i)}}(x_t)$$

Why should EM work?

Steps to show that $\log \mathrm{Lik}(\theta^{(i)}) \geq \log \mathrm{Lik}(\theta^{(i-1)})$:

$$\log P_{\theta^{(i)}}(x_1, \ldots, x_n) \geq \sum_{t=1}^{n} \sum_{c_t=1}^{K} Q^{(i)}(c_t) \log \left(\frac{P_{\theta^{(i)}}(x_t, c_t)}{Q^{(i)}(c_t)} \right)$$

$$\geq \sum_{t=1}^{n} \sum_{c_t=1}^{K} Q^{(i)}(c_t) \log \left(\frac{P_{\theta^{(i-1)}}(x_t, c_t)}{Q^{(i)}(c_t)} \right) \qquad \textbf{M-step}$$

$$= \sum_{t=1}^{n} \sum_{c_t=1}^{K} Q^{(i)}(c_t) \log \left(\frac{P_{\theta^{(i-1)}}(x_t, c_t)}{P_{\theta^{(i-1)}}(c_t | x_t)} \right) \qquad \textbf{E-step}$$

$$= \sum_{t=1}^{n} \sum_{c_t=1}^{K} Q^{(i)}(c_t) \log P_{\theta^{(i)}}(x_t)$$

$$= \sum_{t=1}^{n} \log P_{\theta^{(i)}}(x_t)$$

Why should EM work?

- Likelihood never decreases
- So whenever we converge we converge to a local optima
- However problem is non-convex and can have many local optimal
- In general no guarantee on rate of convergence
- In practice, do multiple random initializations and pick the best one!

EM Algorithm Generally

- More generally, EM can be used to learn any probabilistic model with some Latent (unseen) variables and some observed variables whenever
  - Given all parameters finding distribution for latent variables is easy
  - Its is easy to find parameters given distribution/ observation for all variables

How to choose K (no. of clusters)

- Elbow method:
  - plot Objective versus K, typically it monotonically decreases.
  - Pick point where there is a kink
    - Intuition: look at rate of change
- Add to objective penalty (+ pen(K)) and minimize, pen increases with K
  - intuition we prefer smaller number of clusters
  - Use prior knowledge to pick p
  - (AIC, BIC etc can been seen to be specific cases)
- We can leave the burden of choosing K to the probabilistic model

Mixture of Multinomials

Mixture of Multinomials

10	10	5	2	0	0	0	0	5
1	0	0	1	0	0	0	1	10
0	0	0	0	1	1	0	0	0
20	15	10	5	0	0	0	0	0
10	5	5	2	1	1	1	1	5

Mixture of Multinomials

K buyer types
Each type: distribution over products

10	10	5	2	0	0	0	0	5
1	0	0	1	0	0	0	1	10
0	0	0	0	1	1	0	0	0
20	15	10	5	0	0	0	0	0
10	5	5	2	1	1	1	1	5

Mixture of Multinomials

Mixture of K multinomials

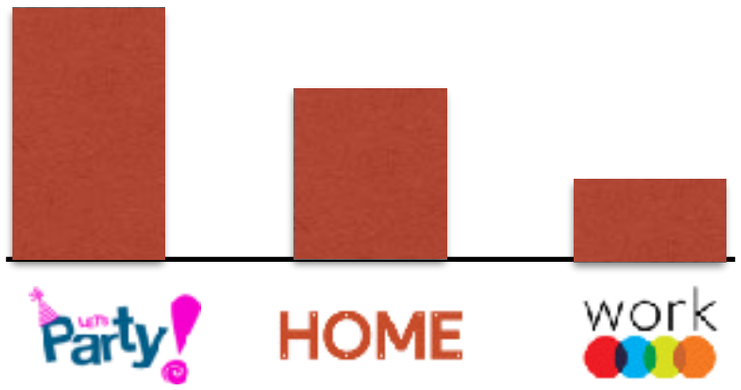
10	10	5	2	0	0	0	0	5
1	0	0	1	0	0	0	1	10
0	0	0	0	1	1	0	0	0
20	15	10	5	0	0	0	0	0
10	5	5	2	1	1	1	1	5

