生物資源学類 実用解析 I レポート課題, 2020/06/08 出題。奈佐原顕郎

以下, 行列 A の (i, j) 成分を $[A]_{ij}$ などと表す（基礎数学 II 教材で使った記法)。

問 1: 行列 A, B について,

(1-1) 積 AB を, 成分を使って定義せよ。すなわち, $[AB]_{ij}$ を, A の成分と B の成分の式であらわせ。

(1-2) 積 AB が定義できる条件を述べよ（ヒント: 行列の形)。

(1-3) A^{T} を, 成分を使って定義せよ（転置行列)。すなわち, $[A^{\mathrm{T}}]_{ij}$ を A の成分であらわせ。

(1-4) 2 つの行ベクトル $\mathbf{x} = (x_1, x_2, x_3)$, $\mathbf{w} = (w_1, w_2, w_3)$ を考える。$\mathbf{x}\mathbf{w}^{\mathrm{T}}$ と $\mathbf{x}^{\mathrm{T}}\mathbf{w}$ をそれぞれ成分で書き表せ。

(1-5) $\mathbf{x} = (x_1, x_2)$, $\mathbf{b} = (b_1, b_2, b_3)$ とし,

$$W = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{bmatrix}$$

とする。$\mathbf{a} = \mathbf{x}W + \mathbf{b}$ とする。$\mathbf{a}$ を成分で書き表せ。

(1-6) $\mathbf{a}$ の第 k 成分を a_k とする。$\partial a_2/\partial w_{12}$ を求めよ。

(1-7) $\partial a_2/\partial w_{23}$ を求めよ。

(1-8) $\partial a_2/\partial x_1$ を求めよ。

(1-9) $\partial a_2/\partial b_2$ を求めよ。

(1-10) $\partial a_2/\partial b_3$ を求めよ。

(1-11) $\partial a_j/\partial x_i = w_{ij}$ であることを示せ。

(1-12) $\partial a_k/\partial w_{ij} = \delta_{kj} x_i$ であることを示せ。ただし, δ_{kj} とは, 2 つの整数 k, j について, $k = j$ のときは 1, $k \neq j$ のときは 0 をとるような記号と定義する。

(1-13) 関数 L を, ベクトル $\mathbf{a}$ の（全成分の）関数とする。次式を示せ (ヒント: 連鎖律):

$$\frac{\partial L}{\partial x_1} = \frac{\partial L}{\partial a_1}\frac{\partial a_1}{\partial x_1} + \frac{\partial L}{\partial a_2}\frac{\partial a_2}{\partial x_1} + \frac{\partial L}{\partial a_3}\frac{\partial a_3}{\partial x_1}, \quad \frac{\partial L}{\partial x_2} = \frac{\partial L}{\partial a_1}\frac{\partial a_1}{\partial x_2} + \frac{\partial L}{\partial a_2}\frac{\partial a_2}{\partial x_2} + \frac{\partial L}{\partial a_3}\frac{\partial a_3}{\partial x_2}$$

(1-14) 次式を示せ (ヒント: 問 (1-11)):

$$\frac{\partial L}{\partial x_1} = \frac{\partial L}{\partial a_1}w_{11} + \frac{\partial L}{\partial a_2}w_{12} + \frac{\partial L}{\partial a_3}w_{13}, \quad \frac{\partial L}{\partial x_2} = \frac{\partial L}{\partial a_1}w_{21} + \frac{\partial L}{\partial a_2}w_{22} + \frac{\partial L}{\partial a_3}w_{23}$$

(1-15) 次式を示せ（ただし, $\partial L/\partial \mathbf{x} = (\partial L/\partial x_1, \partial L/\partial x_2)$, $\partial L/\partial \mathbf{a} = (\partial L/\partial a_1, \partial L/\partial a_2, \partial L/\partial a_3)$ とする）:

$$\frac{\partial L}{\partial \mathbf{x}} = \frac{\partial L}{\partial \mathbf{a}} W^{\mathrm{T}}$$

これはテキスト P148 の式 (5.13) の最初の式の, X がベクトルの場合である（テキストの Y が $\mathbf{a}$ に相当)。

(1-16) 次式を示せ (ヒント: 連鎖律):

$$\frac{\partial L}{\partial w_{ij}} = \frac{\partial L}{\partial a_1}\frac{\partial a_1}{\partial w_{ij}} + \frac{\partial L}{\partial a_2}\frac{\partial a_2}{\partial w_{ij}} + \frac{\partial L}{\partial a_3}\frac{\partial a_3}{\partial w_{ij}}$$

