注：本レジュメの最後には Appendix を準備している。

1 識別関数

1.1 パーセプトロンアルゴリズム

本章の前書きにて、本章では一般化線形モデルを取り扱うと述べてきたが、このサブセクションでそのモデルの例を学ぶ。つまりパラメタ $\boldsymbol{w}$ に対して非線形関数を施すことになる。パーセプトロンアルゴリズムはその 1 例となる手法であり、本手法は 2 クラス分類において活躍するものとなっている。

まず一般化線形モデルは次のように構成できることは既に学んだ。

$$y(\boldsymbol{x}) = f(\boldsymbol{w}^T \phi(\boldsymbol{x})). \tag{1}$$

以下でいくつか必要な定義を行う。パーセプトロンモデルにおいては、上記モデル内の非線形関数を次のようなステップ関数で定義する。

$$f(a) = \begin{cases} +1 & a \geq 0 \\ -1 & a < 0 \end{cases} \tag{2}$$

また、目的変数のラベル表記を $t_n = 0, 1$ から $t_n = -1, 1$ へ変更する。この変更は、今後の議論を簡潔に進めるためである[*1]。

ここからは、これまでの通りパラメタを決定するために誤差関数の最小化を考えていきたい。しかしながら、従来の誤差関数[*2]では任意の点で勾配が 0 になってしまうため、最小化を実行することができない[*3]。したがって、パーセプトロンアルゴリズムにおいてパラメタを最小化するためには、新たな誤差関数を定義する必要がある。この新しく定義する誤差関数のことを、パーセプトロン規準[*4]と呼び、その定義式を以下に与える。

$$E_P(\boldsymbol{w}) = -\sum_{n \in M} \boldsymbol{w}^T \boldsymbol{\phi}_n t_n \tag{3}$$

M は全ての誤分類されたパターンの集合を表す。上記の式を得るために、以下 4 つ必要な約束をしている。

1. クラス C_1 の $\boldsymbol{x}_n$ に対して $\boldsymbol{w}^T \boldsymbol{\phi}_n > 0$、クラス C_2 の $\boldsymbol{x}_n$ に対して $\boldsymbol{w}^T \boldsymbol{\phi}_n < 0$ となるように $\boldsymbol{w}$ を定める。
2. 目的変数のラベルについては先の定義に従う。
3. 約束 1 と 2 によって常に $\boldsymbol{w}^T \boldsymbol{\phi}_n t_n > 0$ が満たされる。
4. 正しく分類できたパターンには誤差 0 を割り当て、誤分類したパターンでは $-\boldsymbol{w}^T \boldsymbol{\phi}_n t_n > 0$ を最小化する。

これはヒンジ損失という損失関数であり、数式で表現すると次のようになる[*5]。

$$loss(\boldsymbol{w}, \boldsymbol{x}, t_n) = max(0, -\boldsymbol{w}^T \boldsymbol{\phi}_n t_n) \tag{4}$$

パーセプトロン規準、つまり今回の誤差関数（式 3）について最小化を実行する。最小化の実行にあたっては確率勾配急降下法を用いることにする。パラメタ $\boldsymbol{w}$ に関する急降下法のアルゴリズムの式は以下で与えられる。

$$\boldsymbol{w}^{(\tau+1)} = \boldsymbol{w}^{(\tau)} - \eta \nabla E_P(\boldsymbol{w}) = \boldsymbol{w}^{(\tau)} + \eta \boldsymbol{\phi}_n t_n \tag{5}$$

---

*1 のちに学ぶ活性化関数との適合が良い。

*2 p.188 式（4.31）を参照せよ。

*3 ごめん意味不明だった。

*4 「規準」or「基準」どちらか正しいのか...

