*VII-2-2. ソフトマックス関数と多クラス判別*

*VII-2-2-1.　ソフトマックス関数*

シグモイド関数を使ったロジット回帰によって、等分散性の問題はある程度解消されました。シグモイド関数の導関数は、あることが起こるか起こらないかという背反事象の確率を表しています。したがって、「そうであるか」、「そうでないか」という２項対立的な判別や分類には利用できます。しかし、一つの式で多クラスの判別が同時に可能ならば、その方がより便利です。３クラス以上の判別でも、「あるクラスであるか」、「それ以外か」という考え方を組み合わせれば、ロジスティック解析的な考え方が可能です。ソフトマックス関数は、シグモイド関数の特性を残して多クラス判別用の関数として発展させたものです。

互いに背反するＪ個の事象があるとします。その事象に０からＪ-1 の番号を与えます。ソフトマックス関数以下の関数で、$\boldsymbol{y}$が与えられた時にその中で事象$i$が起こる確率を表しています。。

$$P(i|\boldsymbol{y}) = \varsigma(\beta_i) = \frac{e^{\beta_i}}{\sum_{j=0}^{J-1} e^{\beta_j}} \qquad \textit{式 85}$$

$\beta_i, \beta_j$はベクトル空間上で、判別平面からそのデータまでの距離で、これが、定数項を成分に加えた法線ベクトルとダミー変数を成分に加えた変数ベクトルの内積になることは前項で説明しました。

$$\beta_i = b_0 + b_1 y_{i1} + b_2 y_{i2} + \cdots + b_m y_{im} = \boldsymbol{b} \cdot \boldsymbol{y}$$

ただし

$$\boldsymbol{b} = (b_0 \quad \cdots \quad b_m)$$

$$\boldsymbol{y} = (y_0 \quad \cdots \quad y_m)$$

$$y_0 = 1$$

ドット･は内積

念のために、累積確率曲線として使えることを確認します。

$$\lim_{\beta_i \to -\infty} \frac{e^{\beta_i}}{\sum_{j=0}^{J-1} e^{\beta_j}} = 0, \qquad \lim_{\beta_i \to \infty} \frac{e^{\beta_i}}{\sum_{j=0}^{J-1} e^{\beta_j}} = 1$$

で、それぞれの事象が起こる確率の総和は

$$\sum_{i=0}^{\mathrm{J}-1} P(i|\boldsymbol{y}) = \frac{1}{\sum_{j=0}^{J-1} e^{\beta_j}} \sum_{i=0}^{J-1} e^{\beta_i} = 1$$

$\beta_i$で偏微分してみます。

$$\varsigma(\beta_i) = \frac{e^{\beta_i}}{\sum_{j=0}^{J-1} e^{\beta_j}} = e^{\beta_i} \cdot \left( \sum_{j=0}^{J-1} e^{\beta_j} \right)^{-1}$$

$$f(\beta_i) = e^{\beta_i}, \quad g(\beta_i) = \left(\sum_{j=0}^{J-1} e^{\beta_{ij}}\right)^{-1}$$

$$\varsigma(\beta_i) = f(\beta_i)g(\beta_i)$$

$$\frac{\partial f(\beta_i)}{\partial \beta_i} = e^{\beta_i}, \ \frac{\partial g(\beta_i)}{\partial \beta_i} = -\left(\sum_{j=0}^{J-1} e^{\beta_j}\right)^{-2} e^{\beta_i}$$

$$\frac{\partial \varsigma(\beta_i)}{\partial \beta_i} = \frac{\partial f(\beta_i)}{\partial \beta_i} g(\beta_i) + f(\beta_i)\frac{\partial g(\beta_i)}{\partial \beta_i} = e^{\beta_i}\left(\sum_{j=0}^{J-1} e^{\beta_j}\right)^{-1} + e^{\beta_i}\left(-\left(\sum_{j=0}^{J-1} e^{\beta_j}\right)^{-2} e^{\beta_i}\right)$$

$$= e^{\beta_i}\left(\sum_{j=0}^{J-1} e^{\beta_j}\right)^{-1}\left(1 - e^{\beta_i}\left(\sum_{j=0}^{J-1} e^{\beta_j}\right)^{-1}\right)$$

