

演習 13.32 について考える。

本当はパターン認識と機械学習の学習 (暗黒通信団) にある、ベクトル、行列の微分をまとめるつもりだったが、できていない。

まず、P156 の一般の EM アルゴリズムを考えると、何かしらの方法で $p(\mathbf{Z}|\mathbf{X}, \theta^{old})$ (もしくはそれに等価なもの) を求め、 Q を最大化する新しい、 θ を求める手順だった。

Q については (9.33) を参考に連続値に展開して、

$$\begin{aligned} Q(\theta, \theta^{old}) &= \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \theta^{old}) \ln p(\mathbf{Z}, \mathbf{X}|\theta) = \int p(\mathbf{Z}|\mathbf{X}, \theta^{old}) \ln p(\mathbf{Z}, \mathbf{X}|\theta) d\mathbf{Z} \\ &= \int \cdots \int p(\mathbf{z}_1, \cdots, \mathbf{z}_N|\mathbf{X}, \theta^{old}) \ln p(\mathbf{z}_1, \cdots, \mathbf{z}_N, \mathbf{X}|\theta) d\mathbf{z}_1 \cdots d\mathbf{z}_N (= \mathbb{E}(\ln p(\mathbf{Z}, \mathbf{X}|\theta))) \end{aligned} \quad (1)$$

ただし、 $d\mathbf{z}_1, \cdots, d\mathbf{z}_N$ はここでは自由に入れ替えられるとする。(極論すると、和なので足す順番を入れ替えても良い。また、この辺の話はガウス分布を仮定しているので、連続関数。)

(13.77) で置き換えた (13.108) を代入して、 \mathbf{P}_0, μ_0 に依存しない項を定数とすると、

$$Q(\theta, \theta^{old}) = \int \cdots \int p(\mathbf{z}_1, \cdots, \mathbf{z}_N|\mathbf{X}, \theta^{old}) \left(-\frac{1}{2} \ln |\mathbf{P}_0^{new}| - \frac{1}{2} (\mathbf{z}_1 - \mu_0^{new})^T \mathbf{P}_0^{new} (\mathbf{z}_1 - \mu_0^{new}) \right) d\mathbf{z}_1 \cdots d\mathbf{z}_N \quad (2)$$

ここでは、まだ、積分を実施していない。(期待値をとっていない)。

先に、 μ_0 で微分し (この辺が、パターン認識と機械学習の学習の 2.4-2.4.3 付近に記載がある。)、これが 0 になる場合を調べる。

$$\begin{aligned} \frac{\partial Q(\theta, \theta^{old})}{\partial \mu_0} &= \int \cdots \int p(\mathbf{z}_1, \cdots, \mathbf{z}_N|\mathbf{X}, \theta^{old}) (\mathbf{P}_0 (\mathbf{z}_1 - \mu_0)) d\mathbf{z}_1 \cdots d\mathbf{z}_N \\ &= \mathbf{P}_0 \int \cdots \int p(\mathbf{z}_1, \cdots, \mathbf{z}_N|\mathbf{X}, \theta^{old}) \mathbf{z}_1 d\mathbf{z}_1, \cdots, d\mathbf{z}_N - \mathbf{P}_0 \int \cdots \int p(\mathbf{z}_1, \cdots, \mathbf{z}_N|\mathbf{X}, \theta^{old}) \mu_0 d\mathbf{z}_1 \cdots d\mathbf{z}_N \\ &= \mathbf{P}_0 \int p(\mathbf{z}_1|\mathbf{X}, \theta^{old}) \mathbf{z}_1 d\mathbf{z}_1 - \mathbf{P}_0 \mu_0 = \mathbf{P}_0^{new} \int \gamma(\mathbf{z}_n)^{old} \mathbf{z}_1 d\mathbf{z}_1 - \mathbf{P}_0^{new} \mu_0^{new} = 0 \end{aligned} \quad (3)$$