*5 つまり、ここでは誤差関数をヒンジ損失で定義しているのである。

上式と図 4.7 を使って、誤分類が起こるたびに決定境界がどのように変化するのか追跡する[*6]。ここは読み合わせ。

パーセプトロンの収束定理/パーセプトロンアルゴリズムの課題についても読み合わせ。

2　確率的生成モデル

本セクションでは、分類を識別関数を用いて行うのではなく、確率的な生成アプローチを用いて行うことにする。サブセクション 1.5.4[*7]でも学んだように、生成アプローチにおいては、ベイズの定理を用いて事後分布を計算する必要がある。

後のサブセクションにおいて様々なクラスの条件付き分布のもとで事後分布を計算したいが、その準備として、一般的に事前分布、クラスの条件付き密度を定義し、そこから事後分布の表式を得ることにしよう。

まず、2 クラス分類問題を考えていく。ベイズの定理を用いると、事後分布の表式はロジスティックシグモイド関数として式（4.57）のように表すことができる。ただし、シグモイド関数の引数 a は

$$a = \ln \frac{p(\boldsymbol{x}|C_1)p(C_1)}{p(\boldsymbol{x}|C_2)p(C_2)} \tag{6}$$

のように定義している。念のため、上式を式（4.57）の a に代入することでベイズの定理の形になることを計算して確かめる[*8]。

$$\begin{aligned}\sigma(\ln \frac{p(\boldsymbol{x}|C_1)p(C_1)}{p(\boldsymbol{x}|C_2)p(C_2)}) &= \frac{1}{1+exp[-\ln \frac{p(\boldsymbol{x}|C_1)p(C_1)}{p(\boldsymbol{x}|C_2)p(C_2)}]} \\ &= \frac{1}{1+exp[\ln \frac{p(\boldsymbol{x}|C_2)p(C_2)}{p(\boldsymbol{x}|C_1)p(C_1)}]} \\ &= \frac{1}{1+\frac{p(\boldsymbol{x}|C_2)p(C_2)}{p(\boldsymbol{x}|C_1)p(C_1)}} \\ &= \frac{p(\boldsymbol{x}|C_1)p(C_1)}{p(\boldsymbol{x}|C_1)p(C_1)+p(\boldsymbol{x}|C_2)p(C_2)}\end{aligned} \tag{7}$$

確かにベイズの定理の形に戻った。

ここで、ロジスティックシグモイド関数の性質を整理する。

・全実数軸（$-\infty < x < \infty$）を限られた区間 $(-1 < y < 1)$ に押し込む。

・以下のような対称性を持つ[*9]。

$$\sigma(-a) = 1 - \sigma(a) \tag{8}$$

・ロジスティックシグモイド関数の逆関数はロジット関数として以下の式で与えられる[*10]。

$$a = ln \frac{\sigma}{1-\sigma} \tag{9}$$

これは対数オッズという名前でも知られている量である[*11]。

[*6] 緑枠で囲まれた点が、なぜ今回特別に考慮すべき誤分類点であるかが理解できなかった。例えば右隣の赤点ではダメなのか？

[*7] p.52 を参照。

[*8] やることはただの算数ですが、a の表式が突拍子もない感じもするので本当に元に戻るかは手を動かして納得しましょう。

[*9] p.111 で学習済み。

[*10] 導出については、式（4.59）を a について解くだけのなので流石にパスします。

[*11] 「自然科学の統計学」p.236 で学習済み。

K クラス分類問題の場合は、事後分布の表式が式（4.62）で与えられ、これを正規化指数関数（ソフトマックス関数）と定義する[*12]。式（4.63）を式（4.62）の最右辺に代入することで、ベイズの定理の形になることは自明であるため、式（4.57-8）で行ったような検算は省略する。

ここで、ソフトマックス関数の性質を整理する。この性質の整理については、非常に説明がわかりやすいサイトがあったので、そちらを参照してほしい[*13]。

本セクションの冒頭でも述べたように、この後様々なクラスの条件付き密度分布について事後分布を求めていく。その際に、このセクションで定義した式（4.59）や式（4.62）が必要となる。

以降では、入力変数 $\boldsymbol{x}$ が連続値の場合と離散値の場合について、事後分布の表式を求めていくことにする[*14]。

2.1　連続値入力

クラスの条件付き確率密度がガウス分布で与えられるときに事後確率の表式を得る。2 クラス分類問題についてはロジスティックシグモイド関数を、K クラス分類問題についてはソフトマックス関数を適用してやれば良い。

