概 要

講義資料の解答をここに書く。中間試験・期末試験の問題はここからしか出ない。

Lecture 2 の回答

総和のキホン

$\sum_{k=1}^{n}(k^2-2k+1)$ を求める。

$$\sum_{k=1}^{n}(a_k+b_k)=\sum_{k=1}^{n}a_k+\sum_{k=1}^{n}b_k$$

より、$\sum_{k=1}^{n}(k^2-2k+1)=\sum_{k=1}^{n}k^2-\sum_{k=1}^{n}2k+\sum_{k=1}^{n}1=\sum_{k=1}^{n}k^2-\sum_{k=1}^{n}2k+n$

$$\sum_{k=1}^{n}ca_k=c\sum_{k=1}^{n}a_k$$

より、$\sum_{k=1}^{n}k^2-\sum_{k=1}^{n}2k+n=\sum_{k=1}^{n}k^2-2\sum_{k=1}^{n}k+n$

$$1^2+2^2+3^2+\cdots+n^2=\sum_{k=1}^{n}k^2=\frac{1}{6}n(n+1)(2n+1)$$

より、$\sum_{k=1}^{n}k^2-2\sum_{k=1}^{n}k+n=\frac{1}{6}n(n+1)(2n+1)-2\sum_{k=1}^{n}k+n$

$$1+2+3+\cdots+n=\sum_{k=1}^{n}k=\frac{1}{2}n(n+1)$$

より、$\frac{1}{6}n(n+1)(2n+1)-2\sum_{k=1}^{n}k+n=\frac{1}{6}n(n+1)(2n+1)-n(n+1)+n=\frac{1}{3}n^3-\frac{1}{2}n^2+\frac{1}{6}n$

min 関数

$$\min\{0,1,2\}=0$$

この場合、min 関数は与えられた集合のうち最小の要素を返す。

$$\min_{x}(x-1)^2=0$$

この場合、min 関数は与えられた関数について、min の下のパラメータ（ここでは x）を調節して得られる関数の最小値を返す。

$$\operatorname*{argmin}_{x}(x-1)^2=1$$

argmin 関数は、与えられた関数を argmin の下のパラメータ（ここでは x）を調節して最小化する時の、パラメータの値を返す。

罰金項を含めた最適化

$\frac{\partial\tilde{E}(\mathbf{w})}{\partial w_i}$ を算出する。ここで、

$$\tilde{E}(\mathbf{w})=\frac{1}{2}\sum_{n=1}^{N}\{y(x_n,\mathbf{w})-t_n\}^2+\frac{\lambda}{2}\|\mathbf{w}\|^2$$

$$\|\mathbf{w}\|^2\equiv\mathbf{w}^\top\mathbf{w}=w_0^2+w_1^2+\cdots+w_M^2$$

$$y(x, \mathbf{w}) = \sum_{j=0}^{M} w_j x^j$$

とする。
$\frac{\partial E(\mathbf{w})}{\partial w_i} = \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\} x_n^i$ より、あとは $\frac{\partial}{\partial w_i}(\frac{\lambda}{2}\|\mathbf{w}\|^2)$ を計算すれば良い。

$$\begin{aligned}
&\frac{\partial}{\partial w_i}(\frac{\lambda}{2}\|\mathbf{w}\|^2) \\
&= \frac{\lambda}{2}\frac{\partial}{\partial w_i}(w_0^2 + w_1^2 + \cdots + w_i^2 + \cdots + w_M^2) \\
&= \lambda w_i
\end{aligned}$$

よって、$\frac{\partial \tilde{E}(\mathbf{w})}{\partial w_i} = \sum_{n=1}^{N} \{y(x_n, \mathbf{w}) - t_n\} x_n^i + \lambda w_i$ となる。

