ニューラルネットワーク

Python 中級
鈴木 敬彦

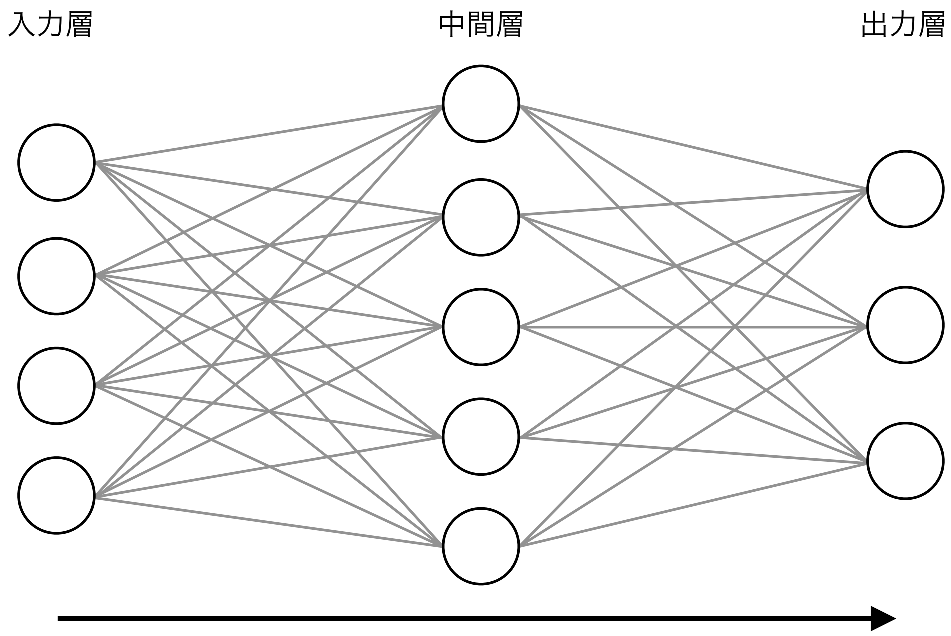
より正確には、「人工神経網 (Artificial Neural Network)」と記すべきでしょう。ニューラルネットワークは、文字通り、我々人間やその他の動物や生物の脳を構成する神経回路網を模したものです。神経回路網は、主に神経細胞/ニューロン (neuron) とシナプス (synapse) から構成され、神経細胞への刺激がその細胞の電位を変化させ、その変化をシナプスが他の神経細胞へと伝達することで情報を処理しています。ニューラルネットワークは、この情報伝達の仕組みを模したもので、神経細胞に相当するノードへの入力に対して、重みとバイアスを加え、その結果を別のノードに伝達していくことで情報を処理しようとする数理モデルです。

神経細胞の結合の仕方によって様々なタイプのニューラルネットワークが存在しますが、ここでは多層パーセプトロン (**M**ulti**l**ayer **P**erceptron、以下 MLP) と呼ばれるニューラルネットワークについて説明します。

MLP は名前の通り、「多層」構造になっています。MLP 以前には多層ではない形式ニューロンや単純パーセプトロンというものがありましたが、単純な問題しか処理できませんでした。その後、MLP とその最適化手法の確立は、1980 年代の 2 度目のニューラルネットワークブームの一翼を担い、そして、2010 年代の 3 度目のブームの発端とも言える「深層学習 / ディープラーニング」においてはその中核を担う技術となっています。