前提として、確率密度の表式は以下のようにガウス分布で与えるとする。

$$p(\boldsymbol{x}|C_k) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} exp\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_k)^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_k)\} \tag{10}$$

まず、2 クラス分類問題について考える。クラス C_1 に対する事後分布をシグモイド関数を用いて表現したい。

[*12] ソフトマックス関数という名前の方が一般的であると思われる。特にニューラルネットワークにおいては、私はこちらの表現しか見たことがない。
[*13] 右記サイトを参照：https://mathtrain.jp/softmax
[*14] とは言うものの、本レジュメは連続値の場合しか取り扱いません。悪しからず。

シグモイド関数の引数部分の計算がポイントとなるので、この計算を実行する。

$$
\begin{aligned}
a &= \ln \frac{exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)\}p(C_1)}{exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_2)^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2)\}p(C_2)} \\
&= -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_2)^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2) + ln\frac{p(C_1)}{p(C_2)} \\
&= -\frac{1}{2}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x} + \frac{1}{2}\boldsymbol{\mu}_1^T\Sigma^{-1}\boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{\mu}_1 - \frac{1}{2}\boldsymbol{\mu}_1^T\Sigma^{-1}\boldsymbol{\mu}_1 \\
&+ \frac{1}{2}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{x} - \frac{1}{2}\boldsymbol{\mu}_2^T\Sigma^{-1}\boldsymbol{x} - \frac{1}{2}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{\mu}_2 + \frac{1}{2}\boldsymbol{\mu}_2^T\Sigma^{-1}\boldsymbol{\mu}_2 + ln\frac{p(C_1)}{p(C_2)} \\
&= \frac{1}{2}\boldsymbol{\mu}_1^T\Sigma^{-1}\boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{\mu}_1 - \frac{1}{2}\boldsymbol{\mu}_1^T\Sigma^{-1}\boldsymbol{\mu}_1 - \frac{1}{2}\boldsymbol{\mu}_2^T\Sigma^{-1}\boldsymbol{x} \\
&- \frac{1}{2}\boldsymbol{x}^T\Sigma^{-1}\boldsymbol{\mu}_2 + \frac{1}{2}\boldsymbol{\mu}_2^T\Sigma^{-1}\boldsymbol{\mu}_2 + ln\frac{p(C_1)}{p(C_2)} \\
&= \frac{1}{2}\boldsymbol{\mu}_1^T\Sigma^{-1}\boldsymbol{x} + \frac{1}{2}\boldsymbol{\mu}_1^T\Sigma^{-1}\boldsymbol{x} - \frac{1}{2}\boldsymbol{\mu}_2^T\Sigma^{-1}\boldsymbol{x} - \frac{1}{2}\boldsymbol{\mu}_2^T\Sigma^{-1}\boldsymbol{x} \\
&- \frac{1}{2}\boldsymbol{\mu}_1^T\Sigma^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_2^T\Sigma^{-1}\boldsymbol{\mu}_2 + ln\frac{p(C_1)}{p(C_2)} \\
&= \boldsymbol{\mu}_1^T\Sigma^{-1}\boldsymbol{x} - \boldsymbol{\mu}_2^T\Sigma^{-1}\boldsymbol{x} + w_0 \ (\because (4.67)) \\
&= (\Sigma^{-1}\boldsymbol{\mu}_1)^T\boldsymbol{x} - (\Sigma^{-1}\boldsymbol{\mu}_2)^T\boldsymbol{x} + w_0 \ (\because \Sigma^{-1}\text{の対称性}) \\
&= (\Sigma^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))^T\boldsymbol{x} + w_0 \\
&= \boldsymbol{w}^T\boldsymbol{x} + w_0 \ (\because (4.66))
\end{aligned} \tag{11}
$$

式（4.47）より事後分布の表式は

$$p(C_1|\boldsymbol{x}) = \sigma(\boldsymbol{w}^T\boldsymbol{x} + w_0) \tag{12}$$

