Esempio EM: apprendimento non supervisionato delle medie di due Gaussiane

L'insieme delle variabili in considerazione è $Y = X \cup Z$, dove X è l'insieme delle variabili osservabili e Z l'insieme delle variabili non osservabili. In particolare, l'i-esima "variabile" è data da

$$y_i = \langle x_i, z_{i1}, z_{i2} \rangle$$

dove $x_i \in \mathbb{R}$ è generato o dalla Gaussiana 1 o dalla Gaussiana 2 (entrambe con varianza conosciuta σ e medie μ_1 e μ_2 sconosciute), e $z_{i1}, z_{i2} \in \{0, 1\}$.

Vediamo come definire in questo caso

$$Q(h'|h) = E[\ln p(Y|h')|h, X]$$

In particolare, la probabilità a posteriori di y_i data l'ipotesi h' è data da

$$p(y_i|h') = p(x_i, z_{i1}, z_{i2}|h') = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(z_{i1}(x_i-\mu_1')^2 + z_{i2}(x_i-\mu_2')^2)}$$

dove $h' = \langle \mu_1', \mu_2' \rangle$. Infatti, se x_i è generato dalla Gaussiana 1, allora $z_{i,1} = 1 \ \wedge \ z_{i,2} = 0$, e quindi

$$p(y_i|h') = p(x_i, 1, 0|h') = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu_1')^2},$$

se x_i è generato dalla Gaussiana 2, allora $z_{i,1} = 0 \ \wedge \ z_{i,2} = 1$, e quindi

$$p(y_i|h') = p(x_i, 0, 1|h') = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu_2')^2}$$

Ricordando che il logaritmo (ln) è una funzione monotona crescente e che le istanze y_i sono estratte indipendentemente, la probabilità a posteriori di Y data l'ipotesi h' è data da

$$\begin{aligned}
\ln p(Y|h') &= ln \prod_{i=1}^{m} p(y_i|h') \\
&= \sum_{i=1}^{m} \ln p(y_i|h') \\
&= \sum_{i=1}^{m} (\ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2}(z_{i1}(x_i - \mu_1')^2 + z_{i2}(x_i - \mu_2')^2))
\end{aligned}$$

e il suo valore aspettato rispetto a X e l'ipotesi corrente h è

$$\begin{aligned}
E[\ln p(Y|h')] &= E\left[\sum_{i=1}^{m} (\ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2}(z_{i1}(x_i - \mu_1')^2 + z_{i2}(x_i - \mu_2')^2))\right] \\
&= \sum_{i=1}^{m} (\ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2}(E[z_{i1}](x_i - \mu_1')^2 + E[z_{i2}](x_i - \mu_2')^2))
\end{aligned}$$

Quindi il passo E dell'algoritmo EM definisce

$$Q(h'|h) = \sum_{i=1}^{m} (\ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2}(E[z_{i1}](x_i - \mu_1')^2 + E[z_{i2}](x_i - \mu_2')^2))$$

dove $E[z_{i1}]$ e $E[z_{i2}]$ sono calcolate sulla base di $h = \langle \mu_1, \mu_2 \rangle$ e X come

$$\begin{aligned}
E[z_{i1}] &= \frac{e^{-\frac{1}{2\sigma^2}(x_i - \mu_1)^2}}{e^{-\frac{1}{2\sigma^2}(x_i - \mu_1)^2} + e^{-\frac{1}{2\sigma^2}(x_i - \mu_2)^2}} \\
E[z_{i2}] &= \frac{e^{-\frac{1}{2\sigma^2}(x_i - \mu_2)^2}}{e^{-\frac{1}{2\sigma^2}(x_i - \mu_1)^2} + e^{-\frac{1}{2\sigma^2}(x_i - \mu_2)^2}}
\end{aligned}$$

Il passo M deve risolvere il problema

$$\arg\max_{h'} Q(h'|h) = \arg\max_{h'} \sum_{i=1}^{m} (\ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{1}{2\sigma^2}(E[z_{i1}](x_i - \mu_1')^2 + E[z_{i2}](x_i - \mu_2')^2))$$

$$
\begin{aligned}
&= \arg\max_{h'} \frac{1}{2\sigma^2} \sum_{i=1}^{m} (-(E[z_{i1}](x_i - \mu_1')^2 + E[z_{i2}](x_i - \mu_2')^2)) \\
&= \arg\min_{h'} \sum_{i=1}^{m} (E[z_{i1}](x_i - \mu_1')^2 + E[z_{i2}](x_i - \mu_2')^2)
\end{aligned}
$$

che si risolve ponendo le derivate parziali rispetto a μ_1' e μ_2' a zero:

$$
\begin{aligned}
\frac{\partial \sum_{i=1}^{m} (E[z_{i1}](x_i - \mu_1')^2 + E[z_{i2}](x_i - \mu_2')^2)}{\partial \mu_1'} &= 0 \\
\\
\frac{\partial \sum_{i=1}^{m} (E[z_{i1}](x_i - \mu_1')^2 + E[z_{i2}](x_i - \mu_2')^2)}{\partial \mu_2'} &= 0
\end{aligned}
$$

o, equivalentemente

$$
\begin{aligned}
\sum_{i=1}^{m} E[z_{i1}](x_i - \mu_1') = 0 \quad &\Rightarrow \quad \sum_{i=1}^{m} E[z_{i1}]x_i = \mu_1' \sum_{i=1}^{m} E[z_{i1}] \\
\sum_{i=1}^{m} E[z_{i1}](x_i - \mu_2') = 0 \quad &\Rightarrow \quad \sum_{i=1}^{m} E[z_{i2}]x_i = \mu_2' \sum_{i=1}^{m} E[z_{i2}]
\end{aligned}
$$

da cui

$$
\begin{aligned}
\mu_1' &= \frac{\sum_{i=1}^{m} E[z_{i1}]x_i}{\sum_{i=1}^{m} E[z_{i1}]} \\
\mu_2' &= \frac{\sum_{i=1}^{m} E[z_{i2}]x_i}{\sum_{i=1}^{m} E[z_{i2}]}
\end{aligned}
$$