

通信理論に特化した深層学習 第4回ゼミ資料 ニューラルネットワーク

豊橋技術科学大学
電気・電子情報工学系
准教授 竹内啓悟

順伝播型ニューラルネットワーク(Feedforward neural network)

脳の機能を模倣した推論モデルで、ニューロン間のシナプス結合の強度を学習することで推論を行う数理モデルである。

ニューロン(以後、ユニットと呼ぶ)

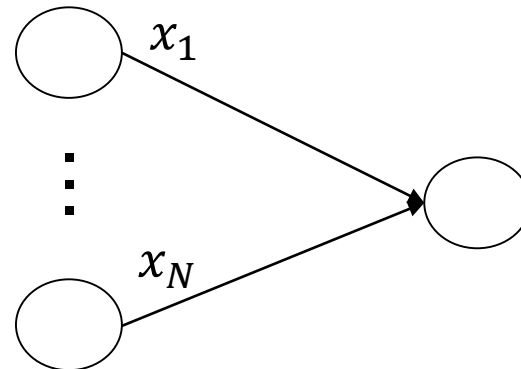
入力 $u \in \mathbb{R}$ を出力 $x \in \mathbb{R}$ に写像する活性化関数 f が付随する。



結合強度(以後、重みと呼ぶ)とバイアス

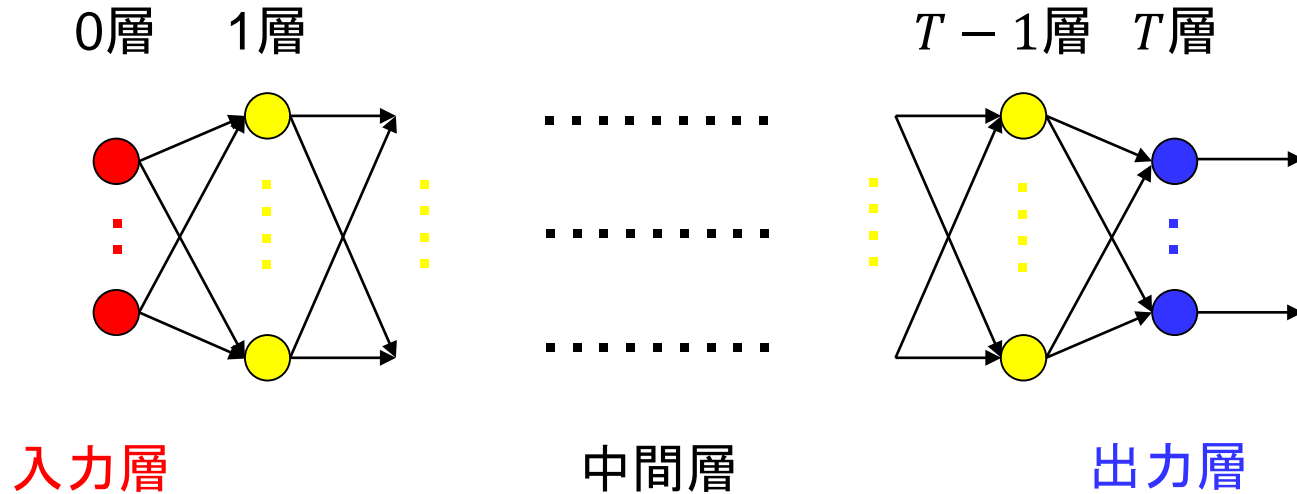
N 個のユニットの出力 $\{x_n\}$ と別のユニットの入力との関係を決める。