多層パーセプトロン / Multilayer Perceptron

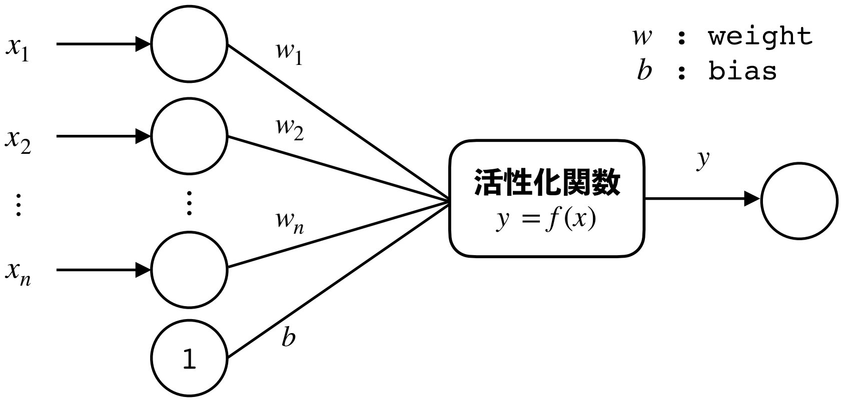
MLP は、入力層と出力層の間に中間層もしくは隠れ層 (hidden layer) と呼ばれる層を 1 層持つ全結合順伝播型のニューラルネットワークです。全結合とは、各層のノード（ニューロンのこと）が次の層の全ノードと結合されるということです。順伝播とは、ニューラルネットワークの計算の方向が、入力層から出力層の方向のみだということです。



各ノードからノードへの結合は係数を持ち、入力値にこの係数を乗じてから、全てのノードからの出力を足し合わせ、更にバイアスを加えます。（閾値を設定しこれを減じる場合もあります。）最後に活性化関数（図には書かれていません）を適用して次のノードへの入力とします。

以下の図は、任意の層への入力が次の層の一つのノードに渡されるまでの計算の流れを図示したものです。入力層最下部の「1」と書かれているノードはバイアス用の便宜的なノードです。

入力：　$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$



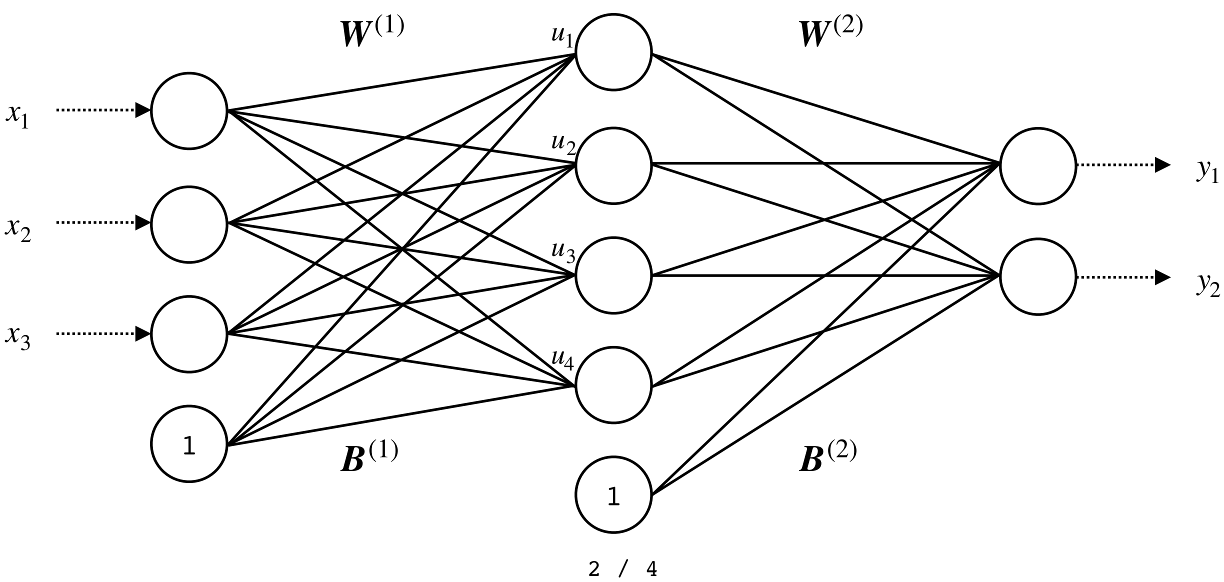
出力：　$\boldsymbol{y} = f(w_1x_1 + w_2x_2 + \cdots + w_nx_n + b)$

$= f(\boldsymbol{x}\boldsymbol{w} + b)$　　　...　$\boldsymbol{w} = (w_1, w_2, \cdots, w_n)^\mathsf{T}$

係数 $\boldsymbol{w}$ はウェイト (weight) と呼ばれ、$\boldsymbol{b}$ はバイアス (bias) と呼ばれます。入力がベクトルであれば、層が持つウェイトとバイアスは、それぞれ行列とベクトルになります。

