RPZ - EM algoritmus

Petr Svec

Pozn. V textu budu kombinovat cesky vety s anglickymi pojmy, jelikoz u nekterych terminu vazne nevim cesky ekvivalenty.
Prerekvizita :) - znalost MLE

K cemu to je ve zkratce: K odhadu skrytych(latentnich) promennych u mixture models (smesi nekolika rozdeleni) a k nim nalezicim parametrum jednotlivych komponent smesi.

Formulace problemu:
Mejme smes rozdeleni

$$S = Mix_{\boldsymbol{w}}(U_1, ..., U_k) = p(\mathbf{x}; \boldsymbol{w}, \theta) = \sum_{k=1}^{K} w_k p_k(\mathbf{x}; \theta_k), \tag{1}$$

kde $\boldsymbol{w}$ je vektor vah(apriornich psti) jednotlivych komponent $\rightarrow \sum w_i = 1, w_i \in \langle 0; 1 \rangle$, θ je matice parametru $\rightarrow \theta_k$ je vektor parametru k-te komponenty smesi S. Komponenty oznacme $c_1, ..., c_K$.
Necht $\mathbf{x_1}, ..., \mathbf{x_n}$ jsou m-rozmerna mereni, ktere jsou z komponent $c_1, ..., c_K$ smesi S. Skutecnost, ze lib. vzorek patri do tridy c_i neni znama. To formalne popiseme pomoci $K \times n$ (latentnich) promennych $z_{ik}; z_{ik} \in \{0; 1\}$, kde $z_{ik} = 1$ pokud $\mathbf{x_i}$ bylo generovano tridou k, s podminkou, ze

$$\forall i \in \{1, .., n\} : \sum_{k=1}^{K} z_{ik} = 1 \tag{2}$$

tj. $\mathbf{x_i}$ nalezi prave do jedne tridy smesi.
Cilem je nalezt θ takove, ktere maximalizuje verohodnost.

Hrube odvozeni EM:
Nejprve je treba uvest Jensenovu nerovnost, ktera bude treba dale v textu(bez dukazu):

Veta 1 (Jensenova nerovnost)**.** *Necht f je konvexni funkce definovana na intervalu I. Pokud* $x_1, ..., x_n \in I$ *a* $\lambda_1, ..., \lambda_n \geq 0, pricemz \sum_{i=1}^{n} \lambda_i = 1,$ *pak*

$$f(\sum_{i=1}^{n} \lambda_i x_i) \leq \sum_{i=1}^{n} \lambda_i f(x_i).$$

U konkavnich fci je nerovnost opacna.

S pouzitim pouze toho co vime(jsme schopni pozorovat) by log-likelihood vypadala nasledovne:

$$\ell(\boldsymbol{w}, \theta; \mathbf{x}) = \sum_{i=1}^{n} \log \sum_{k=1}^{K} w_k p_k(\mathbf{x_i}; \theta_k) \tag{3}$$

Je tu ale mensi problem s tim, ze nezname provazanost promennych pri maximalizaci sumy v logaritmu tj. tohle je pouze pro lidi kteri uz nechteji zit.

Pokud bychom znali z_{ik}, pak ML je:

$$\ell(\boldsymbol{w}, \theta; \mathbf{x}, \mathbf{z}) = \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik} (\log p_k(\mathbf{x_i}; \theta_k) + \log w_k) \tag{4}$$

Misto toho ale muzeme pouzit odhad $Q_{ik} = P[z_{ik} = 1]$ tj. pst ze $\mathbf{x_i}$ patri do c_k.

$$E[\ell(\boldsymbol{w}, \theta; \mathbf{x}, \mathbf{z}), Q] = \sum_{i=1}^{n} \sum_{k=1}^{K} Q_{ik} (\log p_k(\mathbf{x_i}; \theta_k) + \log w_k) \tag{5}$$

Ted prijde magie s pouzitim Jensenovy nerovnosti(aplikovana na logaritmus) a dobra myslenka v jednom:

$$\begin{aligned}
\ell(\boldsymbol{w}, \theta; \mathbf{x}) &= \sum_{i=1}^{n} \log \sum_{k=1}^{K} w_k p_k(\mathbf{x_i}; \theta_k) = \\
&= \sum_{i=1}^{n} \log \sum_{k=1}^{K} Q_{ik} \frac{w_k p_k(\mathbf{x_i}; \theta_k)}{Q_{ik}} = \\
&\geq \sum_{i=1}^{n} \sum_{k=1}^{K} Q_{ik} \log \frac{w_k p_k(\mathbf{x_i}; \theta_k)}{Q_{ik}} = \\
&= \sum_{i=1}^{n} \sum_{k=1}^{K} Q_{ik} \log w_k p_k(\mathbf{x_i}; \theta_k) - \sum_{i=1}^{n} \sum_{k=1}^{K} Q_{ik} \log Q_{ik} = \\
&= E[\ell(\boldsymbol{w}, \theta; \mathbf{x}, \mathbf{z}), Q] - \sum_{i=1}^{n} \sum_{k=1}^{K} Q_{ik} \log Q_{ik} = \\
&= L(w, \theta, Q; \mathbf{x})
\end{aligned} \tag{6}$$

Ona dvojita suma na 5. radku nas nemusi trapit, jelikoz neni zavisla na parametrech w, θ. Maximalizace $L(w, \theta, Q; \mathbf{x})$ je tedy stejna jako maximalizace rovnice #4.

Shrnuti: pointou tedy je maximalizace dolni hranice log-likelihood (resp. hledani Q, ktery ji maximalizuje) nebo jinak receno optimalizace dolni hranice takova, ze rozdil mezi dolni hranici a skutecnou verohodnosti v danem bode je minimalni(zaroven ale vetsi nebo roven nule).
Algoritmus je tvoren nasledujicimi dvema kroky:

1. Expectation step - $Q^{(t+1)} = \arg\max_Q L(w^{(t)}, \theta, Q; \mathbf{x})$
2. Maximization step - $(w^{(t+1)}, \theta^{(t+1)}) = \arg\max_{w,\theta} L(w, \theta, Q^{(t+1)}; \mathbf{x})$

Bezi v nekonecne smycce, ktera je prerusena pokud $(L^{(t)} - L^{(t+1)}) < \epsilon$, kde ϵ je pevne zvolena konstanta.
1. krok - Expectation ma obecne reseni, kterym je aposteriorni pst. $Q^*_{ik} \equiv P(z_{ik} = 1 | \mathbf{x_i}; w, \theta)$, ktera maximalixuje $L(w, \theta, Q; \mathbf{x})$ pri danem w, θ. Dukaz je triv, udelejte si za domaci ukol :D

2. krok odpovida temer klasicke MLE s tim rozdilem, ze je treba zahrnou i skryte promenne

resp. jejich odhady do vypoctu odhadu parametru. Navic, coz je pomerne dulezity rozdil, nehledame parametry, ktere optimalizuji puvodni likelihood, ale dolni hranici.
Priklad: pro odhad str. hodnoty jedne komponenty smesi, ktera se ridi normalnim rozdelenim bude odhad vypadat nasledovne:

$$\hat{\mu}_k = \frac{\sum_{i=1}^{n} Q_{ik}\mathbf{x_i}}{\sum_{i=1}^{n} Q_{ik}} \tag{7}$$