演習 5.25

ノートのタイトル　2014/10/05

$$E = E_0 + \frac{1}{2}(w - w^*)^T H (w - w^*)$$

matrix cookbook p.9

$$\nabla E = \frac{1}{2}\left(w^T H w - w^T H w^* - w^{*T} H w + w^{*T} H w^*\right) \qquad \frac{\partial x^T B x}{\partial x} = (B + B^T)x$$

$$= \frac{1}{2}\left\{(H + H^T)w - Hw^* - (w^{*T}H)^T\right\} \qquad \nabla(x^T B x) = (B + B^T)x$$

(C.19)

$$\frac{\partial}{\partial x}(x^T a) = \frac{\partial}{\partial x}(a^T x) = a$$

$\leftarrow H = H^T$ なので

$$= Hw - Hw^*$$

$$= H(w - w^*)$$

Hが正定値の定義 ≡ Hが対称かつ $x^T H x > 0$

注 Hが対称でないのは定義であり $x^T H x > 0$ の必要条件ではない

(5.196)に入れて

$$w^{(\tau)} = w^{(\tau-1)} - \rho H (w - w^*)$$

両辺に u_j をかけて

$$w^{(\tau)T} u_j = w^{(\tau-1)T} u_j - \{\rho H (w^{(\tau-1)} - w^*)\}^T u_j$$

$$w_j^{(\tau)} = w_j^{(\tau-1)} - \rho (w^{(\tau-1)} - w^*)^T H u_j$$

$$w_j^{(\tau)} = w_j^{(\tau-1)} - \rho (w^{(\tau-1)} - w^*)^T \eta_j u_j$$

$$w_j^{(\tau)} = w_j^{(\tau-1)} - \rho\eta_j (w^{(\tau-1)} - w^*)^T u_j$$

$$= w_j^{(\tau-1)} - \rho\eta_j (w_j^{(\tau-1)} - w_j^*)$$

$$= (1 - \rho\eta_j) w_j^{(\tau-1)} + \rho\eta_j w_j^*$$

$\tau = 1, 2, 3$ について　問題より

$$w_j^{(1)} = \rho\eta_j w_j^* \qquad \leftarrow w^{(0)} = 0 \text{ として}$$

$$w_j^{(2)} = (1 - \rho\eta_j) w_j^{(1)} + \rho\eta_j w_j^*$$

$$= (1 - \rho\eta_j)\rho\eta_j w_j^* + \rho\eta_j w_j^*$$

$$= \{(1 - \rho\eta_j) + 1\}\rho\eta_j w_j^*$$

$$w_j^{(3)} = (1 - \rho\eta_j) w_j^{(2)} + \rho\eta_j w_j^*$$

$$= (1 - \rho\eta_j)\{(1 - \rho\eta_j) + 1\}\rho\eta_j w_j^* + \rho\eta_j w_j^*$$

$$= [(1 - \rho\eta_j)\{(1 - \rho\eta_j) + 1\} + 1]\rho\eta_j w_j^*$$

$$= \{(1 - \rho\eta_j)^2 + (1 - \rho\eta_j) + 1\}\rho\eta_j w_j^*$$

$$\therefore\ w_j^{(\tau)} = \sum_{k=0}^{\tau-1}(1-\rho\eta_j)^k \rho\eta_j w_j^*$$

← 等比級数の和 $S_n = \sum_{k=0}^{n-1} ar^k = \frac{a(1-r^n)}{1-r}$

$$= \frac{\{1-(1-\rho\eta_j)^{\tau-1}\}}{1-(1-\rho\eta_j)}\rho\eta_j w_j^*$$

$$= \{1-(1-\rho\eta_j)^{\tau}\} w_j^* \quad \cdots (5.197)$$

これより

$|1-\rho\eta_j|<1$ ならば $\tau \to \infty$ で $w_j^{(\tau)} = w_j^*$ である。

問題には明記されてないが、この条件が要る

$\eta_j \gg (\rho\tau)^{-1}$ かつ $|1-\rho\eta_j|<1$ かつ $\rho>0$ のとき $w_j^{(\tau)} \simeq w_j^*$ を示す

$\cdots$① $\cdots$② $\cdots$③

ρは学習率で正の値で小さい値

(例) $\eta_j = 1,\ \rho = 0.1,\ \tau = 100$ とすると

$(\rho\tau)^{-1} = 0.1,\ |1-\rho\eta_j| = 0.9$ となり ①, ② を満たす

これを (5.197) に入れると

$$w_j^{(100)} = (1-0.9^{100})w_j^* \simeq w_j^*$$

1より小さい数のべきになるので ~0 である

(証)