となる。

ここで、式 12 と図 4.10 との対応を解釈したかったが、よくわからなかった[15]。

次に、K クラス分類問題について考えたかったが、式（4.68）が導出できなかった。

最後に、3 クラス分類問題について、図 4.11 を例として用いて考える。ポイントは図 4.11 右において、異なるクラス間の決定境界の形状及び表式について解釈を行うところである。決定境界の表式は、次のように両クラスが持つ共分散によって決まる。

・共分散が同じ：決定境界は $\boldsymbol{x}$ の線形関数（直線）

・共分散が異なる：決定境界は $\boldsymbol{x}$ の 2 次関数（放物線）

2.2 最尤解

前サブセクションでは、条件付きクラス確率密度を与えたうえでベイズの定理を適用して事後確率を求めた。そしてその事後確率を決定境界として、クラス分類を行った。

本サブセクションでは、事前確率パラメタを与えたうえで、さらに $\boldsymbol{x}$ の観測値と対応するラベルで構成される学習データを与えることで、尤度関数を定義する。ここから尤度関数の最大化を行い、事前確率/各クラスの平均および分散に関する最尤推定量を求める。この学習を通じて分布のフィッティングを行うことで、クラス分類が実行できるようになる。

[15] 図 4.10 右のグラフの形状について、なんとなくシグモイド関数の形をしていることはわかるのだが、なぜ横軸の値が小さい時の方が、縦軸の値が大きいのか理解できなかった。だって、シグモイド関数は単調増加関数のはずなのに...

最尤法の手続きとしては、これまで学んできた通りであり以下のステップで計算を進める。

1. （対数）尤度関数を定義する。
2. パラメタごとに尤度関数を最大化する、つまり尤度方程式を解く。

まず 2 クラス分類の例で学習を行う。まず、準備として教科書に書かれている通りに、各クラスの条件付き確率密度*16/ラベルのルール/事前確率パラメタを与える。

この準備のもとで、尤度関数は式（4.71）のように与えられる。この式は以下 3 点を意味している。

・ラベルが 1 である場合は、クラス C_1 の学習データを得たことになる。

・ラベルが 0 である場合は、クラス C_2 の学習データを得たことになる。

・トータルで N 個の学習データを得ている。

続いて、以下 3 つ（a〜c）のパラメタに対して（対数）尤度関数の最大化を実行する。まず、事前準備として、対数尤度関数を露わに書いておく*17。

$$
\begin{aligned}
\ln p(\boldsymbol{t}, \boldsymbol{X}|\pi, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}) &= \ln \Pi_{n=1}^{N} [\pi \mathrm{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_1, \boldsymbol{\Sigma})]^{t_n} [(1-\pi)\mathrm{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_2, \boldsymbol{\Sigma})]^{1-t_n} \\
&= \ln \Pi_{n=1}^{N} \pi^{t_n} [\mathrm{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_1, \boldsymbol{\Sigma})]^{t_n} (1-\pi)^{1-t_n} [\mathrm{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_2, \boldsymbol{\Sigma})]^{1-t_n} \\
&= \sum_{n=1}^{N} (\ln \pi^{t_n} + \ln [\mathrm{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_1, \boldsymbol{\Sigma})]^{t_n} + \ln (1-\pi)^{1-t_n} + \ln [\mathrm{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_2, \boldsymbol{\Sigma})]^{1-t_n}) \\
&= \sum_{n=1}^{N} (t_n \ln \pi + t_n \ln [\mathrm{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_1, \boldsymbol{\Sigma})] + (1-t_n) \ln (1-\pi) + (1-t_n) \ln [\mathrm{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_2, \boldsymbol{\Sigma})])
\end{aligned} \tag{13}
$$