ここで

$$e^{\beta_i} \cdot \left(\sum_{j=0}^{J-1} e^{\beta_j}\right)^{-1} = \frac{e^{\beta_i}}{\sum_{j=0}^{J-1} e^{\beta_j}} = \varsigma(\beta_i)$$

だから、

$$\frac{\partial \varsigma(\beta_i)}{\partial \beta_i} = \varsigma(\beta_i)\big(1 - \varsigma(\beta_i)\big)$$

となります。確率で表すと

$$P(i|\boldsymbol{y}) = \varsigma(\beta_i)$$
$$P(not\ i|\boldsymbol{y}) = 1 - \varsigma(\beta_i)$$

なので、

$$\frac{\partial P(i|\boldsymbol{y})}{\partial \beta_i} = P(i|\boldsymbol{y})P(not\ i|\boldsymbol{y})$$

$$\frac{\partial P(not\ i|\boldsymbol{y})}{\partial \beta_i} = -P(i|\boldsymbol{y})P(not\ i|\boldsymbol{y})$$

となります。$P(i|\boldsymbol{y}) \geq 0, \ P(not\ i|\boldsymbol{y}) \geq 0$ですから、

$$\frac{\partial P(i|\boldsymbol{y})}{\partial \beta_i} \geq 0$$

このように、ソフトマックス関数は、単調増加関数で、０～１で変動し、すべての事象が起こる確率の総和が１となるので、あることが起こることの確率として使えます。あることが起こることと起こらないことの２項対立的にとらえると、その確率はシグモイド関数になっています。

### *VII-2-2-2. ニューラルネットワークモデルを使った多クラス同時判別の計算フロー*

図 115 に、データ$\boldsymbol{y}_n = (y_{n0} \quad y_{n1} \quad y_{n2})$が与えられた時、それが、クラス 0 、 1 、 2 に属する確率をソフトマックス関数を使って計算する作業フローを、ニューラルネットワークモデルで使われる神経細胞の結合を模した図で表しました。なお、$y_{n0}$は$y_{n0} = 1$のダミー変数です。中央の楕円は神経細胞で矢印は神経線維です。神経線維の先端がシナプスでシグナルは矢印の方向一方向のみに伝わります。細胞の数は違いますが、一対の 2 クラス判別でも、多クラス判別でもこの流れは同じです。$\boldsymbol{y}_n$はデータですが、神経伝達的には感覚受容器で受け取った刺激と考えれば良いでしょう。刺激(多次元の情報)は神経細胞の矢印の方向に伝わりシナップスの部分に到達します。シナップスでは刺激が到達すると神経伝達物質がシナプス感激に放出され、シナップス後細胞の受容体に結合して刺激が伝達されます。情報を受け取る細胞によって、個々の刺激（情報）に対する受け取る情報の大きさが変わります。この大きさの違いをニューラルネットワークモデルでは「重さ」と言います。それぞれの「重さ」で増幅された刺激の総和を入力総和といいます。

$$\beta_{nk} = \sum_{d=0}^{D} b_{kd} y_{nd} = \boldsymbol{y}_n \cdot \boldsymbol{b}_k$$

前項で説明したように入力総和は以下の式で表される判別境界平面からのデータ$\boldsymbol{y}_n$までの距離です。

$$b_{k0} y_0 + b_{k1} y_1 + \cdots + b_{kD} y_D = 0$$

この入力総和が神経細胞内で変換されて神経細胞から出力されます。この変換に使う関数を活性化関数と言います。一対の判別の時は関数にシグモイド関数を使いましたが、多クラス判別ではソフトマックス関数を使うことにします。

この例では、神経細胞は 1 層しかないので、入力された刺激が入力総和に統合されたのちに変換されて、そのまま確率として出力されます。次に紹介するニューラルネットワークモデルでは神経細胞が 2 層になって、第一層から出力された刺激は次の層の神経細胞に入って刺激として伝わり、他の細胞からの刺激とともに入力総和となって変換後に確率とし

