Affine レイヤ

1. 表記について

i) 行列 A に対し、その (i,j) 成分を $(A)_{ij}$ と書くことにする。
特に、行列が2文字以上で書かれる場合にこの記法を用いる。
例えば、行列 P と Q がこの順に乗算可能とするとき、積 PQ の (i,j) 成分は $(PQ)_{ij}$ となる。

ii) (i,j) 成分が a_{ij} である行列を (a_{ij}) と書く。i) と合わせると、
行列 A について $A=((A)_{ij})$。
また、i) の行列 P, Q を $P=(p_{ij})$, $Q=(q_{ij})$ とすると、

$$PQ=((PQ)_{ij})=\left(\sum_k p_{ik}q_{kj}\right)$$

iii) 行ベクトルは1行の行列とみなし、ii) と同様の記法を用いる。
すなわち、N 次元ベクトル $\mathbb{v}=(v_1, v_2, \cdots, v_N)$ について単純に $\mathbb{v}=(v_i)$ と書くこととする。

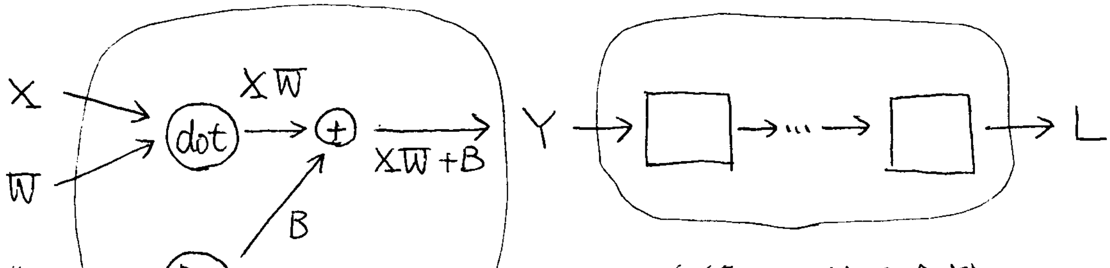
iv) 行列 $A=(a_{ij})$ の転置を A^T と表す。

$$A^T=((A^T)_{ij})=((A)_{ji})=(a_{ji})$$

v) $N\times N'$ 行列 $A=(a_{ij})$ のスカラ関数 $L=L(A)$ があるとき、
L の A による微分を下記のように書く。

$$\frac{\partial L}{\partial A}=\left(\frac{\partial L}{\partial a_{ij}}\right)=\begin{pmatrix} \frac{\partial L}{\partial a_{11}} & \cdots & \frac{\partial L}{\partial a_{1N'}} \\ \vdots & \ddots & \vdots \\ \frac{\partial L}{\partial a_{N1}} & \cdots & \frac{\partial L}{\partial a_{NN'}} \end{pmatrix}$$

2. レイヤの設定



後続のレイヤを合成して
得られるレイヤ： $L = L(Y)$

Affineレイヤ

X, W, b, Dup_N は、次のようになっているものとする。

$X = (x_{ij})$ ： $N \times N'$ 行列

$W = (w_{ij})$ ： $N' \times M$ 行列

$b = (b_i)$ ： M次元行ベクトル

$$B = \mathrm{Dup}_N(b) := \begin{pmatrix} b_1 & b_2 & \cdots & b_M \\ b_1 & b_2 & \cdots & b_M \\ \vdots & \vdots & & \vdots \\ b_1 & b_2 & \cdots & b_M \end{pmatrix}$$

： ベクトルbを縦にN個並べて得られる $N \times M$ 行列

3. 証明

・レイヤ設定より、Bの(i,j)成分は $(B)_{ij} = b_j$ なので、

$$Y = (y_{ij}) = XW + B = \left(\sum_k x_{ik} w_{kj} + b_j \right)$$

となる。これを用いて、$L = L(Y) = L(y_{11}, \cdots, y_{NM})$ の微分を計算する。

・まず、$\frac{\partial L}{\partial X}$ の(i,j)成分を計算する。$L = L(y_{11}, \cdots, y_{NM})$ なので、

$$\left(\frac{\partial L}{\partial X}\right)_{ij} = \frac{\partial L}{\partial x_{ij}} = \sum_{s,t} \frac{\partial L}{\partial y_{st}} \frac{\partial y_{s,t}}{\partial x_{ij}}$$

$$= \sum_{s,t} \frac{\partial L}{\partial y_{st}} \cdot \frac{\partial}{\partial x_{ij}} \left(\sum_u x_{su} w_{ut} + b_t \right)$$

$$= \sum_{s,t} \frac{\partial L}{\partial y_{st}} \left(\sum_u w_{ut} \frac{\partial x_{su}}{\partial x_{ij}} + \frac{\partial b_t}{\partial x_{ij}} \right)$$

