PRML輪講3章計算ノート

担当　大木俊幸

平成 29 年 6 月 19 日

このノートについて

輪講において筆者が担当した (日本語版)PRML の 3 章で、スライドには載せていない計算の詳しい内容についてまとめたものです。

正規方程式 (3.15) の導出

二乗和誤差 (あるいは対数尤度関数) の極値を求めるため、(3.12)

$$E_D(\mathbf{w}) = \frac{1}{2}(\mathbf{t}^{\mathrm{T}}\mathbf{t} - \mathbf{w}^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{w} + \mathbf{w}^{\mathrm{T}}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\mathbf{w})$$

を微分して [1]

$$\begin{aligned}\frac{\partial}{\partial \mathbf{w}}E_D(\mathbf{w}) &= \frac{1}{2}(-\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{t} - (\mathbf{t}^{\mathrm{T}}\boldsymbol{\Phi})^{\mathrm{T}} + \mathbf{2}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\mathbf{w}) \\ &= -\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{t} + \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\mathbf{w}\end{aligned}$$

$\mathbf{w} = \mathbf{w}_{ML}$ でこれは 0 だから

$$\mathbf{w}_{ML} = (\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi})^{\mathbf{-1}}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{t} \tag{3.15}$$

演習 3.5

工事中。

事後分布 (3.49) について (演習 3.7)

モデルパラメータ $\mathbf{w}$ の共役事前分布を設定するときに (3.48) のガウス分布を用いる理由→線形回帰における尤度関数が (3.10) の $\mathbf{w}$ の二次で与えられること、ベイズの定理より (事後確率)$\propto$(尤度)$\times$(事前確率) ということから共役な事前分布は同じく $\mathbf{w}$ の二次形式になるガウス分布とするのが妥当と考えられます。

[1] $\mathbf{a}$ を定ベクトル、A を $\mathbf{w}$ によらない行列として、$\partial_{\mathbf{w}}(\mathbf{w}^{\mathrm{T}}\mathbf{a}) = \partial_{\mathbf{w}}(\mathbf{w}^{\mathrm{T}}\mathbf{a})^{\mathrm{T}} = \mathbf{a}, \partial_{\mathbf{w}}\mathbf{w}^{\mathrm{T}}\mathrm{A}\mathbf{w} = 2\mathbf{a}$ となることに注意。

事後分布は、尤度 (3.10)、事前分布 (3.48) を用いて

$$\begin{aligned} p(\mathbf{w}|\mathbf{t}) &\propto \prod_{n=1}^{N} \mathcal{N}(t_n|\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_n), \beta^{-1})\mathcal{N}(\mathbf{w}|\mathbf{m}_0, \mathbf{S}_0) \\ &\propto \exp\left\{-\frac{1}{2}\big(\beta(\mathbf{t}-\mathbf{\Phi w})^{\mathrm{T}}(\mathbf{t}-\mathbf{\Phi w}) + (\mathbf{w}-\mathbf{m})\big)\right\} \end{aligned}$$

展開してまとめると、以下 $\mathbf{w}$ によらない項を無視して

$$\propto \exp\Big\{-\frac{1}{2}\big(\mathbf{w}^{\mathrm{T}}\underbrace{(\mathbf{S}_0^{-1}-\beta\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi})}_{\mathbf{S}_N^{-1}}\mathbf{w} - \mathbf{w}^{\mathrm{T}}\underbrace{(\mathbf{S}_0^{-1}\mathbf{m}_0-\beta\mathbf{\Phi}^{\mathrm{T}}\mathbf{t})}_{\mathbf{S}_N^{-1}\mathbf{m}_N} - (\mathbf{m}_0^{\mathrm{T}}\mathbf{S}_0^{-1}+\beta\mathbf{t}^{\mathrm{T}}\mathbf{\Phi})\mathbf{w}\big)\Big\}$$

ここで、分散共分散行列が対称行列であることと

$$(\mathbf{x}-\mathbf{a})^{\mathrm{T}}\mathrm{A}(\mathbf{x}-\mathbf{a}) = \mathbf{x}^{\mathrm{T}}\mathrm{A}\mathbf{x} - \mathbf{a}^{\mathrm{T}}\mathrm{A}\mathbf{x} - \mathbf{x}^{\mathrm{T}}\mathrm{A}\mathbf{a} + \mathbf{a}^{\mathrm{T}}\mathrm{A}\mathbf{a}$$