Mixture of Multinomials

10	10	5	2	0	0	0	0	5
1	0	0	1	0	0	0	1	10
0	0	0	0	1	1	0	0	0
20	15	10	5	0	0	0	0	0
10	5	5	2	1	1	1	1	5

Mixture of Multinomials

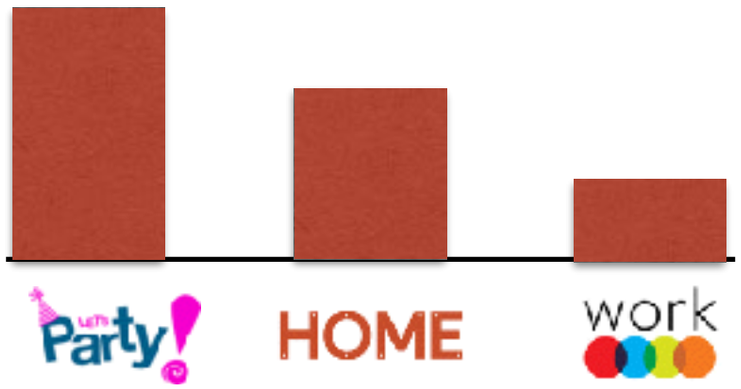
$\pi =$



10	10	5	2	0	0	0	0	5
1	0	0	1	0	0	0	1	10
0	0	0	0	1	1	0	0	0
20	15	10	5	0	0	0	0	0
10	5	5	2	1	1	1	1	5

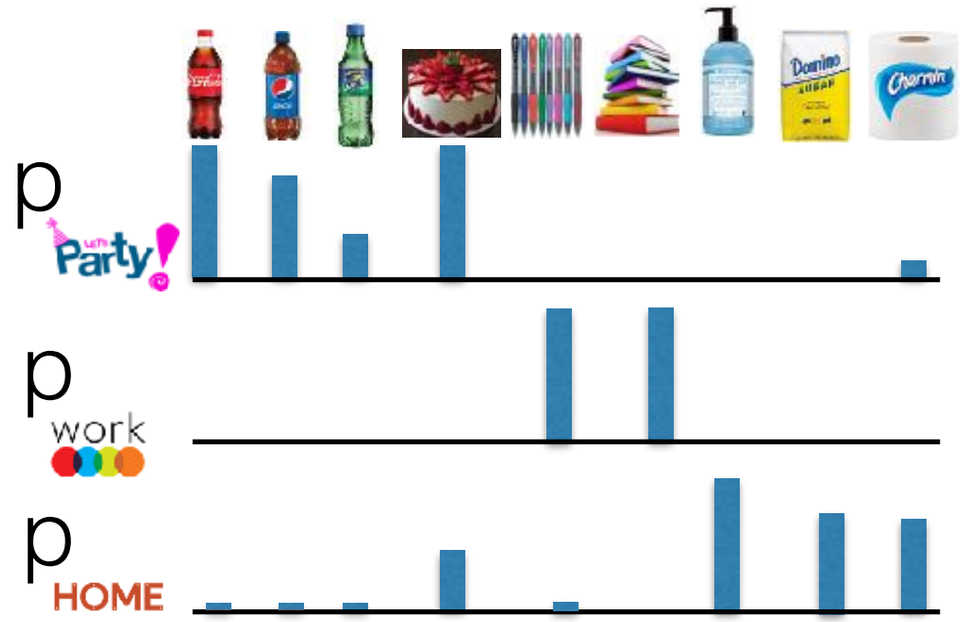
Mixture of Multinomials

$\pi =$



10	10	5	2	0	0	0	0	5
1	0	0	1	0	0	0	1	10
0	0	0	0	1	1	0	0	0
20	15	10	5	0	0	0	0	0
10	5	5	2	1	1	1	1	5