Lecture 3 の回答

対数関数と総乗

$\ln \prod_{n=1}^{N} x_n^2$ を計算する。$\ln(a \cdot b) = \ln a + \ln b$ という性質を総乗の形で一般化すると、

$$\ln \prod_{n=1}^{N} a_n = \sum_{n=1}^{N} \ln a_n$$

となるので、$\ln \prod_{n=1}^{N} x_n^2 = \sum_{n=1}^{N} \ln x_n^2 = \sum_{n=1}^{N} 2 \ln x_n$ となる。

対数関数と argmax

実関数 $f(x)$ において、$\operatorname{argmax}_x f(x)$ と $\operatorname{argmax}_x \ln f(x)$ の値が同じになることを示す。

実関数 $f(x)$ が $x = a$ において最大になるとする。つまり、a の任意の前後の値、$b < a < c$ において、$f(b) < f(a)$、$f(c) < f(a)$ となる。さて、対数関数は単調増加性を持つので、$x < y \Rightarrow \log x < \log y$ となる。これらを組み合わせると、$f(b) < f(a) \Rightarrow \ln f(b) < \ln f(a)$、$f(c) < f(a) \Rightarrow \ln f(c) < \ln f(a)$ となる。つまり、対数関数が適用されても、$f(x)$ を最大化する値 (ここでは a) は変わらない、ということである。よって、$\operatorname{argmax}_x f(x) = \operatorname{argmax}_x \ln f(x)$ が示された。

対数関数とガウス分布

$N(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2 \right)$ に対して対数関数を適用する。

$$
\begin{aligned}
&\ln N(x|\mu, \sigma^2) \\
&\quad = \ln \left(\frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2 \right) \right) \\
&\quad = \ln \frac{1}{(2\pi\sigma^2)^{1/2}} + \ln \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2 \right) \quad (\because \log_c(a \cdot b) = \log_c a + \log_c b) \\
&\quad = \ln(2\pi\sigma^2)^{-1/2} + \ln \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2 \right) \\
&\quad = -\frac{1}{2} \ln(2\pi\sigma^2) + \ln \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2 \right) \\
&\quad = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^2 + \ln \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2 \right) \\
&\quad = -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^2 - \frac{1}{2\sigma^2}(x-\mu)^2 \quad (\because \ln \exp x = \ln e^x = x)
\end{aligned}
$$

最尤推定

$$L(\mu, \sigma^2) = -\frac{N}{2} \ln 2\pi - \frac{N}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (x_i - \mu)^2$$

とし、$\frac{\partial L}{\partial \mu} = 0, \frac{\partial L}{\partial \sigma^2} = 0$ を解くと、

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i, \quad \sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$$

となることを示す。

$$
\begin{aligned}
&\frac{\partial L}{\partial \mu} \\
&\quad = \frac{d}{d\mu}\Bigg(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2\Bigg) \\
&\quad = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}\frac{d(x_i-\mu)^2}{d\mu} \quad (\because \text{微分の線形性}) \\
&\quad = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}-2(x_i-\mu) \\
&\quad = \frac{1}{\sigma^2}\sum_{i=1}^{N}x_i-\mu
\end{aligned}
$$

$\frac{\partial L}{\partial \mu}=0$ を解くと、

$$
\begin{aligned}
&\frac{1}{\sigma^2}\sum_{i\,=1}^{N}x_i-\mu=0 \\
&\Leftrightarrow \sum_{i=1}^{N}x_i=\sum_{i=1}^{N}\mu \\
&\Leftrightarrow \sum_{i=1}^{N}x_i=N\mu \\
&\Leftrightarrow \mu=\frac{1}{N}\sum_{i=1}^{N}x_i
\end{aligned}
$$