(1-17) 次式を示せ (ヒント: 問 (1-12)):

$$\frac{\partial L}{\partial w_{ij}} = \frac{\partial L}{\partial a_j} x_i$$

(1-18) 次式を示せ (ただし, $\partial L/\partial W$ は, (i, j) 成分が $\partial L/\partial w_{ij}$ であるような行列):

$$\frac{\partial L}{\partial W} = \mathbf{x}^{\mathrm{T}} \frac{\partial L}{\partial \mathbf{a}}$$

これはテキスト P148 の式 (5.13) の 2 番めの式で, X がベクトルの場合である。

問 2: ニューラルネットのひとつの層（Affine レイヤー）を行列表記で実装することを考えよう。ある層への入

力が $n \times m$ 行列（つまり n 行 m 列の行列）である X, 出力が $n \times l$ 行列である A で表されるとする。この層の挙動は, 適当な行列 W, B を使って, $A = XW + B$ と表されるとする（この表記は P61 の式 (3.9) と同様とする。ただし行列の形やサイズは P61 と同じではない)。

(2-1) 行列 W と行列 B は, それぞれ何行何列か?

(2-2) この層を構成するニューロンは何個か? ヒント: n, m, l のうちどれかだろう。

(2-3) ひとつのニューロンに入ってくる「神経」は何本あるか?

(2-4) B を構成する行ベクトルどうしは互いにおなじである。なぜか?

ヒント: $A = XW + B$ は, 問 1 の $\mathbf{a} = \mathbf{x}W + \mathbf{b}$ の拡張である。入力を表す行ベクトル $\mathbf{x}$ が複数あるとき（つまり複数のデータがあるとき), バッチ処理するために, それらを縦に並べて作ったのが行列 X である。

問 3: 上の問 2 の状況で, 損失関数 L を考える。L は上で定義した行列 A の（各成分の）関数とみなせる。そこで, L を A の各成分で偏微分してできる行列を $\partial L/\partial A$ と書く。つまり, $\partial L/\partial A$ の (i,j) 成分は, $\partial L/\partial a_{ij}$ である (a_{ij} は A の (i,j) 成分)。

(3-1) X の (i,j) 成分を x_{ij} とする。a_{ip} は, $p \neq j$ のときは x_{ij} に依存しないことを示せ。

(3-2) 次式を示せ (ヒント: 連鎖律):

$$\frac{\partial L}{\partial x_{ij}} = \sum_p \frac{\partial L}{\partial a_{ip}} \frac{\partial a_{ip}}{\partial x_{ij}}$$

(3-3) $\partial a_{ip}/\partial x_{ij} = w_{jp}$ であることを示せ。

(3-4) 次式を示せ（ヒント: 行列の積の定義（Σ を使う式）をリメディアル教材などで再確認せよ）:

$$\frac{\partial L}{\partial X} = \frac{\partial L}{\partial A} W^{\mathrm{T}}$$

これはテキスト P148 の式 (5.13) の最初の式である（テキストでは A が Y になっている)。

(3-5) 次式を示せ (ヒント: 連鎖律):

$$\frac{\partial L}{\partial w_{ij}} = \sum_q \frac{\partial L}{\partial a_{qj}} \frac{\partial a_{qj}}{\partial w_{ij}}$$

(3-6) $\partial a_{qj}/\partial w_{ij} = x_{qi}$ であることを示せ。

(3-7) 次式を示せ:

$$\frac{\partial L}{\partial W} = X^{\mathrm{T}} \frac{\partial L}{\partial A}$$

これはテキスト P148 の式 (5.13) の 2 番めの式である（テキストでは A が Y になっている)。

問 4: ニューラルネットワークの出力層と損失関数について考えよう。分類クラスが n 個あり, それに応じて出力層では n 個のニューロンからひとつずつ出力があり, それをまとめると $\mathbf{a} := (a_1, a_2, \cdots, a_n)$ となるとする（P67 の図 3-22 を拡張して考えよう)。これがソフトマックス関数によって, $\mathbf{y} := (y_1, y_2, \cdots, y_n)$ と変換される（その式は教科書にあるとおり, $y_k := \exp(a_k)/\{\sum_j \exp(a_j)\}$)。これと, one-hot 表現された正解データ $\mathbf{t} := (t_1, t_2, \cdots, t_n)$ を用いて, 損失関数 $L = -\sum_k t_k \ln y_k$ を求める（ミニバッチ処理では, この L を複数のデータについて求めてその平均を扱うのだが, ここでは単純化のためにデータは 1 個だけとする)。次式を示せ:

$$L = -\sum_k t_k (a_k - \ln \sum_j \exp(a_j)),$$

$$\frac{\partial L}{\partial a_l} = y_l - t_l, \quad \frac{\partial L}{\partial \mathbf{a}} = (y_1 - t_1, y_2 - t_2, \cdots, y_n - t_n)$$

これが P154 あたりで述べられている, Softmax-with-Loss レイヤからの”キレイ”な逆伝播である。