p

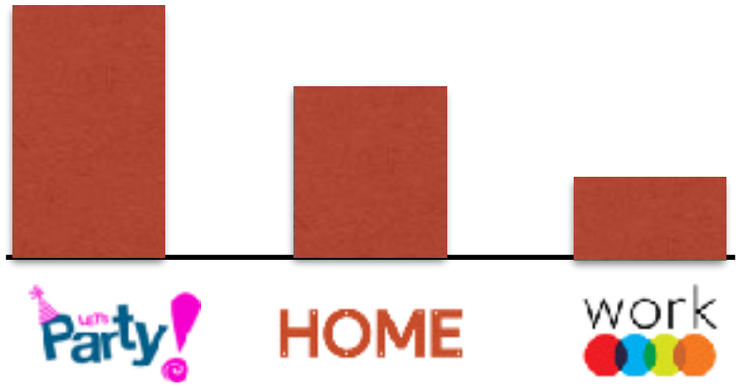


p

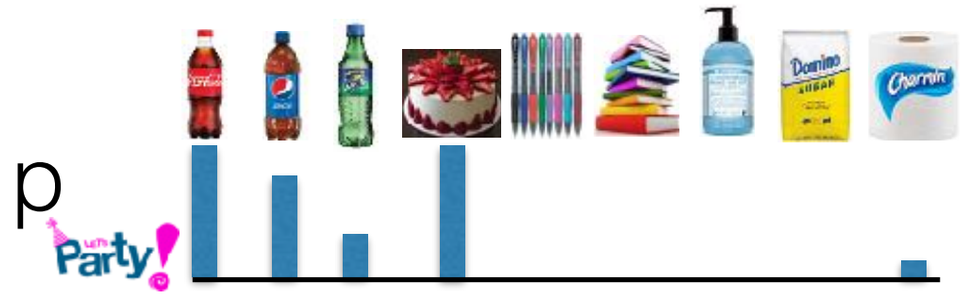
p

Mixture of Multinomials

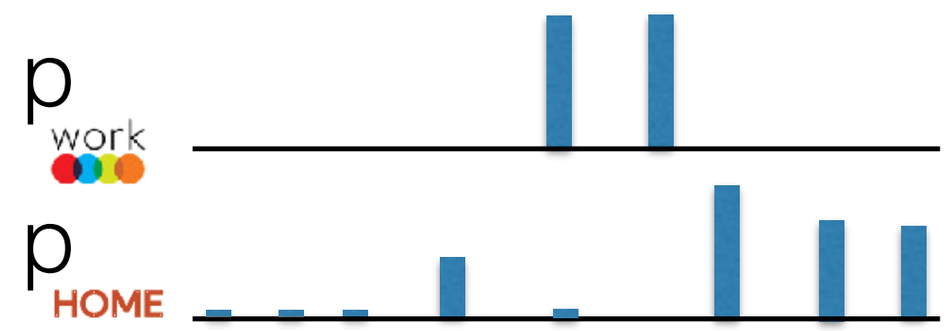
$\pi =$



p



p



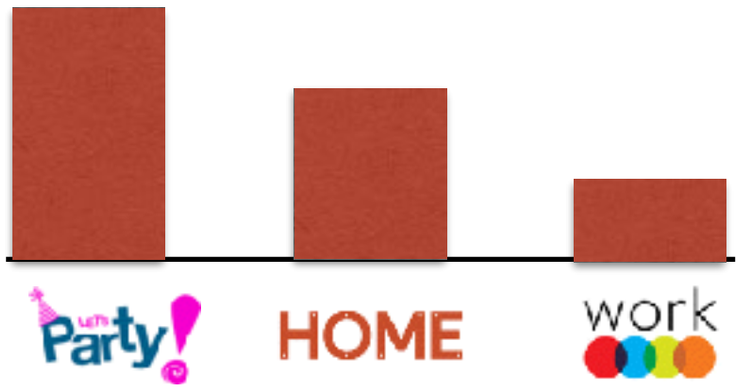
p

$\pi \sim$

Party!	10	10	5	2	0	0	0	0	5
HOME	1	0	0	1	0	0	0	1	10
work	0	0	0	0	1	1	0	0	0
Party!	20	15	10	5	0	0	0	0	0
work	10	5	5	2	1	1	1	1	5