$$u = \sum_{n=1}^N w_n x_n + b$$



T層順伝播型ネットワーク

ユニットを直列にT層にわたって結合させる。



注意1

入力層は、0層目とし、層の数に含めない。

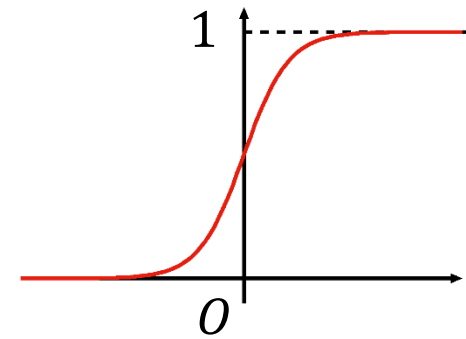
注意2

各層ごとにユニット数や活性化関数が異なってもよい。

活性化関数 (Activation function) の例

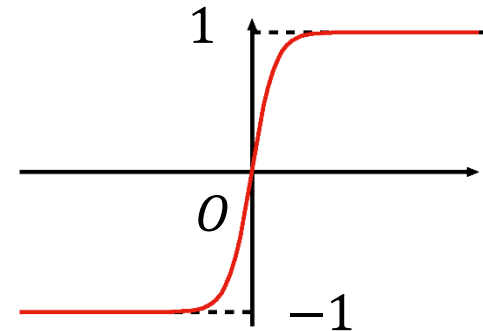
シグモイド関数

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



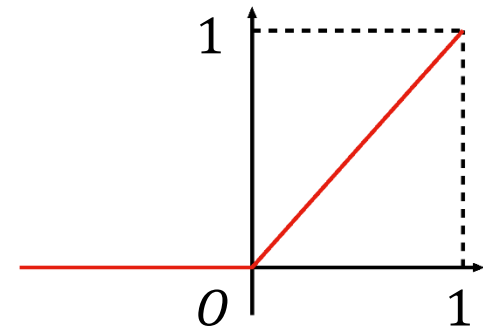
tanh関数

$$f(u) = \tanh u.$$



正規化線形 (Rectified linear) 関数

$$f(u) = \max\{u, 0\}$$



正規化線形関数を持つユニットをReLUと呼ぶ。

T層順伝播型ネットワーク

$t - 1$ 層目と t 層目の関係

$$\mathbf{x}_t = f_t(\mathbf{u}_t), \quad \mathbf{u}_t = \mathbf{W}_t \mathbf{x}_{t-1} + \mathbf{b}_t.$$

活性化関数 f_t は同一層では同一関数を使用し、要素ごとに適用する。

N_t : t 層目のユニット数

$\mathbf{W}_t \in \mathbb{R}^{N_t \times N_{t-1}}$: $t - 1$ 層目と t 層目の間の重み行列

$\mathbf{b}_t \in \mathbb{R}^{N_t}$: t 層目のバイアスベクトル

損失関数 L と学習目標

二乗誤差

$$L(\mathbf{t}, \mathbf{y}) = \frac{1}{2} \|\mathbf{t} - \mathbf{y}\|^2$$

\mathbf{t} : 出力目標とする教師データベクトル

\mathbf{y} : 出力層から出てくる出力ベクトル

学習の目標

未知の入力データ(評価用データ)が入力されたときに、平均二乗誤差を最小にするという意味で出力が目標に最も近づくように、訓練用データを使ってユニット数、層の数、重み、バイアス等の設計パラメータを最適化する。

評価用データに対する誤差を汎化誤差と呼ぶ。

訓練用データに適合しすぎて、訓練用データに対する誤差と汎化誤差とに著しい乖離が生じる状態を過学習と呼ぶ。

2層順伝播型ネットワークの表現力

定理4.1

中間層の活性化関数 f_1 は以下の条件を満たす任意の関数とする。

- f_1 は有界かつ可測である。
- $\lim_{u \rightarrow \infty} f_1(u) = 1, \lim_{u \rightarrow -\infty} f_1(u) = 0.$

任意の入力次元 $N_0 \in \mathbb{N}$ と任意の連続関数 $f: [0, 1]^{N_0} \rightarrow \mathbb{R}$ に対して、中間層のユニット数 N_1 を十分に大きくすれば、出力層の活性化関数に恒等写像を使用した2層順伝播型ネットワークで、任意の精度で関数 f を一様に近似できるものが存在する。

中間層のユニット数を無限大にすれば、2層で十分な表現力が得られる。

[4-1] G. Cybenko, "Approximation by Superpositions of a Sigmoidal Function,"
Math. Control Signals Systems, vol. 2, no. 4, pp. 303-314, 1989.

深層ニューラルネットワークの表現力

定理4.2

任意の入力次元 $N_0 \in \mathbb{N}$ と任意の可測関数 $f: \mathbb{R}^{N_0} \rightarrow \mathbb{R}$ に対して、層の数 T を十分に大きくすれば、中間層のユニット数が $N_0 + 4$ 以下でユニットにReLUを用いた深層ニューラルネットワークで、 L_1 ノルムの意味で関数 f を任意の精度で近似できるものが存在する。

層数を無限大にすれば、層当たり有限のユニット数で十分な表現力が得られる。

ユニット数と層数の適切なバランスについては未解決である。

[4-2] Z. Lu *et al.*, “The Expressive Power of Neural Networks: A View from the Width,” *Proc. 31st Conf. Neural Inf. Process. Syst.*, Long Beach, CA, USA, Dec. 2017.