Applied Machine Learning

EM for Mixtures of Multinomial Distributions - Topic Models

EM for Mixtures of Multinomial Distributions
Model Topics

- Modeling clusters through Normal Distributions
- EM algorithm for mixture of normals: E-Step
- EM algorithm for mixture of normals: M-Step

EM Algorithm

1. Initialize probability distributions
2. While ($\theta^{(n)}$ has not reached convergence)
    1. E-step
        - $p(\delta \,|\, \theta^{(n)}, \mathbf{x})$
        - $$\begin{aligned} \mathcal{Q}(\theta; \theta^{(n)}) &= \sum_\delta \mathscr{L}(\theta; \mathbf{x}, \delta) p(\delta \,|\, \theta^{(n)}, \mathbf{x}) \\ &= \mathbb{E}_{p(\delta|\theta^{(n)}, \mathbf{x})}[\mathscr{L}(\theta; \mathbf{x}, \delta)] \end{aligned}$$
        - $w_{i,j}$ to associate each item $\mathbf{x}_i$ to cluster center j
    2. M-step
        - $\theta^{(n+1)} = \underset{\theta}{\operatorname{argmax}}\, Q(\theta; \theta^{(n)})$

Topics with Multinomial Distributions

- One document => One topic
- t possible topics
- Word counts conditioned on the topic
  - d possible words
  - Similarity measured through counts
  - Independent and Identically Distributed (IID) samples of multinomial distribution
- Documents result from
  - selection of topic j with probability π_j
  - draw words as IID samples for topic j
  - d-dimensional vector stores counts per word

Probability Model for Mixture of Multinomials

- documents: t different topics, d different words
- data item i: $\mathbf{x}_i$
  - $x_{i,k}$: count of word k in item i
- word probabilities for topic j: $\mathbf{p}_j$
  - $p_{j,k}$: probability of word k for topic j
- Mixture of t multinomials (topic models)
  - $p(\mathbf{x}_i | \mathbf{p}_j) = \frac{(\mathbf{x}_i^\top \mathbf{1})!}{\prod_v x_{i,v}!} \prod_u p_{j,u}^{x_{i,u}}$
- Parameters $\theta = (\mathbf{p}_1, \ldots, \mathbf{p}_t, \pi_1, \ldots, \pi_j)$
- $$\begin{aligned} p(\mathbf{x}_i, \theta) &= \textstyle\sum_l p(\mathbf{x}_i | \text{topic} = j) p(\text{topic} = j | \theta) \\ &= \textstyle\sum_j \left[\frac{(\mathbf{x}_i^\top \mathbf{1})!}{\prod_v x_{i,v}!} \prod_u p_{j,u}^{x_{i,u}} \right] \pi_j \end{aligned}$$

- $\delta_{i,j} = \begin{cases} 1 & \text{if } \mathbf{x}_i \text{ comes from topic } j \\ 0 & \text{otherwise} \end{cases}$
  - $p(\delta_{i,j} = 1 | \theta) = \pi_j$
  - $p(\delta_i | \theta) = \prod_j [\pi_j]^{\delta_{i,j}}$
- $p(\mathbf{x}, \delta_i | \theta) = \prod_j \left[\pi_j \frac{(\mathbf{x}_i^\top \mathbf{1})!}{\prod_v x_{i,v}!} \prod_u (p_{j,u})^{x_{i,u}} \right]^{\delta_{i,j}}$
- $\mathscr{L}(\theta; \mathbf{x}, \delta) = \sum_{i,j} \left[log \pi_j + \sum_u x_{i,u} log p_{j,u} \right] \delta_{i,j} + K$

EM for Topic Models

1. Initialize probability distributions
2. While ($\theta^{(n)}$ has not reached convergence)
   1. E-step
      - $p(\delta \mid \theta^{(n)}, \mathbf{x})$
      - $$\begin{aligned} \mathcal{Q}(\theta; \theta^{(n)}) &= \sum_\delta \mathscr{L}(\theta; \mathbf{x}, \delta) p(\delta \mid \theta^{(n)}, \mathbf{x}) \\ &= \mathbb{E}_{p(\delta \mid \theta^{(n)}, \mathbf{x})}[\mathscr{L}(\theta; \mathbf{x}, \delta)] \end{aligned}$$
      - $w_{i,j}$ to associate each item $\mathbf{x}_i$ to cluster center j
   2. M-step
      - $\theta^{(n+1)} = \underset{\theta}{\operatorname{argmax}}\, Q(\theta; \theta^{(n)})$