ひとまずここで式変形を止めておく。ここまで変形すれば a. については最尤法を適用できる。

a. 事前確率パラメタ π

式 13 において、π に依存する項のみを取り出すと

$$
\sum_{n=1}^{N} t_n \ln \pi + \sum_{n=1}^{N} (1-t_n) \ln (1-\pi) \tag{14}
$$

この対数尤度に対して尤度方程式を解くことは初等的であるため省略する。π の最尤推定量が式（4.73）のように得られることも直観通りである。

b. 各クラスの期待値*18$\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$

式 13 から μ_1 に依存する項のみを取り出すと次のようになる。

$$
\begin{aligned}
\sum_{n=1}^{N} t_n \ln [\mathrm{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_1, \boldsymbol{\Sigma})] &= \sum_{n=1}^{N} t_n \ln p(\boldsymbol{x}_n|C_1) \\
&= \sum_{n=1}^{N} t_n \ln [\frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} exp\{-\frac{1}{2}(\boldsymbol{x}_n - \boldsymbol{\mu}_1)^T \Sigma^{-1} (\boldsymbol{x}_n - \boldsymbol{\mu}_1)\}] \\
&= \sum_{n=1}^{N} t_n [-\frac{1}{2}(\boldsymbol{x}_n - \boldsymbol{\mu}_1)^T \Sigma^{-1} (\boldsymbol{x}_n - \boldsymbol{\mu}_1)] + const.
\end{aligned} \tag{15}
$$

a. 同様に、このこの対数尤度に対して尤度方程式を解くことは初等的であるため省略する*19。式（4.75）の結果も直観通りであり、$\boldsymbol{\mu}_1$ に関する最尤推定量はクラス C_1 のラベルを持つ入力ベクトルについて平均を取っ

*16 各クラスの共分散行列は共通とする。

*17 正規分布の式をローマン体のエヌで表記しています。グニャっとしたエヌにしたかったのですが、やり方がわかりませんでした。

*19 行列の微分は含みますが、もはや我々にとっては算数と同じレベルで理解できているはず。

たものとなっている。$\boldsymbol{\mu}_2$ も全く同様の議論によって最優推定量を求めることができるが、本レジュメでは導出を省略する。

c. 共分散行列 $\boldsymbol{\Sigma}$

ここは tex で書くのを諦めた（> Go to App.1)。

さいごに

・K クラス分類の取り扱いは割愛

・頑健性の話については p.101 を参照

[Cの計算（式(4.77)導入）]

Gauss分布の式から、次の2つの部分に注目してやればいい。

・分母の共分散行列のdet部分

・exp内の共分散行列の部分
逆

したがって考えるべき対数尤度は次のようになる。

$$\sum t_n \ln \frac{1}{|\Sigma|^{1/2}} + \sum t_n \ln \left[\exp\left\{ -\frac{1}{2}(x_n-\mu_1)^T \Sigma^{-1} (x_n-\mu_1) \right\} \right]$$

$$+\sum (1-t_n) \ln \frac{1}{|\Sigma|^{1/2}} + \sum (1-t_n) \ln \left[\exp\left\{ -\frac{1}{2}(x_n-\mu_2)^T \Sigma^{-1} (x_n-\mu_2) \right\} \right]$$

$$= \underline{-\frac{1}{2}\sum t_n \ln|\Sigma|} \underline{-\frac{1}{2}\sum t_n (x_n-\mu_1)^T \Sigma^{-1} (x_n-\mu_1)}$$

$$\underline{-\frac{1}{2}\sum (1-t_n) \ln|\Sigma|} \underline{-\frac{1}{2}\sum (1-t_n) (x_n-\mu_2)^T \Sigma^{-1} (x_n-\mu_2)}$$

（赤線部）$= -\frac{1}{2}\sum t_n \ln|\Sigma| - \frac{1}{2}\sum (1-t_n) \ln|\Sigma|$

$$= \boxed{-\frac{N}{2} \ln|\Sigma|}$$

（青線部）$= -\frac{1}{2}\sum t_n (x_n-\mu_1)^T \Sigma^{-1} (x_n-\mu_1) - \frac{1}{2}\sum (1-t_n)(x_n-\mu_2)^T \Sigma^{-1} (x_n-\mu_2)$

$$= -\frac{1}{2}\left\{ \sum t_n (x_n-\mu_1)^T \Sigma^{-1} (x_n-\mu_1) + \sum (1-t_n)(x_n-\mu_2)^T \Sigma^{-1} (x_n-\mu_2) \right\}$$

