

0 定義、公式、約束事

0.1 アインシュタインの縮約

本ノートではアインシュタインの縮約記法を用いる。

すなわち、成分を表す同じ添字がついた文字の積は、その添字について和をとっているものとみなす。

例を挙げると、 $\boldsymbol{x} = (x_1, x_2, x_3)$ というベクトルと $\boldsymbol{y} = (y_1, y_2, y_3)$ というベクトルの内積は以下のようになる。

$$\boldsymbol{x} \cdot \boldsymbol{y} = x_i y_i = \sum_{i=1}^3 x_i y_i \quad (1)$$

記法になれるまではスムーズに読めないかもしれないが、この記法を使わないと行列やベクトルの積を成分表示したときに和の記号がたくさん出てきてしまい、かえって混乱を招くように思う。

また、この記法を使うと、あえて和をとらない $x_i y_i$ がほしいときどうするのか疑問に思うかもしれない。その場合はクロネッカーのデルタを用いて $\delta_{ij} x_j y_j$ などとすればよい。

0.2 行列・ベクトルを含む式の微分

行列やベクトルを含む式でも、スカラーの場合のような微分に関する公式が成り立つことを確認しておく。

なお、 $\frac{\partial \boldsymbol{u}}{\partial x_i}$ は、ベクトル \boldsymbol{u} の各成分に関して x_i で微分したベクトルを表しており、 $\frac{\partial f(\boldsymbol{x})}{\partial x_i}$ は、ベクトル \boldsymbol{x} に依存するスカラー関数 f をベクトル \boldsymbol{x} の i 番目の成分で微分したものを i 番目の成分に持つベクトル $(\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \frac{\partial f(\boldsymbol{x})}{\partial x_2}, \dots)$ を表している

0.2.1 定数行列がかかったベクトルの微分

ベクトル $\boldsymbol{x} (\in \mathbb{R}^n)$ に依存するベクトル値関数 $\boldsymbol{u}(\boldsymbol{x})$ と定数行列 $A (\in \mathbb{R}^{n \times n})$ に関して以下がなりたつ

$$\frac{\partial}{\partial x_i} A \boldsymbol{u} = A \frac{\partial \boldsymbol{u}}{\partial x_i} \quad (2)$$

実際に成分で計算してみると

$$\frac{\partial}{\partial x_i} A \boldsymbol{u} = \frac{\partial}{\partial x_i} A_{\alpha\beta} u_\beta \quad (3)$$

$$= A_{\alpha\beta} \frac{\partial u_\beta}{\partial x_i} \quad (A_{\alpha\beta} \text{ はスカラーの定数なので微分の影響を受けない}) \quad (4)$$

$$= A \frac{\partial \boldsymbol{u}}{\partial x_i} \quad (\text{行列とベクトルの積の定義より}) \quad (5)$$

0.2.2 積の微分

ベクトル $\mathbf{x}(\in \mathbb{R}^n)$ に依存する 2 つのベクトル値関数 $\mathbf{u}(\mathbf{x}), \mathbf{v}(\mathbf{x})(\in \mathbb{R}^n)$ に関して、 \mathbf{x} の第 i 成分を x_i 、として以下が成り立つ。

$$\frac{\partial}{\partial x_i} \mathbf{u}^T \cdot \mathbf{v} = \left(\frac{\partial \mathbf{u}}{\partial x_i} \right)^T \cdot \mathbf{v} + \mathbf{u} \cdot \left(\frac{\partial \mathbf{v}}{\partial x_i} \right) \quad (6)$$

実際に成分ごとに計算して確認すると

$$\frac{\partial}{\partial x_i} \mathbf{u}^T \mathbf{v} = \frac{\partial}{\partial x_i} (u_\alpha v_\alpha) \quad (7)$$

$$= \frac{\partial u_\alpha}{\partial x_i} v_\alpha + u_\alpha \frac{\partial v_\alpha}{\partial x_i} \quad (\text{スカラーに対する積の微分}) \quad (8)$$

$$= \left(\frac{\partial \mathbf{u}}{\partial x_i} \right)^T \cdot \mathbf{v} + \mathbf{u} \cdot \left(\frac{\partial \mathbf{v}}{\partial x_i} \right) \quad (9)$$

特別な場合として、定数行列 $A(\in \mathbb{R}^{n \times n})$ に関して、行列 A の i, j 成分を A_{ij} として以下が成り立つ。

$$\frac{\partial}{\partial x_i} \mathbf{u}^T A \mathbf{v} = \left(\frac{\partial \mathbf{u}}{\partial x_i} \right)^T A \mathbf{v} + \mathbf{u}^T A \left(\frac{\partial \mathbf{v}}{\partial x_i} \right) \quad (10)$$

定数行列がかかったベクトルの微分と積の微分を両方使うことで確認できる。

0.3 多変数の平方完成

$\mathbf{x}, \mathbf{a}, \mathbf{b}$ を n 成分のベクトル、 A, B を $n \times n$ 成分の行列として、以下のような多項式を平方完成しなくてはならないとき、うんざりした気持ちになったりしないだろうか。

$$(\mathbf{x} - \mathbf{a})^T A (\mathbf{x} - \mathbf{a}) + (\mathbf{x} - \mathbf{b})^T B (\mathbf{x} - \mathbf{b}) \quad (11)$$

愚直に展開しても当然平方完成できる。ただ項が増えると計算ミスしやすくなる。そこで、平方完成という操作を多項式の項の整理ではなく、1 次の微分係数が消えている点の周りでのテイラー展開を求める操作と考えると計算が簡単になることがある。一般に適当な滑らかさを持つ、ベクトル変数スカラー値の関数 $f(\mathbf{x})$ の、ある点 \mathbf{a} のまわりでの、2 次で打ち切ったテイラー展開は以下の通りである。

$$f(\mathbf{x}) = f(\mathbf{a}) + \frac{\partial f(\mathbf{a})}{\partial x_i} (x_i - a_i) + \frac{1}{2} \frac{\partial^2 f(\mathbf{a})}{\partial x_i \partial x_j} (x_i - a_i)(x_j - a_j) \quad (12)$$

平方完成が目的なのだから、 f は x_i の 2 次式であるはずなので、テイラー展開の 3 次以上の項はない。平方完成のアルゴリズムは以下の通りである。

(1) まず、 $\frac{\partial f(\mathbf{a})}{\partial x_i} = 0$ を満たす点 $\hat{\mathbf{a}}$ を求める

(2) $\frac{\partial^2 f(\hat{\mathbf{a}})}{\partial x_i \partial x_j}$ を計算する

そうすれば、テイラー展開のうち \mathbf{x} の 0 次の項と 2 次の項のみが残る形に変形できる。

$$f(\boldsymbol{x}) = f(\hat{\boldsymbol{a}}) + \frac{1}{2}(\boldsymbol{x} - \hat{\boldsymbol{a}})^T \frac{\partial^2 f(\hat{\boldsymbol{a}})}{\partial \boldsymbol{x}^2} (\boldsymbol{x} - \hat{\boldsymbol{a}}) \quad (13)$$

$$= f(\hat{\boldsymbol{a}}) + \frac{1}{2}(\boldsymbol{x} - \hat{\boldsymbol{a}})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \hat{\boldsymbol{a}}) \quad (14)$$

最後の変形は多変量正規分布でよく使われる形に合わせて行った。なお、多変量正規分布の指数関数の肩に現れる場合には 0 次の項は正規化因子に含めることができるので、最終的な分布の計算を終えるまで計算を先送りできる。