活性化関数には様々なタイプの関数があり、後半でいくつかの活性化関数を紹介しますが、以後のニューラルネットワークの計算の説明においては省きます。

MLP　全体の順伝播の計算は以下のようになります。ここでは簡便のため、入力層、中間層、出力層のノード数をそれぞれ 3、4、2 とします。入力と出力をそれぞれ $\boldsymbol{x}$、$\boldsymbol{y}$ とし、入力層から中間層への入力を $\boldsymbol{u}$ とします。入力層と中間層の結合のウェイトとバイアスを $\boldsymbol{W}^{(1)}$、$\boldsymbol{B}^{(1)}$ で表し、中間層と出力層の結合のウェイトとバイアスを $\boldsymbol{W}^{(2)}$、$\boldsymbol{B}^{(2)}$ で表します。ここでは、入力 $\boldsymbol{x}$ を行ベクトルとして、$\boldsymbol{W}^{(1)}$ を右から乗じます。（実装において入力値の形状やウェイトをどちらから乗じる



かは設計次第です。）これらより $\boldsymbol{W}^{(1)}$と $\boldsymbol{W}^{(2)}$の形状は、それぞれ 3×4、4×2 とわかります。この MLP の出力は以下のようになります。

$$
\begin{aligned}
\boldsymbol{u} &= \boldsymbol{x}\boldsymbol{W}^{(1)} + \boldsymbol{B}^{(1)} \\
\boldsymbol{y} &= \boldsymbol{u}\boldsymbol{W}^{(2)} + \boldsymbol{B}^{(2)} = \left(\boldsymbol{x}\boldsymbol{W}^{(1)} + \boldsymbol{B}^{(1)}\right)\boldsymbol{W}^{(2)} + \boldsymbol{B}^{(2)}
\end{aligned}
$$

これを成分毎の記述に展開すると以下のようになります。

$$
\boldsymbol{x} = (x_1 \quad x_2 \quad x_3)\ , \qquad \boldsymbol{u} = (u_1 \quad u_2 \quad u_3 \quad u_4)\ , \qquad \boldsymbol{y} = (y_1 \quad y_2)
$$

$$
\boldsymbol{W}^{(1)} = \begin{pmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} & w_{14}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} & w_{24}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} & w_{34}^{(1)} \end{pmatrix}\ , \qquad \boldsymbol{B}^{(1)} = \left(b_1^{(1)} \quad b_2^{(1)} \quad b_3^{(1)} \quad b_4^{(1)}\right)
$$

$$
\boldsymbol{W}^{(2)} = \begin{pmatrix} w_{11}^{(2)} & w_{21}^{(2)} & w_{31}^{(2)} & w_{41}^{(2)} \\ w_{12}^{(2)} & w_{22}^{(2)} & w_{32}^{(2)} & w_{42}^{(2)} \end{pmatrix}^{\mathsf{T}}\ , \qquad \boldsymbol{B}^{(2)} = \left(b_1^{(2)} \quad b_2^{(2)}\right)
$$