\mathbf{P}_0^{new} はガウス分布の精度行列なので、正則とする。また、 $\gamma(\mathbf{z}_n)^{old}$ は正規分布なので、積分部分は $\hat{\mu}_0^{old}$ になる。よって、

$$\hat{\mu}_0^{old} - \mu_0^{new} = 0 \quad (4)$$

つまり、

$$\mathbb{E}(\mathbf{z}_0) = \hat{\mu}_0^{old} = \mu_0^{new} \quad (5)$$

次に、(2) を \mathbf{P}_0 について、微分し、0 になる場合を考える。

$\frac{\partial \ln |\mathbf{P}_0|}{\partial \mathbf{P}_0}$ に関しては (C.28) もしくは、パターン認識と機械学習の学習の 2.4.6 に、また、 $\frac{\partial (\mathbf{z}_1 - \mu_0^{new})^T \mathbf{P}_0^{new} (\mathbf{z}_1 - \mu_0^{new})}{\partial \mathbf{P}_0}$ に関してはパターン認識と機械学習の学習の 2.5 記載がある。その結果、

$$\frac{\partial Q(\theta, \theta^{old})}{\partial \mathbf{P}_0^{new}} = \int \cdots \int p(\mathbf{z}_1, \cdots, \mathbf{z}_N|\mathbf{X}, \theta^{old}) \left(-(\mathbf{P}_0^{new-1})^T + (\mathbf{P}_0^{new-1} (\mathbf{z}_1 - \mu_0^{new}) (\mathbf{z}_1 - \mu_0^{new})^T \mathbf{P}_0^{new-1})^T \right) d\mathbf{z}_1 \cdots d\mathbf{z}_N = 0 \quad (6)$$

\mathbf{P}_0 は対称行列で、その逆行列も対称行列となる。また、 $(\mathbf{z}_1 - \mu_0^{new}) (\mathbf{z}_1 - \mu_0^{new})^T$ も対称行列。よって、

$$\mathbf{P}_0^{new-1} = \mathbf{P}_0^{new-1} \int p(\mathbf{z}_1|\mathbf{X}, \theta^{old}) (\mathbf{z}_1 - \mu_0^{new}) (\mathbf{z}_1 - \mu_0^{new})^T d\mathbf{z}_1 \mathbf{P}_0^{new-1} \quad (7)$$

積分部分は

$$\begin{aligned} \int p(\mathbf{z}_1|\mathbf{X}, \theta^{old})(\mathbf{z}_1 - \mu_0^{new})(\mathbf{z}_1 - \mu_0^{new})d\mathbf{z}_1 &= \int p(\mathbf{z}_1|\mathbf{X}, \theta^{old})\mathbf{z}_1\mathbf{z}_1^T d\mathbf{z}_1 - 2 \int p(\mathbf{z}_1|\mathbf{X}, \theta^{old})\mathbf{z}_1 d\mathbf{z}_1 \mu_0^{newT} + \mu_0^{new} \mu_0^{newT} \\ &= \mathbb{E}(\mathbf{z}_1\mathbf{z}_1^T) - 2\mathbb{E}(\mathbf{z}_1)\mu_0^{newT} + \mu_0^{new} \mu_0^{newT} = \mathbb{E}(\mathbf{z}_1\mathbf{z}_1^T) - 2\mathbb{E}(\mathbf{z}_1)\mathbb{E}(\mathbf{z}_1)^T + \mathbb{E}(\mathbf{z}_1)\mathbb{E}(\mathbf{z}_1)^T = \mathbb{E}(\mathbf{z}_1\mathbf{z}_1^T) - \mathbb{E}(\mathbf{z}_1)\mathbb{E}(\mathbf{z}_1)^T \end{aligned} \quad (8)$$

これを (7) に代入して、両側から、 \mathbf{P}_0^{new} をかけると、

$$\mathbf{P}_0^{new} = \mathbb{E}(\mathbf{z}_1\mathbf{z}_1^T) - \mathbb{E}(\mathbf{z}_1)\mathbb{E}(\mathbf{z}_1)^T \quad (9)$$

よく見ると $\mathbf{P}_0^{new} = \hat{\mathbf{V}}_1^{old}$ になりそう。

なお、(13.104) の式を求める際に、

$$\mathbb{E}((\mathbf{z}_1 - \hat{\mu}_1)(\mathbf{z}_1 - \hat{\mu}_1)^T) = \hat{\mathbf{V}}_1 \quad (10)$$

が求まっている。(13.106) は (13.104) の転置、(1.41) より、(13.107) は上記と、(1.40) より求まる。