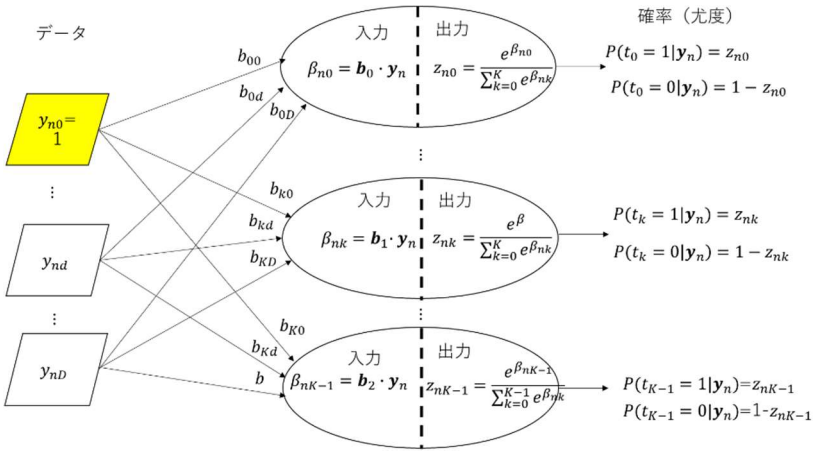


図 115 ソフトマックス関数による 3 クラス同時判別のフロー

て出力されます。もっと複雑なモデルでは、何層ものネットワーク構造になっていて、刺激が次々と次の層に伝わっていきます。そうしたネットワークの作り方は別に学ばなければなりませんが、一つの層での作業がデータ→入力総和→変換→出力という内容だということを頭に入れておいてください。大きな流れを理解しておかないと、次の項の解析学的な計算プロセスの説明でどの部分の説明かわからなくなります。

それはそれとして、この解説で使っている記号の使い方は一般的ではありません。多くの教科書では重さを$\boldsymbol{w}$とか$\boldsymbol{v}$で表し、入力総和を$\boldsymbol{a}$や$\boldsymbol{b}$で表して

$$\boldsymbol{a}_n{}^{\boldsymbol{T}} = \boldsymbol{w}\boldsymbol{x}_n{}^{\boldsymbol{T}}$$

$$\boldsymbol{w} = \begin{bmatrix} w_{10} & \cdots & w_{1m} & \cdots & w_{1M} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{j0} & \cdots & w_{jm} & \cdots & w_{jM} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{J0} & \cdots & w_{Jm} & \cdots & w_{JM} \end{bmatrix}$$

のような記述の仕方をしています。他の教科書を参考にする場合は注意してください。計算上求めているのは未知数で、未知数には$w, x, y, z$などを使うのが暗黙のルールだという理屈もわかるのですが、この解説は線形判別分析から出発しています。「重さ」というのはベクトルを線形結合する際の個々のベクトルの倍率という意味でが、線形判別分析ではこの「重さ」は境界平面を決めている式の係数です。だとすれば、係数には$a, b, c, d$などを使うというのも、数学の暗黙の了解です。線形判別分析のところで、平面と点の距離は、それぞれ定数項とダミー変数を加えた、平面の法線ベクトルと点の変数ベクトルの内積だと説明しました。流れを大切にして、一般的ではない記号のつけ方をしました。

### *VII-2-2-3. 多クラス同時判別の解析学的説明*

ソフトマックス関数は、１つのクラスに関しては、データがそのクラスに属する確率とそのクラスに属さない確率の関係がシグモイド関数になっています。そして、そのデータがあるクラスに属する確率の総計が１になっています。ですから、あるクラスについて個々のデータがそのクラスに属する場合はその確率、属さない場合はそのクラスに属さない確率を尤度として、すべての尤度を掛け合わせ、さらにそれらをすべてのクラスについてかけ合わせれば、すべてのデータをそのように分類することの妥当性（総尤度）が計算できます。実際の計算では、尤度の対数をとって対数尤度とし、正負の符号を入れ替えて、総データ数で割ったものを平均交差エントロピー誤差として、勾配降下法によって最小の平均交差エントロピー誤差を与える係数を推定します。クラス数をＫ、変数の数をＤとすると、定数項も含めて、$K \times (D+1)$この係数を推定します。そうした違いはありますが、素二マックス関数は本質的にはシグモイド関数ですから、交差エントロピー誤差な微分などはシグモイド関数の場合と同じです。ですからこの項の解説は前項までのいくつかの解説の繰り返しになります。