① より $\tau \gg (\rho\eta_j)^{-1}$

② より $|\rho\eta_j|<1$

また ① より $\eta_j \gg (\rho\tau)^{-1} > 0$ なので $\rho\eta_j > 0$

よって $0 < \rho\eta_j < 1$

$\therefore\ (\rho\eta_j) > 1$

よって $\tau \gg (\rho\eta_j)^{-1} > 1$

(5.197)

$$w_j^{(\tau)} = \{1-(1-\rho\eta_j)^{\tau}\} w_j^*$$

$|1-\rho\eta_j| < 1,\ \tau \gg 1$ なので $(1-\rho\eta_j)^{\tau} \to 0$

よって

$$w_j^{(\tau)} \simeq w_j^*$$

を得る。

次に $\eta_j \ll (\rho\tau)^{-1}$ かつ $|1-\rho\eta_j|<1$ かつ $\rho>0$ のとき $|w_j^{(\tau)}| \ll |w_j^*|$

(例) $\rho=0.1,\ \tau=10,\ \eta_j=0.1$ のとき

(5.197)　　　この巾乗よりもここの値がうんと小さいという条件になっている $\eta_j \ll (\rho\tau)^{-1}$

$$w_j^{(\tau)} = \{1-(1-\underline{0.01})^{(10)}\} w_j^* = \underbrace{0.0956}_{\text{小さい}} w_j^*$$

なので $|w_j^{(\tau)}| \ll |w_j^*|$ といえる。

(証) $\{1-(1-\rho\eta_j)^\tau\} \sim 0$ であることを示す

$\downarrow$

$(1-\rho\eta_j)^\tau \sim 1$ であることを示す

$\downarrow$

τは反復回数なのである程度大きい値100とかになる。

$\downarrow$

$\rho\eta_j$ がうんと小さいときならOK

$\eta_j \ll (\rho\tau)^{-1}$ より $\rho\eta_j \ll \frac{1}{\tau}$

$|1-\rho\eta_j|<1$、$\rho>0$ より $\eta_j>0$、また $\tau>1$ として良いので

$$\rho\eta_j = \frac{a}{\tau},\quad 0<a<1$$

と書ける。

$\rho\eta_j \ll \frac{1}{\tau}$ なので $a \sim 0$ である

なので

$$(1-\rho\eta_j)^\tau = \left(1-\frac{a}{\tau}\right)^\tau \simeq 1$$

を得る。

3.5.3の図3.15の結論

$$\begin{cases} \lambda_i \ll \alpha \text{ のとき } w_i \sim 0 \\ \lambda_i \gg \alpha \text{ のとき } w_i \sim w_{MLi} \end{cases}$$

ここでの結論

$$\begin{cases} \eta_j \ll (\rho\tau)^{-1}\alpha \text{ のとき } |w_j^{(\tau)}| \ll |w_j^*| \\ \eta_j \gg (\rho\tau)^{-1}\alpha \text{ のとき } w_j^{(\tau)} \simeq w_j^* \end{cases}$$

ここで

ゴサ関数

w^* は (5.195) $E = E_0 + \frac{1}{2}(w-w^*)^T H (w-w^*)$ を最小にする値

なので図3.15の w_{MAP} に相当する

なので

$|w_j^{(\tau)}| \ll |w_j^*| \iff w_j \sim 0$ ← 充分小さいというのも0に近いというのも同じこと

また

$w_j^{(\tau)} \simeq w_j^* \iff w_j^{(\tau)} \simeq w_{MAP}$ ← w_{ML}に近付くのもw_{MAP}に近付くのも同じこと

である。

次に λ_i と η_j を比較する

3.5.3での誤差関数は

(3.80) $E(w) = E(m_N) + \frac{1}{2}(w - m_N) A (w - m_N)$

ここで $A = \alpha I + \beta \Phi^T \Phi$

$\beta \Phi^T \Phi u_i = \lambda_i u_i$

なので、ヘッセ行列Aの固有値は $\alpha + \lambda_i$ である。

この章での誤差関数は

(5.195) $E = E_0 + \frac{1}{2}(w - w^*)^T H (w - w^*)$

で、η_j は Hの固有値である

なので、$\lambda_i \ll \alpha$ と $\eta_i \ll (\rho\tau)^{-1}$ は若干似ているが相当するとは言えないと思う