- E-Step: find weights $\mathcal{Q}(\theta; \theta^{(n)})$ from data items and $\theta^{(n)}$

  $$\mathcal{Q}(\theta; \theta^{(n)}) = \sum_\delta \mathscr{L}(\theta; \mathbf{x}, \delta) p(\delta \mid \theta^{(n)}, \mathbf{x})$$

  - $$= \sum_{i,j} \left(\left[log\pi_j + \sum_u x_{i,u} log p_{j,u} \right] w_{i,j} \right) + K$$
  - where
  - $$\begin{aligned} w_{i,j} &= p(\delta_{i,j} = 1 \mid \theta^{(n)}, \mathbf{x}) \\ &= \frac{p(\mathbf{x}, \delta_{i,j} = 1 \mid \theta^{(n)})}{\sum_l p(\mathbf{x}, \delta_{i,l} = 1 \mid \theta^{(n)})} \\ &= \frac{[\prod_k (\mathbf{p}_{j,k})^{x_{i,k}}]\pi_j}{\sum_l [\prod_k (\mathbf{p}_{j,k})^{x_{i,k}}]\pi_k} \end{aligned}$$
  - M-Step: parameters θ that maximize $\mathcal{Q}(\theta; \theta^{(n)})$
  - $$\begin{aligned} \mathbf{p}_j &= \frac{\sum_i \mathbf{x}_i w_{i,j}}{\sum_i (\mathbf{x}_i^\top \mathbf{1}) w_{i,j}} \\ \pi_j &= \frac{\sum_i w_{i,j}}{N} \end{aligned}$$

EM for Topic Models

1. Initialize probability distributions
2. While ($\theta^{(n)}$ has not reached convergence)
   1. E-step
      - weights to associate each item $\mathbf{x}_i$ to cluster centers j
      - $$w_{i,j}^{(n)} = \frac{[\prod_k (\mathbf{p}_{j,k})^{x_{i,k}}]\pi_j^{(n)}}{\sum_l [\prod_k (\mathbf{p}_{j,k})^{x_{i,k}}]\pi_k^{(n)}}$$
   2. M-step
      - parameters $\theta^{(n+1)}$)
      - $$\mathbf{p}_j^{(n+1)} = \frac{\sum_i \mathbf{x}_i w_{i,j}^{(n)}}{\sum_i (\mathbf{x}_i^\top \mathbf{1}) w_{i,j}^{(n)}}$$
        $$\pi_j^{(n+1)} = \frac{\sum_i w_{i,j}^{(n)}}{N}$$

- E-Step: find weights $\mathcal{Q}(\theta; \theta^{(n)})$ from data items and $\theta^{(n)}$

  $$\mathcal{Q}(\theta; \theta^{(n)}) = \sum_\delta \mathscr{L}(\theta; \mathbf{x}, \delta) p(\delta \mid \theta^{(n)}, \mathbf{x})$$
  - $$= \sum_{i,j} \left(\left[log \pi_j + \sum_u x_{i,u} log p_{j,u} \right] w_{i,j} \right) + K$$
  - where
  - $$w_{i,j} = p(\delta_{i,j} = 1 \mid \theta^{(n)}, \mathbf{x})$$
    $$= \frac{p(\mathbf{x}, \delta_{i,j} = 1 \mid \theta^{(n)})}{\sum_l p(\mathbf{x}, \delta_{i,l} = 1 \mid \theta^{(n)})}$$
    $$= \frac{[\prod_k (\mathbf{p}_{j,k})^{x_{i,k}}]\pi_j}{\sum_l [\prod_k (\mathbf{p}_{j,k})^{x_{i,k}}]\pi_k}$$
- M-Step: parameters θ that maximize $\mathcal{Q}(\theta; \theta^{(n)})$
  - $$\mathbf{p}_j = \frac{\sum_i \mathbf{x}_i w_{i,j}}{\sum_i (\mathbf{x}_i^\top \mathbf{1}) w_{i,j}}$$
    $$\pi_j = \frac{\sum_i w_{i,j}}{N}$$

EM for Mixtures of Normal Distributions

- Modeling clusters through Normal Distributions
- EM algorithm for mixture of normals: E-Step
- EM algorithm for mixture of normals: M-Step

Applied Machine Learning

EM for Mixtures of Multinomial Distributions - Topic Models