判別分析の数学とは、どのクラスに分類されるのかがわかっているデータがあって、その尤度（正しいとされる確率）が最大になるようにすべての係数を決めるということです。

あるクラス$k$に着目すると、あるデータ$n$について、正しいクラスの判別には２つの場合があります。そのクラスに属するという判定が正しいという場合と、そのクラスに属さないという判定が正しいという場合です。その両方を含めて、あるクラスにあるデータが属するか否かが正しい確率を$P_{nk}$と表します。全データ・全クラスについて、$NK$の尤度があり、そのすべての積が全尤度$L$です。

$$L = \prod_{k=1}^{K} \prod_{n=1}^{N} P_{nk}$$

このままでは数値が大きすぎてコンピュータの計算限界を超えてしまうから対数尤度

$LL$*を取ります。*

$$LL = \sum_{k=1}^{K} \sum_{n=1}^{N} \log_e P_{nk}$$

勾配降下法で微分するために平均交差エントロピー誤差にします。

$$CEE = -\frac{1}{KN} \sum_{k=1}^{K} \sum_{n=1}^{N} \log_e P_{nk}$$

負号をΣの中に入れて、Σの中だけを考えます。

$$cee = -\log_e P_{nk}$$

この偏微分は以下の通り

$$\frac{\partial cee}{\partial P_{nk}} = \frac{-1}{P_{nk}}$$

$P_{nk}$には２つの場合があります。

データがクラス$k$に属する場合

$$P_{nk} = P_{nk}(k|\boldsymbol{y})$$

$$\frac{\partial cee}{\partial P_{nk}} = \frac{-1}{P_{nk}(k|\boldsymbol{y})}$$

クラス$k$に属さない場合

$$P_{nk} = P_{nk}(not\ k|\boldsymbol{y})$$

$$\frac{\partial cee}{\partial P_{nk}} = \frac{-1}{P_{nk}(not\ k|\boldsymbol{y})}$$

それぞれの場合の確率をソフトマックス関数で表すと

$$P_{nk}(k|\boldsymbol{y}) = \frac{e^{\beta_{nk}}}{\sum_{j=1}^{K} e^{\beta_{nj}}}$$

$$P_{nk}(not\ k|\boldsymbol{y}) = 1 - P_{nk}(k|\boldsymbol{y}) = 1 - \frac{e^{\beta_{nk}}}{\sum_{j=1}^{K} e^{\beta_{nj}}}$$

それぞれを、$y_{nk}$で偏微分すると

$$\frac{\partial P_{nk}(k|\boldsymbol{y})}{\partial \beta_{nk}} = P_{nk}(k|\boldsymbol{y})P_{nk}(not\ k|\boldsymbol{y})$$

$$\frac{\partial P_{nk}(not\ k|\boldsymbol{y})}{\partial \beta_{nk}} = -P_{nk}(k|\boldsymbol{y})P_{nk}(not\ k|\boldsymbol{y})$$

データがクラス$k$に属する場合

$$\frac{\partial cee}{\partial \beta_{nk}} = \frac{\partial cee}{\partial P_{nk}(k|\boldsymbol{y})}\frac{P_{nk}(k|\boldsymbol{y})}{\partial \beta_{nk}} = \frac{-1}{P_{nk}(k|\boldsymbol{y})}P_{nk}(k|\boldsymbol{y})P_{nk}(not\ k|\boldsymbol{y}) = -P_{nk}(not\ k|\boldsymbol{y})$$

$$= -\big(1 - P_{nk}(\ k|\boldsymbol{y})\big) = P_{nk}(\ k|\boldsymbol{y}) - 1$$

データがクラス$k$に属さない場合

$$\frac{\partial cee}{\partial \beta_{nk}} = \frac{\partial cee}{\partial P_{nk}(not\ k|\boldsymbol{y})}\frac{P_{nk}(not\ k|\boldsymbol{y})}{\partial \beta_{nk}} = \frac{-1}{P_{nk}(not\ k|\boldsymbol{y})}\big(-P_{nk}(k|\boldsymbol{y})P_{nk}(not\ k|\boldsymbol{y})\big)$$

$$= P_{nk}(\ k|\boldsymbol{y})$$

です。クラス$k$に属する場合を$t = 1$、属さない場合を$t = 0$と表現して、場合分けしないでこの結果を表すと、

$$\frac{\partial cee}{\partial \beta_{nk}} = P_{nk}(\ k|\boldsymbol{y}) - t_{nk}$$