同様にして、$\frac{\partial L}{\partial \sigma^2}$ を解く。

$$
\begin{aligned}
&\frac{\partial L}{\partial \sigma^2} \\
&\quad = \frac{d}{d\sigma^2}(-\frac{N}{2}\ln\sigma^2)+\frac{d}{d\sigma^2}(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2) \\
&\quad = -\frac{N}{2}\frac{1}{\sigma^2}+\frac{d}{d\sigma^2}(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2) \quad (\because \frac{d\ln x}{dx}=\frac{1}{x}) \\
&\quad = -\frac{N}{2}\frac{1}{\sigma^2}-\frac{1}{2}\sum_{i=1}^{N}(x_i-\mu)^2\frac{d}{d\sigma^2}\frac{1}{\sigma^2} \quad (\because \text{微分の線形性}) \\
&\quad = -\frac{N}{2}\frac{1}{\sigma^2}+\frac{1}{2}\frac{1}{\sigma^4}\sum_{i=1}^{N}(x_i-\mu)^2
\end{aligned}
$$

$\frac{\partial L}{\partial \sigma^2}=0$ を解く。

$$
\begin{aligned}
&-\frac{N}{2}\frac{1}{\sigma^2}+\frac{1}{2}\frac{1}{\sigma^4}\sum_{i\ =1}^{N}(x_i-\mu)^2=0\\
&\quad\Leftrightarrow \frac{N}{2}\frac{1}{\sigma^2}=\frac{1}{2}\frac{1}{\sigma^4}\sum_{i=1}^{N}(x_i-\mu)^2\\
&\quad\Leftrightarrow N\sigma^2=\sum_{i=1}^{N}(x_i-\mu)^2\\
&\quad\Leftrightarrow \sigma^2=\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2
\end{aligned}
$$

β についての最尤推定

$\ln p(\mathbf{t}|\mathbf{x},\mathbf{w},\beta)=-\frac{N}{2}\ln(2\pi)+\frac{N}{2}\ln\beta-\frac{\beta}{2}\sum_{n=1}^{N}(t_n-y)^2$ とし、$\frac{\partial\ln p(\mathbf{t}|\mathbf{x},\mathbf{w},\beta)}{\partial\beta}(\beta^\star)=0$ となる $\beta^\star$ を求める。

$$
\begin{aligned}
&\frac{\partial\ln p(\mathbf{t}|\mathbf{x},\mathbf{w},\beta)}{\partial\beta}(\beta^\star)\\
&\quad=\frac{\partial}{\partial\beta}\left(\frac{N}{2}\ln\beta-\frac{\beta}{2}\sum_{n=1}^{N}(t_n-y)^2\right)\\
&\quad=\frac{N}{2\beta}-\frac{1}{2}\sum_{n=1}^{N}(t_n-y)^2
\end{aligned}
$$

$\frac{\partial\ln p(\mathbf{t}|\mathbf{x},\mathbf{w},\beta)}{\partial\beta}=0$ を解くと、

$$
\begin{aligned}
&\frac{N}{2\beta}-\frac{1}{2}\sum_{n\ =1}^{N}(t_n-y)^2=0\\
&\quad\Leftrightarrow \frac{N}{2\beta}=\frac{1}{2}\sum_{n=1}^{N}(t_n-y)^2\\
&\quad\Leftrightarrow \frac{1}{\beta}=\frac{1}{N}\sum_{n=1}^{N}(t_n-y)^2
\end{aligned}
$$

Lecture 4の回答

スカラ関数のベクトル微分

y が $\mathbf{w}$ の関数のとき、$\frac{\partial}{\partial \mathbf{w}} y^2 = 2\left(\frac{\partial y}{\partial \mathbf{w}}\right) y$ となることを示す。ただし、$\mathbf{w} = (w_1, w_2)^T$ とする。

左辺は、

$$\frac{\partial}{\partial \mathbf{w}} y^2 = \begin{pmatrix} \frac{\partial y^2}{\partial w_1} \\ \frac{\partial y^2}{\partial w_2} \end{pmatrix} = \begin{pmatrix} 2y\frac{\partial y}{\partial w_1} \\ 2y\frac{\partial y}{\partial w_2} \end{pmatrix}$$

右辺は、

$$2\left(\frac{\partial y}{\partial \mathbf{w}}\right) y = 2\begin{pmatrix} \frac{\partial y}{\partial w_1} \\ \frac{\partial y}{\partial w_2} \end{pmatrix} y = \begin{pmatrix} 2y\frac{\partial y}{\partial w_1} \\ 2y\frac{\partial y}{\partial w_2} \end{pmatrix}$$

