二次形式の微分

2019 年 6 月 28 日

$n \in \mathbb{N}, 1 \leq n < \infty$ として n 次元実ベクトル $\boldsymbol{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$ と $n \times n$ 実行列 $A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}$ に関する二次形式 $\boldsymbol{w}^\mathsf{T} A \boldsymbol{w}$ の $\boldsymbol{w}$ に関する微分が $\boldsymbol{w}^\mathsf{T}(A+A^\mathsf{T})$ となることを示します。

まず各要素 w_k ごとに微分する。

$$
\begin{aligned}
\frac{\partial}{\partial w_k} \boldsymbol{w}^\mathsf{T} A \boldsymbol{w} &= \frac{\partial}{\partial w_k} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} w_i w_j \right) && \text{二次形式を展開} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial}{\partial w_k} (a_{ij} w_i w_j) && (f+g)' = f' + g' \text{ の一般化} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{\partial}{\partial w_k} a_{ij} w_i \right) w_j + a_{ij} w_i \frac{\partial w_j}{\partial w_k} && (fg)' = f'g + fg' \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{\partial}{\partial w_k} a_{ij} w_i \right) w_j + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} w_i \frac{\partial w_j}{\partial w_k} && \text{ばらす} \\
&= \sum_{j=1}^{n} a_{kj} w_j + \sum_{i=1}^{n} a_{ik} w_i && \tfrac{\partial w_i}{\partial w_k} = \begin{cases} 1 & (k = i) \\ 0 & (k \neq i) \end{cases} \\
&= \sum_{j=1}^{n} a_{kj} w_j + \sum_{j=1}^{n} a_{jk} w_j && \text{添字 } i \to j \text{ の付け替え} \\
&= \sum_{j=1}^{n} a_{kj} w_j + a_{jk} w_j \\
&= \sum_{j=1}^{n} (a_{kj} + a_{jk}) w_j
\end{aligned}
$$

チュートリアルに従い、ベクトルによる微分は、要素ごとの微分を横に並べた行ベクトルとすると、

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{w}} \boldsymbol{w}^\top A \boldsymbol{w} &= \begin{bmatrix} \frac{\partial}{\partial w_1} \boldsymbol{w}^\top A \boldsymbol{w} & \dots & \frac{\partial}{\partial w_n} \boldsymbol{w}^\top A \boldsymbol{w} \end{bmatrix} \\
&= \begin{bmatrix} \sum_{j=1}^{n} (a_{1j} + a_{j1}) w_j & \dots & \sum_{j=1}^{n} (a_{nj} + a_{jn}) w_j \end{bmatrix} \\
&= \begin{bmatrix} \begin{bmatrix} w_1 & \dots & w_n \end{bmatrix} \begin{bmatrix} a_{11} + a_{11} \\ \vdots \\ a_{1n} + a_{n1} \end{bmatrix} & \dots & \begin{bmatrix} w_1 & \dots & w_n \end{bmatrix} \begin{bmatrix} a_{n1} + a_{1n} \\ \vdots \\ a_{nn} + a_{nn} \end{bmatrix} \end{bmatrix} \\
&= \begin{bmatrix} w_1 & \dots & w_n \end{bmatrix} \begin{bmatrix} a_{11} + a_{11} & a_{21} + a_{12} & \dots & a_{n1} + a_{1n} \\ a_{12} + a_{21} & a_{22} + a_{22} & \dots & a_{n2} + a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} + a_{n1} & a_{2n} + a_{n2} & \dots & a_{nn} + a_{nn} \end{bmatrix} \\
&= \boldsymbol{w}^\top (A + A^\top)
\end{aligned}
$$

おまけ

ベクトルによる微分は縦？横？

実ベクトルをとって実数を返す関数 $f(\boldsymbol{x})$ について、$\frac{\partial f}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$ とするか $\frac{\partial f}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \dots & \frac{\partial f}{\partial x_n} \end{bmatrix}$ と定義するかは場合（や分野など）によるようです。Chainer チュートリアルでは後者を採用しています。なぜそうしているかというと、後者のほうが、ベクトル値関数の Jacobi 行列と同じ形になるので、統一的に扱えるからのようです。ベクトル値関数 $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ $\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ \vdots \\ f_m(\boldsymbol{x}) \end{bmatrix}$ の Jacobi 行列は $m \times n$ 行列

$$
\mathcal{J}\boldsymbol{f} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}
$$