このように極めて簡単な形にまとまります。この微分の結果もシグモイド関数と同じです。

この表記の仕方を使って、$t_1$を人種 A であるかそうでないか、$t_2$を人種 B であるかそうでないか、$t_3$を人種 C であるかそうでないかという形で表して、データを整理すると、表 63 のようになります。今表の中で T の列のようなクラス別を表す表現を 1 of K 符号化法あるいは one-hot エンコーディングと呼びます。データ表記のテクニックの一つですが、表計

表 63. One-hot エンコーディングによる表記と確率・微分

| data No | T | | | $\boldsymbol{y}$ | | 確率 | | | $beta_{nk}$での微分 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | $t_1$ | $t_2$ | $t_3$ | 体長 | 体重 | $P_1$ | $P_2$ | $P_3$ | $dE_1$ | $dE_2$ | $dE_3$ |
| 1 | 0 | 1 | 0 | 144.8 | 78.0 | P(not A) | P(B) | P(not C) | P(A) | P(B)-1 | P(C) |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | | |
| 38 | 1 | 0 | 0 | 173.7 | 78.0 | P(A) | P(not B) | P(not C) | P(A)-1 | P(B) | P(C) |
| 39 | 1 | 0 | 0 | 173.9 | 69.1 | P(A) | P(not B) | P(not C) | P(A)-1 | P(B) | P(C) |
| 40 | 0 | 0 | 1 | 175.9 | 44.1 | P(not A) | P(not B) | P(C) | P(A) | P(B) | P(C)-1 |
| 41 | 0 | 1 | 0 | 176.7 | 145.5 | P(not A) | P(B) | P(not C) | P(A) | P(B)-1 | P(C) |
| 42 | 1 | 0 | 0 | 176.8 | 68.4 | P(A) | P(not B) | P(not C) | P(A)-1 | P(B) | P(C) |
| 43 | 1 | 0 | 0 | 176.8 | 70.7 | P(A) | P(not B) | P(not C) | P(A)-1 | P(B) | P(C) |
| 44 | 0 | 0 | 1 | 177.6 | 46.5 | P(not A) | P(not B) | P(C) | P(A) | P(B) | P(C)-1 |
| 45 | 0 | 1 | 0 | 178.8 | 122.7 | P(not A) | P(B) | P(not C) | P(A) | P(B)-1 | P(C) |
| 46 | 0 | 0 | 1 | 179.0 | 48.6 | P(not A) | P(not B) | P(C) | P(A) | P(B) | P(C)-1 |
| 47 | 0 | 0 | 1 | 179.2 | 48.4 | P(not A) | P(not B) | P(C) | P(A) | P(B) | P(C)-1 |
| 48 | 0 | 1 | 0 | 180.7 | 138.8 | P(not A) | P(B) | P(not C) | P(A) | P(B)-1 | P(C) |
| ⋮ | | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 71 | 0 | 0 | 1 | 211 | 64.5 | P(not A) | P(not B) | P(C) | P(A) | P(B) | P(C)-1 |

算の時には便利なテクニックです。
一般化して書くと

$$t = (t_1 \quad \cdots \quad t_k \quad \cdots \quad t_K) \quad 、t_k = 0\ or\ 1、\sum_{k=1}^{K} t_k = 1$$

というルールになりますが、例に挙げた人種の場合の具体例は次のようになります。

人種 A, $T = (1 \quad 0 \quad 0)$、人種 B, $T = (0 \quad 1 \quad 0)$、人種 C, $T = (0 \quad 0 \quad 1)$

これを使うと、人種 B である確率は、$P_{nk}(\,(0 \quad 1 \quad 0)|\boldsymbol{y})$と書けます
最後に$\beta_{nk}$の係数$\boldsymbol{b}$の要素による偏微分です。

$$\boldsymbol{\beta} = \boldsymbol{y}\boldsymbol{b}^T + \boldsymbol{e}$$

$$\boldsymbol{b} = \begin{pmatrix} b_{10} & b_{11} & \cdots & b_{1D} \\ \vdots & \vdots & \ddots & \vdots \\ b & b_{K1} & \cdots & b_{KD} \end{pmatrix}$$

$$\boldsymbol{y} = \begin{pmatrix} y_{10} & y_{11} & \cdots & y_{1D} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N0} & y_{N1} & \cdots & y_{ND} \end{pmatrix}$$

$$\text{ただし、} y_{n0} = 1$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{11} & \cdots & \beta_{1K} \\ \vdots & \ddots & \vdots \\ \beta_{N1} & \cdots & \beta_{N1} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_N \end{pmatrix}$$