よって、両辺は等しい。同じようにして、$\mathbf{w}$ が N 次元ベクトル場合の場合も示せる。

Lecture 5の回答

対数関数と指数関数

$$\ln \frac{\exp\Big(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)\Big)p(C_1)}{\exp\Big(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)\Big)p(C_2)} = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1) + \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2) + \ln\frac{p(C_1)}{p(C_2)}$$

となることを示す。$x = \ln(\exp(x)) = \exp(\ln x)$ より、

$$\begin{aligned}
&\ln \frac{\exp\Big(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)\Big)p(C_1)}{\exp\Big(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)\Big)p(C_2)} \\
&= \ln\left(\exp\Big(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)\Big)p(C_1)\right)\left(\exp\Big(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)\Big)p(C_2)\right)^{-1} \\
&= \ln\left(\exp\Big(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)\Big)p(C_1)\right) - \ln\left(\exp\Big(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)\Big)p(C_2)\right) \\
&= -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1) + \ln p(C_1) + \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2) - \ln p(C_2) \\
&= -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1) + \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2) + \ln\frac{p(C_1)}{p(C_2)}
\end{aligned}$$

線形回帰モデルへの変形

$$a = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1) + \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2) + \ln\frac{p(C_1)}{p(C_2)}$$

の時、

$$\mathbf{w} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$$

$$w_0 = -\frac{1}{2}\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_2^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2 + \ln\frac{p(C_1)}{p(C_2)}$$

とおくと、

$$a = \mathbf{w}^T\mathbf{x} + w_0$$

と表記できることを示す。

$$\begin{aligned}
a &= -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_1) + \frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}_2) + \ln\frac{p(C_1)}{p(C_2)} \\
&= -\frac{1}{2}(\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x} - \mathbf{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_1^T\boldsymbol{\Sigma}^{-1}\mathbf{x} + \boldsymbol{\mu}_1^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1) + \\
&\quad -\frac{1}{2}(\mathbf{x}^T\boldsymbol{\Sigma}^{-1}\mathbf{x} - \mathbf{x}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2 - \boldsymbol{\mu}_2^T\boldsymbol{\Sigma}^{-1}\mathbf{x} + \boldsymbol{\mu}_2^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_2) + \ln\frac{p(C_1)}{p(C_2)}
\end{aligned}$$

ここで、$\mathbf{x}^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_1$ はスカラなので転置をかけても等しくなるので、$(\mathbf{x}^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_1)^T = \boldsymbol{\mu}_1^T\mathbf{\Sigma}^{-1T}\mathbf{x} = \boldsymbol{\mu}_1^T\mathbf{\Sigma}^{-1}\mathbf{x}$ となる。また、$(ABC)^T = C^TB^TA^T$ であることと、Σ は対称行列で、対称行列の逆行列もまた対称行列となることから、$\mathbf{\Sigma}^{-1T} = \mathbf{\Sigma}^{-1}$ を使った。これらを使い、

$$\begin{aligned}
a &= -\frac{1}{2}\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x} + \boldsymbol{\mu}_1^T\mathbf{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_1^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_1 \\
&\quad + \frac{1}{2}\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x} - \boldsymbol{\mu}_2^T\mathbf{\Sigma}^{-1}\mathbf{x} + \frac{1}{2}\boldsymbol{\mu}_2^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_2 + \ln\frac{p(C_1)}{p(C_2)} \\
&= \boldsymbol{\mu}_1^T\mathbf{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_1^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2^T\mathbf{\Sigma}^{-1}\mathbf{x} + \frac{1}{2}\boldsymbol{\mu}_2^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_2 + \ln\frac{p(C_1)}{p(C_2)}
\end{aligned}$$