$$= -\frac{N}{2}\left\{ \frac{1}{N}\sum t_n (x_n-\mu_1)^T \Sigma^{-1} (x_n-\mu_1) + \frac{1}{N}\sum (1-t_n)(x_n-\mu_2)^T \Sigma^{-1} (x_n-\mu_2) \right\}$$

$$= -\frac{N}{2} \mathrm{Tr}\left\{ \frac{1}{N}\sum t_n (x_n-\mu_1)^T \Sigma^{-1} (x_n-\mu_1) \right.$$
$$\left. + \frac{1}{N}\sum (1-t_n)(x_n-\mu_2)^T \Sigma^{-1} (x_n-\mu_2) \right\}$$ （∵ Tr(スカラー) =(スカラー)）

$$= -\frac{N}{2} \mathrm{Tr}\left\{ \frac{1}{N}\sum_{n\in C_1} (x_n-\mu_1)^T \Sigma^{-1} (x_n-\mu_1) \right.$$
$$\left. + \frac{1}{N}\sum_{n\in C_2} (x_n-\mu_1)^T \Sigma^{-1} (x_n-\mu_2) \right\}$$

$$= -\frac{N}{2}\mathrm{Tr}\left\{\Sigma^{-1}\frac{1}{N}\sum_{n\in C_1}(x_n-\mu_1)(x_n-\mu_1)^T\right.$$

$$\left.+\Sigma^{-1}\frac{1}{N}\sum_{n\in C_2}(x_n-\mu_2)(x_n-\mu_2)^T\right\} \quad (\because \mathrm{Tr}(ABC)=\mathrm{Tr}(CAB))$$

$$= -\frac{N}{2}\mathrm{Tr}\left[\Sigma^{-1}\left\{\frac{1}{N}\sum_{n\in C_1}(x_n-\mu_1)(x_n-\mu_1)^T+\frac{1}{N}\sum_{n\in C_2}(x_n-\mu_2)(x_n-\mu_2)^T\right\}\right]$$

($\because$ Trの線形性)

$$= -\frac{N}{2}\mathrm{Tr}\left\{\Sigma^{-1}\left(\frac{N_1}{N}\cdot\frac{1}{N_1}\sum_{n\in C_1}(x_n-\mu_1)(x_n-\mu_1)^T+\frac{N_2}{N}\cdot\frac{1}{N_2}\sum_{n\in C_2}(x_n-\mu_2)(x_n-\mu_2)^T\right)\right\}$$

$$= -\frac{N}{2}\mathrm{Tr}\left\{\Sigma^{-1}\left(\frac{N_1}{N}S_1+\frac{N_2}{N}S_2\right)\right\} \quad (\because (4.79-80))$$

$$= \boxed{-\frac{N}{2}\mathrm{Tr}(\Sigma^{-1}S)} \quad (\because (4.78))$$

以上から対数尤度は次のようになる。

$$-\frac{N}{2}\ln|\Sigma|-\frac{N}{2}\mathrm{Tr}(\Sigma^{-1}S)$$

ここから尤度方程式を作って解く。その際にp.316の公式を用いる。

$$\frac{\partial}{\partial\Sigma}\left\{-\frac{N}{2}\ln|\Sigma|-\frac{N}{2}\mathrm{Tr}(\Sigma^{-1}S)\right\}=0$$

$$\frac{\partial}{\partial\Sigma}\ln|\Sigma|+\frac{\partial}{\partial\Sigma^{-1}}\frac{\partial\Sigma^{-1}}{\partial\Sigma}\mathrm{Tr}(\Sigma^{-1}S)=0 \quad (\because \text{Chain Rule})$$

$$(\Sigma^{-1})^T+\underline{\frac{\partial\Sigma^{-1}}{\partial\Sigma}}\,\underline{\frac{\partial}{\partial\Sigma^{-1}}\mathrm{Tr}((\Sigma^{-1})^TS)}=0 \quad (\because (C.28))$$

$$\Sigma^{-1}+\underline{(-\Sigma^{-1}\Sigma^{-1})}\cdot\underline{S^T}=0$$

左からΣを2回かけて

$$\Sigma-S^T=0$$

$$\Sigma=S \quad (\because \Sigma\text{が対称})$$

したがいΣの最尤推定量はSとなる。