$\boldsymbol{\beta}$は$\boldsymbol{y}$を重さとさとした$\boldsymbol{b}$の線形結合ですから、

$$\frac{\partial \beta_{nk}}{\partial a_{kd}} = \frac{\sum_{d=0}^{D} a_{kd} y_{nd}}{\partial a_{kd}} = y_{nd}$$

となります。$cee$は入れ子関数ですから、$cee$の$\boldsymbol{b}$による微分は

$$\frac{\partial cee(b_{kd})}{\partial b_{kd}} = \frac{\partial cee}{\partial \beta_{nk}} \frac{\partial \beta_{nk}}{\partial a_{kd}} = (P_{nk}(\,k|\boldsymbol{y}) - t_{nk}) y_{nd}$$

これをこれらの総和として平均エントロピー誤差の微分を求めます

$$CEE = -\frac{1}{KN} \sum_{k=1}^{K} \sum_{n=1}^{N} \log_e P_{nk}$$

$$cee = -\log_e P_{nk}$$

$$CEE = \frac{1}{KN} \sum_{k=1}^{K} \sum_{n=1}^{N} cee$$

$$\frac{\partial CEE}{\partial b_{kd}} = \frac{1}{KN} \sum_{k=1}^{K} \sum_{n=1}^{N} \frac{\partial cee}{\partial b_{kd}} = \frac{1}{KN} \sum_{k=1}^{K} \sum_{n=1}^{N} (P_{nk}(\,k|\boldsymbol{y}) - t_{nk}) y_{nd}$$

となります。勾配降下法では、

$$\boldsymbol{b} = \begin{pmatrix} b_{10} & b_{11} & \cdots & b_{1D} \\ \vdots & \vdots & \ddots & \vdots \\ b_{K0} & b_{K1} & \cdots & b_{KD} \end{pmatrix}$$

に適当な初期値をあたえて、すべての$\frac{1}{N}\sum_{n=1}^{N} \frac{\partial cee}{\partial b_{kd}}$が近似的に 0 になるまで以下のように$b_{kd}$

を更新して、すべてが近似的に 0 になった時の$b_{kd}$を最適値とします。

$$b_{kd} \leftarrow b_{kd} - lr\frac{1}{N}\sum_{n=1}^{N}(P_{nk}(\ k|\boldsymbol{y}) - t_{nk})y_{nd}$$

$lr$は学習率

例に挙げた人種を体格で判別する例では、人種の数が 3 で説明変数が 2 ですから、推定する係数の数は$3 \times (2+1) = 9$個です。つまり、 k は 1 から 3 まで、d は 0 から 2 までの組み合わせのすべてについて、近似的になるまで計算を繰り返すということです。ソフトマックス関数の場合、極値が 1 つなので問題がないのですが、この勾配降下法を他の関数の係数の最適化に用いる場合には若干注意が必要です。極値をいくつか持つ関数の場合、全体最小とならない部分最小を与える極値に落ち込んでしまった場合、そこから出られず、最小値とならない係数で止まってしまう場合があります。そういう場合は、初期値を与えなおして計算しなければなりません。また、学習率をどのくらいにするかも、計算速度と正確性にかかわる問題です。ここに示したのは、最も単純な勾配降下法です。学習率をどのようにするかも含めて、様々な工夫がされて、改良された勾配降下法がたくさん提案されていますが、その詳細をここで解説することはここではできませんので、出版された本やネットで調べてください。

### *VII-2-2-4. ソフトマックス関数による３クラス同時判別の計算*

表 58 のデータで 3 つの人種の同時判別をソフトマックス関数でやってみました。
ここでは結果のみを示します。表 64 は推定された係数です。
これらの係数から、

$$\beta_A = 41.40477 + 0.075043\ 体長 - 0.56176\ 体重$$
$$\beta_B = 0.012199 - 0.8743\ 体長 + 1.949759\ 体重$$
$$\beta_C = -41.417 + 0.79926\ 体長 - 1.388\ 体重$$