で定義されます。実数値関数 f を特別なベクトル値関数 $f : \mathbb{R}^n \to \mathbb{R}^1$ として見ると、

$$
\mathcal{J}f = \frac{\partial f}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \dots & \frac{\partial f}{\partial x_n} \end{bmatrix}
$$

と $1 \times n$ 行列のように扱えます。

特に `numpy` や `Chainer` で行列計算するプログラムを実装するときに統一的に扱えるほうが楽で嬉しい。

前回の失敗について

前回は、積の微分公式を早い段階で適応してしまったのが問題でした。

$$\frac{\partial}{\partial w_k}\boldsymbol{w}^\top A\boldsymbol{w} = \left(\frac{\partial}{\partial w_k}\boldsymbol{w}^\top\right)A\boldsymbol{w} + \boldsymbol{w}^\top\left(\frac{\partial}{\partial w_k}A\boldsymbol{w}\right)$$

ここで、$\frac{\partial}{\partial w_k}\boldsymbol{w}^\top$ってどうなるんだろう？となって躓いてしまいます。上と照らしてみると、$\frac{\partial}{\partial w_k}\boldsymbol{w}^\top = \begin{bmatrix}\frac{\partial w_1}{\partial w_k} & \dots & \frac{\partial w_n}{\partial w_k}\end{bmatrix}$と行ベクトルで定義すればいいことに気づきます（またいろいろ整合がつく）。

$A\boldsymbol{w} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn}\end{bmatrix}\begin{bmatrix} w_1 \\ \vdots \\ w_n\end{bmatrix} = \begin{bmatrix}\sum_{i=1}^n a_{1i}w_i \\ \vdots \\ \sum_{i=1}^n a_{ni}w_i\end{bmatrix}$ に注意すれば、

$$\begin{aligned}
\frac{\partial}{\partial w_k}\boldsymbol{w}^\top A\boldsymbol{w} &= \left(\frac{\partial}{\partial w_k}\boldsymbol{w}^\top\right)A\boldsymbol{w} + \boldsymbol{w}^\top\left(\frac{\partial}{\partial w_k}A\boldsymbol{w}\right) \\
&= \begin{bmatrix}\frac{\partial w_1}{\partial w_k} & \dots & \frac{\partial w_k}{\partial w_k} & \dots & \frac{\partial w_n}{\partial w_k}\end{bmatrix}A\boldsymbol{w} + \boldsymbol{w}^\top\left(\frac{\partial}{\partial w_k}\begin{bmatrix}\sum_{i=1}^n a_{1i}w_i \\ \vdots \\ \sum_{i=1}^n a_{ni}w_i\end{bmatrix}\right) \\
&= \underbrace{\begin{bmatrix}0 & \dots & 1 & \dots & 0\end{bmatrix}}_{\text{k 番目だけ 1}}\begin{bmatrix}\sum_{i=1}^n a_{1i}w_i \\ \vdots \\ \sum_{i=1}^n a_{ni}w_i\end{bmatrix} + \boldsymbol{w}^\top\begin{bmatrix}\frac{\partial}{\partial w_k}\left(\sum_{i=1}^n a_{1i}w_i\right) \\ \vdots \\ \frac{\partial}{\partial w_k}\left(\sum_{i=1}^n a_{ni}w_i\right)\end{bmatrix} \\
&= \sum_{i=1}^n a_{ki}w_i + \boldsymbol{w}^\top\begin{bmatrix}a_{1k} \\ \vdots \\ a_{nk}\end{bmatrix} \\
&= \sum_{i=1}^n a_{ki}w_i + \sum_{i=1}^n a_{ik}w_i \\
&= \sum_{i=1}^n (a_{ki} + a_{ik})w_i \qquad \text{（あとはいっしょ）}
\end{aligned}$$