$$
\begin{aligned}
\boldsymbol{u} &= (u_1 \quad u_2 \quad u_3 \quad u_4) = \boldsymbol{x}\boldsymbol{W}^{(1)} + \boldsymbol{B}^{(1)} \\
&= (x_1 \quad x_2 \quad x_3)\begin{pmatrix} w_{11}^{(1)} & w_{12}^{(1)} & w_{13}^{(1)} & w_{14}^{(1)} \\ w_{21}^{(1)} & w_{22}^{(1)} & w_{23}^{(1)} & w_{24}^{(1)} \\ w_{31}^{(1)} & w_{32}^{(1)} & w_{33}^{(1)} & w_{34}^{(1)} \end{pmatrix} + \left(b_1^{(1)} \quad b_2^{(1)} \quad b_3^{(1)} \quad b_4^{(1)}\right) \\
&= \begin{pmatrix} x_1w_{11}^{(1)} + x_2w_{21}^{(1)} + x_3w_{31}^{(1)} + b_1^{(1)} \\ x_1w_{12}^{(1)} + x_2w_{22}^{(1)} + x_3w_{32}^{(1)} + b_2^{(1)} \\ x_1w_{13}^{(1)} + x_2w_{23}^{(1)} + x_3w_{33}^{(1)} + b_3^{(1)} \\ x_1w_{14}^{(1)} + x_2w_{24}^{(1)} + x_3w_{34}^{(1)} + b_4^{(1)} \end{pmatrix}^{\mathsf{T}} \qquad \texttt{i.e.} \quad \begin{cases} u_1 = x_1w_{11}^{(1)} + x_2w_{21}^{(1)} + x_3w_{31}^{(1)} + b_1^{(1)} \\ u_2 = x_1w_{12}^{(1)} + x_2w_{22}^{(1)} + x_3w_{32}^{(1)} + b_2^{(1)} \\ u_3 = x_1w_{13}^{(1)} + x_2w_{23}^{(1)} + x_3w_{33}^{(1)} + b_3^{(1)} \\ u_4 = x_1w_{14}^{(1)} + x_2w_{24}^{(1)} + x_3w_{34}^{(1)} + b_4^{(1)} \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{y} &= (y_1 \quad y_2) = \boldsymbol{u}\boldsymbol{W}^{(2)} + \boldsymbol{B}^{(2)} \\
&= (u_1 \quad u_2 \quad u_3 \quad u_4)\begin{pmatrix} w_{11}^{(2)} & w_{21}^{(2)} & w_{31}^{(2)} & w_{41}^{(2)} \\ w_{12}^{(2)} & w_{22}^{(2)} & w_{32}^{(2)} & w_{42}^{(2)} \end{pmatrix}^{\mathsf{T}} + \left(b_1^{(2)} \quad b_2^{(2)}\right) \\
&= \begin{pmatrix} u_1w_{11}^{(2)} + u_2w_{21}^{(2)} + u_3w_{31}^{(2)} + u_4w_{41}^{(2)} + b_1^{(2)} \\ u_1w_{12}^{(2)} + u_2w_{22}^{(2)} + u_3w_{32}^{(2)} + u_4w_{42}^{(2)} + b_2^{(2)} \end{pmatrix}^{\mathsf{T}} \\
&= \begin{pmatrix} (x_1w_{11}^{(1)} + x_2w_{21}^{(1)} + x_3w_{31}^{(1)} + b_1^{(1)})w_{11}^{(2)} + (x_1w_{12}^{(1)} + x_2w_{22}^{(1)} + x_3w_{32}^{(1)} + b_2^{(1)})w_{21}^{(2)} \\ +(x_1w_{13}^{(1)} + x_2w_{23}^{(1)} + x_3w_{33}^{(1)} + b_3^{(1)})w_{31}^{(2)} + (x_1w_{14}^{(1)} + x_2w_{24}^{(1)} + x_3w_{34}^{(1)} + b_4^{(1)})w_{41}^{(2)} + b_1^{(2)} \\ \\ (x_1w_{11}^{(1)} + x_2w_{21}^{(1)} + x_3w_{31}^{(1)} + b_1^{(1)})w_{12}^{(2)} + (x_1w_{12}^{(1)} + x_2w_{22}^{(1)} + x_3w_{32}^{(1)} + b_2^{(1)})w_{22}^{(2)} \\ +(x_1w_{13}^{(1)} + x_2w_{23}^{(1)} + x_3w_{33}^{(1)} + b_3^{(1)})w_{32}^{(2)} + (x_1w_{14}^{(1)} + x_2w_{24}^{(1)} + x_3w_{34}^{(1)} + b_4^{(1)})w_{42}^{(2)} + b_2^{(2)} \end{pmatrix}^{\mathsf{T}}
\end{aligned}
$$

i.e.