ここで、$\mathbf{w}^T\mathbf{x} + w_0$ を計算すると、

$$\begin{aligned}
\mathbf{w}^T\mathbf{x} + w_0 &= (\boldsymbol{\mu}_1^T - \boldsymbol{\mu}_2^T)\mathbf{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_1^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_2^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_2 + \ln\frac{p(C_1)}{p(C_2)} \\
&= \boldsymbol{\mu}_1^T\mathbf{\Sigma}^{-1}\mathbf{x} - \boldsymbol{\mu}_2^T\mathbf{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}\boldsymbol{\mu}_1^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_1 + \frac{1}{2}\boldsymbol{\mu}_2^T\mathbf{\Sigma}^{-1}\boldsymbol{\mu}_2 + \ln\frac{p(C_1)}{p(C_2)}
\end{aligned}$$

となるので、$a = \mathbf{w}^T\mathbf{x} + w_0$ が示せた。

シグモイド関数の微分

$\sigma(x) = \frac{1}{1+\exp(-x)}$ の時、$\frac{d\sigma(x)}{dx} = \sigma(x)(1-\sigma(x))$ を示そう。

$$\begin{aligned}
\frac{d\sigma(x)}{dx} &= \frac{d(1+e^{-x})^{-1}}{dx} \\
&= -(1+e^{-x})^{-2}(-1)e^{-x} \quad (\because \frac{de^{-x}}{dx} = -e^{-x}) \\
&= (1+e^{-x})^{-2}e^{-x}
\end{aligned}$$

ここで、右辺を計算すると、

$$\begin{aligned}
\sigma(x)(1-\sigma(x)) &= (1+e^{-x})^{-1}\frac{1+e^{-x}-1}{1+e^{-x}} \\
&= (1+e^{-x})^{-2}e^{-x}
\end{aligned}$$

よって、$\frac{d\sigma(x)}{dx} = \sigma(x)(1-\sigma(x))$ を示した。

交差エントロピー誤差関数の微分

$\frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} = \sum_{n=1}^{N}(y_n - t_n)\phi_n$ を示そう。

$$\begin{aligned}
\frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} &= -\sum_{n=1}^{N}\frac{\partial}{\partial \mathbf{w}}(t_n \ln y_n + (1-t_n)\ln(1-y_n)) \\
&= -\sum_{n=1}^{N}(t_n\frac{\partial}{\partial \mathbf{w}}\ln y_n + (1-t_n)\frac{\partial}{\partial \mathbf{w}}\ln(1-y_n))
\end{aligned}$$

まずは $\frac{\partial}{\partial \mathbf{w}}\ln y_n$ を計算しよう。

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{w}} \ln y_n &= \frac{1}{y_n} \frac{\partial}{\partial \mathbf{w}} y_n \quad (\because \frac{d \ln x}{dx} = \frac{1}{x}) \\
&= \frac{1}{\sigma(\mathbf{w}^T \boldsymbol{\phi}_n)} \frac{\partial}{\partial \mathbf{w}} \sigma(\mathbf{w}^T \boldsymbol{\phi}_n) \\
&= \frac{1}{\sigma(\mathbf{w}^T \boldsymbol{\phi}_n)} \sigma(\mathbf{w}^T \boldsymbol{\phi}_n)(1 - \sigma(\mathbf{w}^T \boldsymbol{\phi}_n)) \boldsymbol{\phi}_n \quad (\because \frac{d\sigma(x)}{dx} = \sigma(x)(1 - \sigma(x))) \\
&= (1 - y_n) \boldsymbol{\phi}_n
\end{aligned}
$$

次に、$\frac{\partial}{\partial \mathbf{w}} \ln(1 - y_n)$ を計算しよう。

$$
\begin{aligned}
\frac{\partial}{\partial \mathbf{w}} \ln(1 - y_n) &= \frac{1}{1 - y_n} \frac{\partial}{\partial \mathbf{w}} (1 - y_n) \\
&= \frac{1}{1 - y_n} \cdot - \frac{\partial}{\partial \mathbf{w}} y_n \\
&= \frac{1}{1 - y_n} \cdot - y_n (1 - y_n) \boldsymbol{\phi}_n \\
&= - y_n \boldsymbol{\phi}_n
\end{aligned}
$$