ここで、

$$\frac{\partial x_{su}}{\partial x_{ij}} = \begin{cases} 1 & (s=i \text{ かつ } u=j) \\ 0 & (\text{それ以外}) \end{cases}$$

$$\frac{\partial b_t}{\partial x_{ij}} = 0 \quad (\text{すべての } t)$$

なので、和に寄与するのは $s=i$ かつ $u=j$ の項のみ。従って、

$$\left(\frac{\partial L}{\partial X}\right)_{ij} = \sum_t \frac{\partial L}{\partial y_{it}} \cdot w_{jt}$$

$$= \sum_t \left(\frac{\partial L}{\partial Y}\right)_{it} \cdot \left(W^T\right)_{tj}$$

$$= \left(\frac{\partial L}{\partial Y} \cdot W^T\right)_{ij}$$

よって

$$\frac{\partial L}{\partial X} = \frac{\partial L}{\partial Y} \cdot W^T$$

- $\frac{\partial L}{\partial W}$ も同様に計算できる。

$$\left(\frac{\partial L}{\partial W}\right)_{ij} = \frac{\partial L}{\partial w_{ij}} = \sum_{s,t} \frac{\partial L}{\partial y_{st}} \frac{\partial y_{st}}{\partial w_{ij}}$$

$$= \sum_{s,t} \frac{\partial L}{\partial y_{st}} \left(\sum_u x_{su} \frac{\partial w_{ut}}{\partial w_{ij}} + \frac{\partial b_t}{\partial w_{ij}} \right)$$

$$= \sum_s \frac{\partial L}{\partial y_{sj}} \cdot x_{si}$$

$$= \sum_s \left(X^T\right)_{is} \cdot \left(\frac{\partial L}{\partial Y}\right)_{sj}$$

$$= \left(X^T \cdot \frac{\partial L}{\partial Y}\right)_{ij}$$

$$\therefore \frac{\partial L}{\partial W} = X^T \cdot \frac{\partial L}{\partial Y}$$

・ほぼ同様にして $\frac{\partial L}{\partial \mathbb{b}}$ も計算することができる。

$$\left(\frac{\partial L}{\partial \mathbb{b}}\right)_i = \frac{\partial L}{\partial b_i} = \sum_{s,t} \frac{\partial L}{\partial y_{st}} \frac{\partial y_{st}}{\partial b_i}$$

$$= \sum_{s,t} \frac{\partial L}{\partial y_{st}} \cdot \left(\sum_u \frac{\partial (x_{su} W_{ut})}{\partial b_i} + \frac{\partial b_t}{\partial b_i} \right)$$

ここで

$$\frac{\partial (x_{su} W_{ut})}{\partial b_i} = 0 \quad (\text{すべての } u)$$

$$\frac{\partial b_t}{\partial b_i} = \begin{cases} 1 & (t = i) \\ 0 & (\text{それ以外}) \end{cases}$$

より、

$$\left(\frac{\partial L}{\partial \mathbb{b}}\right)_i = \sum_s \frac{\partial L}{\partial y_{si}} = \sum_s \left(\frac{\partial L}{\partial Y}\right)_{si}$$

これは $\frac{\partial L}{\partial Y}$ の第 i 列成分のすべての行についての和。すなわち、

$\frac{\partial L}{\partial \mathbb{b}} = \frac{\partial L}{\partial Y}$ のすべての行を足し合わせて得られるベクトル

・データを1件ずつ入力する場合、X, Y はともに行ベクトルとなるから、$\mathbb{b}$ は行ベクトル $\frac{\partial L}{\partial \mathbb{b}} = \frac{\partial L}{\partial Y}$ によって順次更新されていく。
(学習係数により重みづけされる)

先の結果は、バッチデータ入力では各データを個々に入力していく場合の毎回の更新が一括で行われることを表している。

※ 実際には、個別入力の場合には入力のたびにネットワークが更新されるため、バッチデータで一度に更新する場合と結果が完全に一致するわけではない。しかし、充分に小さな学習係数を選べば結果は大きくは異ならないと考えられる。