Mixture of Multinomials

$\pi =$

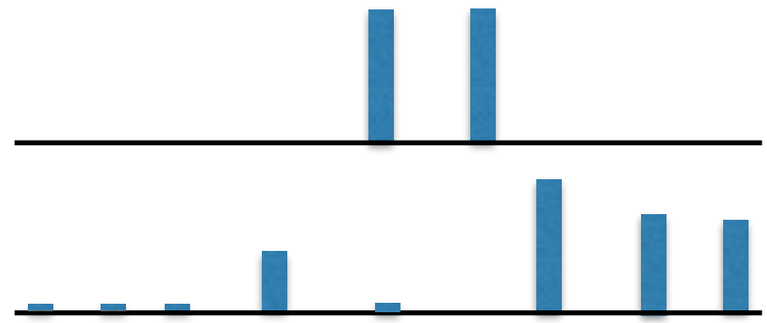


π
~

Party!	~	10	10	5	2	0	0	0	0	5
HOME	~	1	0	0	1	0	0	0	1	10
work	~	0	0	0	0	1	1	0	0	0
Party!	~	20	15	10	5	0	0	0	0	0
work	~	10	5	5	2	1	1	1	1	5

p Party!

p work



p HOME

Mixture of Multinomials

- Eg. Model purchases of each customer
- K-types of customers, each designated with distribution over the d items to buy
- Generative model:
  - π is mixture distribution over the K-types of buyers
  - $p_1, \ldots, p_K$ are the K distributions over the d items, one for each customer type
  - Generative process, each round draw customer type $c_t \sim \pi$
  - Next given c_t draw list of purchases as $x_t \sim \text{multinomial}(p_{c_t})$

Multinomial Distribution

$$P(x|p) = \frac{m!}{x[1]! \cdot \ldots \cdot x[d]!} p[1]^{x_t[1]} \cdot \ldots \cdot p[d]^{x_t[d]}$$

Probability of purchase vector x while drawing products independently m times from p

E-step

$$Q_t^{(i)}(c_t) \propto P(x_t|c_t, \theta^{(i-1)})P(c_t|\theta^{(i-1)})$$

E-step

$$Q_t^{(i)}(c_t) \propto P(x_t|c_t, \theta^{(i-1)})P(c_t|\theta^{(i-1)})$$

$$= \frac{P(x_t|p_{c_t}^{(i-1)})\pi^{(i-1)}(c_t)}{\sum_{k=1}^{K} P(x_t|p_k^{(i-1)})\pi^{(i-1)}(k)}$$

E-step

$$Q_t^{(i)}(c_t) \propto P(x_t|c_t, \theta^{(i-1)})P(c_t|\theta^{(i-1)})$$

$$= \frac{P(x_t|p_{c_t}^{(i-1)})\pi^{(i-1)}(c_t)}{\sum_{k=1}^{K} P(x_t|p_k^{(i-1)})\pi^{(i-1)}(k)}$$

$$= \frac{p_{c_t}[1]^{x_t[1]} \cdot \ldots \cdot p_{c_t}[d]^{x_t[d]} \cdot \pi_{c_t}^{(i-1)}}{\sum_{k=1}^{K} p_k[1]^{x_t[1]} \cdot \ldots \cdot p_{c_t}[d]^{x_t[d]} \cdot \pi_k^{(i-1)}}$$

M-step

$$\theta^{(i)} = \text{argmax}_\theta \sum_{t=1}^{n} \sum_{k=1}^{K} Q_t^{(i)}(k) \log \left(P(x_t | c_t = k, \theta) P(c_t = k | \theta) \right)$$