より、平方完成と規格化によって

$$p(\mathbf{w}|\mathbf{t}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N) \tag{3.49}$$

と求まります。

逐次学習における事後分布の更新式 (演習 3.8)

共役な事前・事後分布は逐次学習 [2] において便利です。既にデータが N セットあり、事後分布が (3.49) で与えられていたとします。今、新たに $(\mathbf{x}_{N+1}, t_{N+a})$ というデータが与えられた時、新たな事後分布は (3.49) を事前分布として

$$\begin{aligned} p(\mathbf{w}, t_{N+1}) &\propto \mathcal{N}(t_{N+1}|\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_{N+1}), \beta^{-1})\mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N) \\ &\propto \exp\left\{-\frac{1}{2}\big(\beta(t_{N+1}-\mathbf{w}^{\mathrm{T}}\phi(\mathbf{x}_{N+1})^2) + (\mathbf{w}-\mathbf{m}_N)^{\mathrm{T}}\mathbf{S}_N^{-1}(\mathbf{w}-\mathbf{m}_N)\big)\right\} \end{aligned}$$

以下演習 3.7 と同様に平方完成すれば、

$$\begin{gathered} \mathbf{S}_{N+1}^{-1} = \mathbf{S}_N^{-1} + \beta\phi(\mathbf{x}_{N+1})\phi(\mathbf{x}_{N+1})^{\mathrm{T}}, \\ \mathbf{m}_{N+1} = \mathbf{S}_{N+1}(\mathbf{S}_N^{-1}\mathbf{m}_N + \beta\phi(\mathbf{x}_{N+1})t_{N+1}) \end{gathered}$$

として、

$$p(\mathbf{w}|t_{N+1}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_{N+1}, \mathbf{S}_{N+1})$$

と書くことができます。つまり、逐次学習では新たなデータから事後分布を更新するとき、これまでのデータから計算した事後分布を事前分布として扱い計算します。共役性から事前分布の形と事後分布の形は同じなので、このようにシステマティックに計算できるのです。

演習 3.11

新たなデータ点が追加されるたびに、事後分布は鋭くなっていきます。(PRML 図 3.7 参照) これは予測分布の広がりがだんだんと狭まること、つまり $\sigma_{N+1}^2 \leq \sigma_N^2$ が成り立つことを意味します。これを証明していきます。

[2] データがリアルタイムで与えられ、パラメータを逐次的に学習するもの

証明

$$\left(\mathrm{M} + \mathbf{v}\mathbf{v}^{\mathrm{T}}\right)^{-1} = \mathrm{M}^{-1} - \frac{(\mathrm{M}^{-1}\mathbf{v})(\mathbf{v}^{\mathrm{T}}\mathrm{M}^{-1})}{1 + \mathbf{v}^{\mathrm{T}}\mathrm{M}^{-1}\mathbf{v}} \tag{3.110}$$

という公式を用います。

$$\sigma_{N+1}^2 \leq \sigma_N^2 \quad = \quad \phi^{\mathrm{T}}(\mathbf{x})(\mathbf{S}_N - \mathbf{S}_{N+1})\phi(\mathbf{x})$$

ここで演習 3.8 の結果と上の公式 (3.110) から、$\phi_N = \phi(\mathbf{x}_N)$ として

$$\begin{aligned} \mathbf{S}_{N+1} = \mathbf{S}_N - \frac{(\mathbf{S}_N^{\mathrm{T}}\phi_{N+1})(\phi_{N+1}^{\mathrm{T}}\mathbf{S}_N)}{1 + \phi_{N+1}^{\mathrm{T}}\mathbf{S}_N\phi_{N+1}} \\ \therefore \phi^{\mathrm{T}}(\mathbf{x})(\mathbf{S}_N - \mathbf{S}_{N+1})\phi(\mathbf{x}) = \frac{\phi^{\mathrm{T}}(\mathbf{S}_N^{\mathrm{T}}\phi_{N+1})(\phi_{N+1}^{\mathrm{T}}\mathbf{S}_N)\phi}{1 + \phi_{N+1}^{\mathrm{T}}\mathbf{S}_N\phi_{N+1}} \end{aligned}$$

分散共分散行列の半正定値性から、最後の式は 0 以上となり、証明終了です。

演習 3.12/3.13

工事中