これにより、

$$
\begin{aligned}
\frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} &= - \sum_{n=1}^{N} (t_n (1 - t_n) \boldsymbol{\phi}_n + (1 - t_n)(- y_n \boldsymbol{\phi}_n)) \\
&= - \sum_{n=1}^{N} (t_n - y_n) \boldsymbol{\phi}_n \\
&= \sum_{n=1}^{N} (y_n - t_n) \boldsymbol{\phi}_n
\end{aligned}
$$

と示せた。

意地悪な問題

$\frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} = \sum_{n=1}^{N} (y_n - t_n) \phi_n = 0$ となるような $\mathbf{w}$ を解析的に導出してみよう。

$$
\begin{aligned}
\sum_{n\ =1}^{N} t_n \phi_n &= \sum_{n=1}^{N} y_n \phi_n \\
\sum_{n=1}^{N} t_n \phi_n &= \sum_{n=1}^{N} \sigma(\mathbf{w}^T \boldsymbol{\phi}_n) \phi_n
\end{aligned}
$$

さて、線形回帰の時には、

$$
\sum_{n=1}^{N} t_n \boldsymbol{\phi}_n^\top = \mathbf{w}^\top \sum_{n=1}^{N} \boldsymbol{\phi}_n \boldsymbol{\phi}_n^\top \tag{1}
$$

となり、$\Phi = \begin{pmatrix} \boldsymbol{\phi}(x_1) \\ \boldsymbol{\phi}(x_2) \\ \cdots \\ \boldsymbol{\phi}(x_N) \end{pmatrix}$ とおくことによって、$\Phi^\top \mathbf{t} = \mathbf{w}^\top \Phi^\top \Phi$ と変形できた。しかし、今回は非線形の関数 σ があるために $\mathbf{w}$ を Σ の外側に移動できない！(非線形関数は行列では表現できないため。)

線形識別モデルのヘッセ行列

$\nabla E(\mathbf{w}) = \sum_{n=1}^{N}(y_n - t_n)\boldsymbol{\phi}_n$ の時に $\mathbf{H} = \nabla\nabla E(\mathbf{w}) = \sum_{n=1}^{N} y_n(1-y_n)\boldsymbol{\phi}_n\boldsymbol{\phi}_n^\top$ となることを示そう。ただし、$\frac{\partial x\mathbf{a}}{\partial\mathbf{w}} = \mathbf{a}^T\frac{\partial x}{\partial\mathbf{w}}$ （x は $\mathbf{w}$ の関数)。

$y_n' = y_n(1-y_n)\boldsymbol{\phi}_n$ より、$\mathbf{H} = \sum\frac{\partial y_n\boldsymbol{\phi}_n}{\partial\mathbf{w}} = \sum\boldsymbol{\phi}_n^T y_n(1-y_n)\boldsymbol{\phi}_n$

softmax 関数の微分

$y_k(\boldsymbol{\phi}) = p(C_k|\boldsymbol{\phi}) = \frac{\exp(a_k)}{\sum_j\exp(a_j)}$ $(a_k = \mathbf{w}_k^T\boldsymbol{\phi})$ に対し、$\frac{\partial y_k}{\partial a_j} = y_k(I_{kj} - y_j)$ を示そう。

$$\frac{\partial}{\partial a_k}y_k = \frac{\exp(a_k)(\sum\exp(a_j)) - \exp(a_k)\exp(a_k)}{(\sum\exp(a_j))^2} = y_k - y_k^2 \tag{2}$$

$k \neq j$ として

$$\frac{\partial}{\partial a_j}y_k = -\frac{\exp(a_k)\exp(a_j)}{(\sum\exp(a_j))^2} = y_k y_j \tag{3}$$

よってこの二つをまとめて、$\frac{\partial y_k}{\partial a_j} = y_k(I_{kj} - y_j)$