$$
\begin{cases}
y_1 = (x_1w_{11}^{(1)} + x_2w_{21}^{(1)} + x_3w_{31}^{(1)} + b_1^{(1)})w_{11}^{(2)} + (x_1w_{12}^{(1)} + x_2w_{22}^{(1)} + x_3w_{32}^{(1)} + b_2^{(1)})w_{21}^{(2)} \\
\quad + \ (x_1w_{13}^{(1)} + x_2w_{23}^{(1)} + x_3w_{33}^{(1)} + b_3^{(1)})w_{31}^{(2)} + (x_1w_{14}^{(1)} + x_2w_{24}^{(1)} + x_3w_{34}^{(1)} + b_4^{(1)})w_{41}^{(2)} + b_1^{(2)} \\
y_2 = (x_1w_{11}^{(1)} + x_2w_{21}^{(1)} + x_3w_{31}^{(1)} + b_1^{(1)})w_{12}^{(2)} + (x_1w_{12}^{(1)} + x_2w_{22}^{(1)} + x_3w_{32}^{(1)} + b_2^{(1)})w_{22}^{(2)} \\
\quad + \ (x_1w_{13}^{(1)} + x_2w_{23}^{(1)} + x_3w_{33}^{(1)} + b_3^{(1)})w_{32}^{(2)} + (x_1w_{14}^{(1)} + x_2w_{24}^{(1)} + x_3w_{34}^{(1)} + b_4^{(1)})w_{42}^{(2)} + b_2^{(2)}
\end{cases}
$$

活性化関数 / activator function

活性化関数は伝達関数 (transfer function) とも呼ばれています。
なぜ活性化関数が必要なのかという点については、先程の MLP の式をもう少し整理してみると理解しやすいでしょう。活性化関数を持たない MLP は単なる線形変換に過ぎず、ウェイトが $\boldsymbol{W}^{(1)}\boldsymbol{W}^{(2)}$、バイアスが $\boldsymbol{B}^{(1)}\boldsymbol{W}^{(2)}+\boldsymbol{B}^{(2)}$ である 2 層のネットワークと等価です。

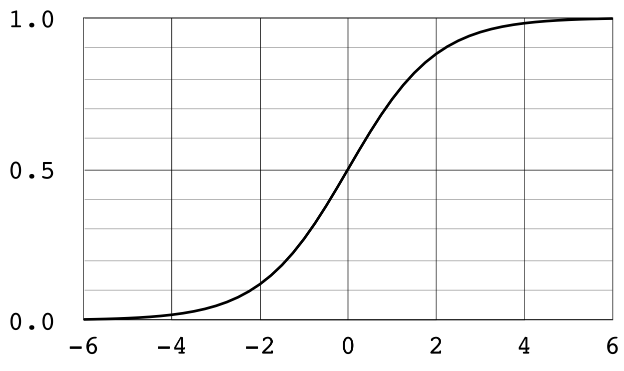
$$\begin{aligned}\boldsymbol{y} &= \left(\boldsymbol{x}\boldsymbol{W}^{(1)}+\boldsymbol{B}^{(1)}\right)\boldsymbol{W}^{(2)}+\boldsymbol{B}^{(2)} = \boldsymbol{x}\boldsymbol{W}^{(1)}\boldsymbol{W}^{(2)}+\boldsymbol{B}^{(1)}\boldsymbol{W}^{(2)}+\boldsymbol{B}^{(2)} \\ &= \boldsymbol{x}\left(\boldsymbol{W}^{(1)}\boldsymbol{W}^{(2)}\right)+\left(\boldsymbol{B}^{(1)}\boldsymbol{W}^{(2)}+\boldsymbol{B}^{(2)}\right)\end{aligned}$$

この理由から、活性化関数には非線形関数が用いられ、一般的なニューラルネットワークの構成においては出力層に活性化関数を用いません。また、後述（ニューラルネットワーク II）の理由から、活性化関数は微分可能でなければなりません。

活性化関数には様々な関数がこれまでに提案され用いられてきました。ここでは、MLP 興盛時に多様されていたシグモイド関数と、深層学習でも多用される ReLU を紹介します。
また、ニューラルネットワークの計算は行列演算ですが、活性化関数の適用は、個々の成分に対しての適用なので注意してください。

シグモイド関数 / Sigmoid Function

$$f(x) = \frac{1}{1+e^{-x}}$$



ReLU (Rectified Linear Unit)

$$f(x) = \begin{cases} x & \texttt{if } x > 0, \\ 0 & \texttt{otherwise.} \end{cases}$$