M-step

$$\theta^{(i)} = \text{argmax}_\theta \sum_{t=1}^{n}\sum_{k=1}^{K} Q_t^{(i)}(k) \log\left(P(x_t|c_t = k, \theta)P(c_t = k|\theta)\right)$$

$$= \text{argmax}_{\pi, p_1, \ldots, p_K} \left\{ \sum_{t=1}^{n}\sum_{k=1}^{K} Q_t^{(i)}(k) \log\left(\frac{m!}{x_t[1]! \cdot \ldots \cdot x_t[d]!} p_k[1]^{x_t[1]} \cdot \ldots \cdot p_k[d]^{x_t[d]}\right)\right.$$

$$\left. + \sum_{t=1}^{n}\sum_{k=1}^{K} Q_t^{(i)}(k) \log \pi_k \right\}$$

M-step

$$\theta^{(i)} = \text{argmax}_\theta \sum_{t=1}^{n}\sum_{k=1}^{K} Q_t^{(i)}(k) \log\left(P(x_t|c_t = k, \theta)P(c_t = k|\theta)\right)$$

$$= \text{argmax}_{\pi,p_1,\ldots,p_K} \left\{ \sum_{t=1}^{n}\sum_{k=1}^{K} Q_t^{(i)}(k) \log\left(\frac{m!}{x_t[1]! \cdot \ldots \cdot x_t[d]!} p_k[1]^{x_t[1]} \cdot \ldots \cdot p_k[d]^{x_t[d]}\right) \right.$$

$$\left. + \sum_{t=1}^{n}\sum_{k=1}^{K} Q_t^{(i)}(k) \log \pi_k \right\}$$

$$= \text{argmax}_{\pi,p_1,\ldots,p_K} \left\{ \sum_{t=1}^{n}\sum_{k=1}^{K} Q_t^{(i)}(k) \log\left(p_k[1]^{x_t[1]} \cdot \ldots \cdot p_k[d]^{x_t[d]}\right) \right.$$

$$\left. + \sum_{t=1}^{n}\sum_{k=1}^{K} Q_t^{(i)}(k) \log \pi_k \right\}$$

M-step

$$\theta^{(i)} = \text{argmax}_\theta \sum_{t=1}^{n}\sum_{k=1}^{K} Q_t^{(i)}(k) \log\left(P(x_t|c_t = k, \theta)P(c_t = k|\theta)\right)$$

$$= \text{argmax}_{\pi,p_1,\ldots,p_K} \left\{ \sum_{t=1}^{n}\sum_{k=1}^{K} Q_t^{(i)}(k) \log\left(\frac{m!}{x_t[1]! \cdot \ldots \cdot x_t[d]!} p_k[1]^{x_t[1]} \cdot \ldots \cdot p_k[d]^{x_t[d]}\right) \right.$$

$$\left. + \sum_{t=1}^{n}\sum_{k=1}^{K} Q_t^{(i)}(k) \log \pi_k \right\}$$

$$= \text{argmax}_{\pi,p_1,\ldots,p_K} \left\{ \sum_{t=1}^{n}\sum_{k=1}^{K} Q_t^{(i)}(k) \log\left(p_k[1]^{x_t[1]} \cdot \ldots \cdot p_k[d]^{x_t[d]}\right) \right.$$

$$\left. + \sum_{t=1}^{n}\sum_{k=1}^{K} Q_t^{(i)}(k) \log \pi_k \right\}$$

$$= \text{argmax}_{\pi,p_1,\ldots,p_K} \left\{ \sum_{t=1}^{n}\sum_{k=1}^{K}\sum_{j=1}^{d} Q_t^{(i)}(k) x_t[j] \log\left(p_k[j]\right) + \sum_{t=1}^{n}\sum_{k=1}^{K} Q_t^{(i)}(k) \log \pi_k \right\}$$

M-step

$$\pi_k^{(i)} = \frac{\sum_{t=1}^{n} Q_t^{(i)}(k)}{n}$$

$$p_k[j] = \frac{\sum_{t=1}^{n} x_t[j] Q_t^{(i)}(k)}{m \sum_{t=1}^{n} Q_t^{(i)}(k)}$$

M-step

$$\pi_k^{(i)} = \frac{\sum_{t=1}^{n} Q_t^{(i)}(k)}{n}$$

proportion of weights for each type

$$p_k[j] = \frac{\sum_{t=1}^{n} x_t[j] Q_t^{(i)}(k)}{m \sum_{t=1}^{n} Q_t^{(i)}(k)}$$

weighted average of jth product