という式が出来ます。シグモイド関数の場合は 2 項対立的で、互いに排反事象を扱っているので、$\beta = 0$とおいた式はは確率 1/2 の判別直線ですが、ソフトマックス関数では$\beta = 0$と置いても判別直線は得られません。$\beta$の式の意味を図形的に説明するのは難しいのですが、たとえば身長 160cm、体重 76kg の人の場合、表 65 のように計算さ入れて、人種 A である確率が 82%、人種 B である確率は 8%という推測が出来ます。それぞれの人種である確率の分布は、図 116 の 3d 画像のようになっていて、これを使えば、図 117 の等高線図を描くことが出来ます。

表 64. 推定された係数。

| 人種 | $b_0$ | $b_1$ | $b_2$ |
|---|---|---|---|
| A | 41.40477 | 0.075043 | -0.56176 |
| B | 0.012199 | -0.8743 | 1.949759 |
| C | -41.417 | 0.79926 | -1.388 |

表 65.確率の計算

| 人種 | beta | exp(beta) | P |
|---|---|---|---|
| A | 10.71799 | 45161.1 | 0.917787 |
| B | 8.305337 | 4045.407 | 0.082213 |
| C | -19.0233 | 5.47E-0.9 | 1.11E-13 |
| 合計 | | 49206.51 | 1 |

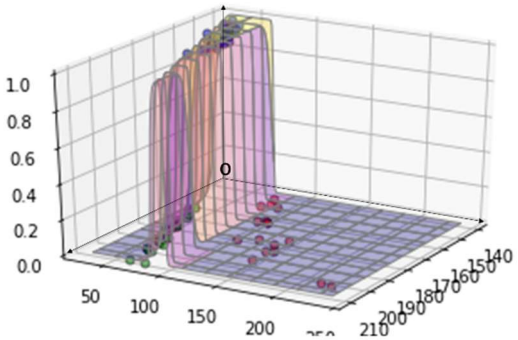


図 116-1.人種 A である確率の分布

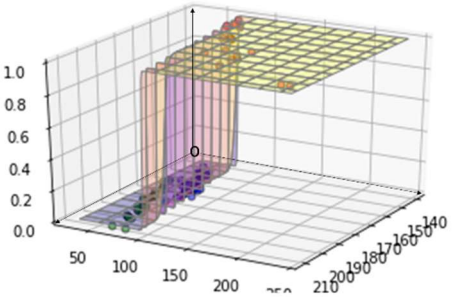


図 116-2.人種 B である確率の分布

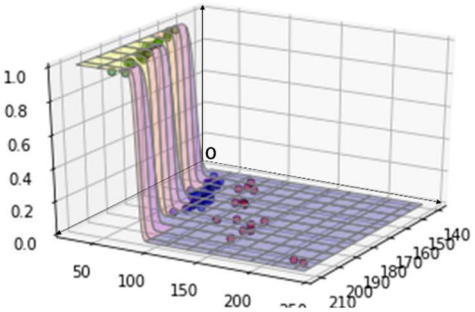


図 116-3.人種 C である確率の分布

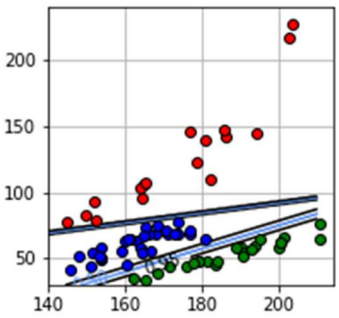


図 117. 各人種である確率の等高線
青線：確率 0.5、黒線:確率 0.95,0.05

*VII-2-3-5. Python を使った多クラス同時判別*

図 116 や図 117 は Python を使って作りました。計算は Python でやりコードリストは Jupyter notebook で書きました。サンプルデータは Excel の csv で作りました。これらは「参考資料」-「python」に「VII-2-２．ソフトマックス」で見ることが出来ます（サンプルデータは「Sample data」にあります)。結果は Jupyter Notebook のホルダーに、parameterBLBW1.csv という名前で推定された係数が保存されます。また、作った図も、BLBWscatter.png、BLBWcontour.png、BLBW3d1.png などの名前で保存されます。