

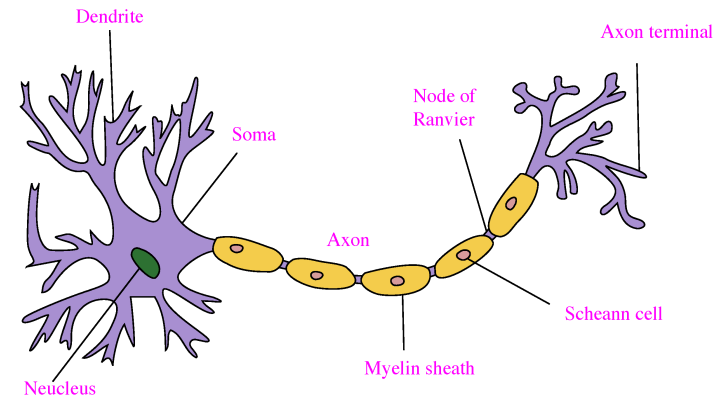
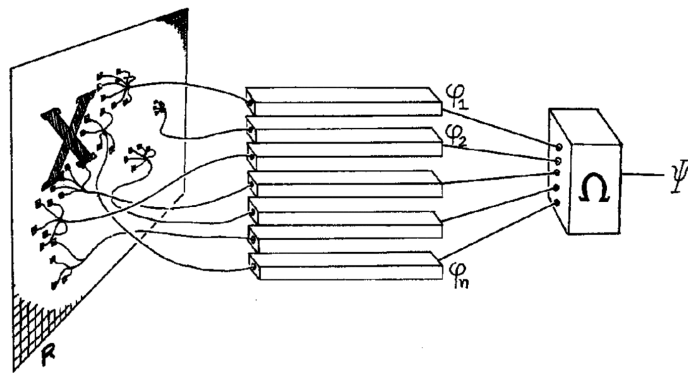
02 ニューラルネットワーク

浅川伸一

パーセプトロン perceptrons (Rosenblatt 1958)



フランク・ローゼンブラット



左：パーセプトロンの模式図 ミンスキーとパパート「パーセプトロン」より 右：ニューロンの模式図 wikipedia より

パーセプトロンの学習

$$\mathbf{w} \leftarrow \mathbf{w} + (y - \hat{y}) \mathbf{x} \quad (1)$$

パーセプトロン perceptron は 3 層の階層型ネットワークでそれぞれ S(sensory layer), A(associative layer), R(response layer) と呼ぶ。 $S \rightarrow A \rightarrow R$ のうち パーセプトロンの本質的な部分は $A \rightarrow R$ の間の学習にある。

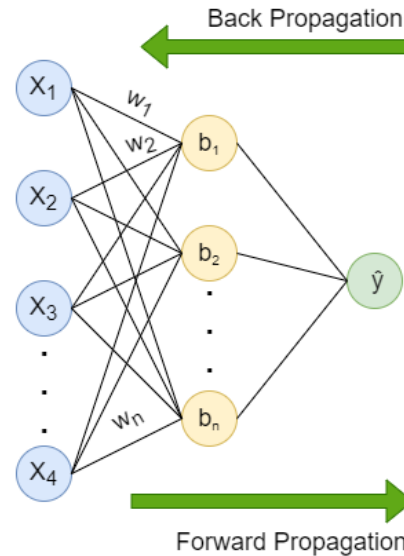
入力パターンに P^+ と P^- とがある。 パーセプトロンは P^+ が入力されたとき 1, P^- のとき 0 を出力する 機械である。 出力層 (R) の i 番目のニューロンへの入力 (膜電位の変化) u_i は

$$y_i = [u_i] \quad \begin{cases} 1 & \text{if } u_i \geq 0, \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

と表される。

ニューラルネットワークの学習

ニューラルネットワークの学習過程には 前向き伝播と 逆向き伝播 の 2 つの部分があります。

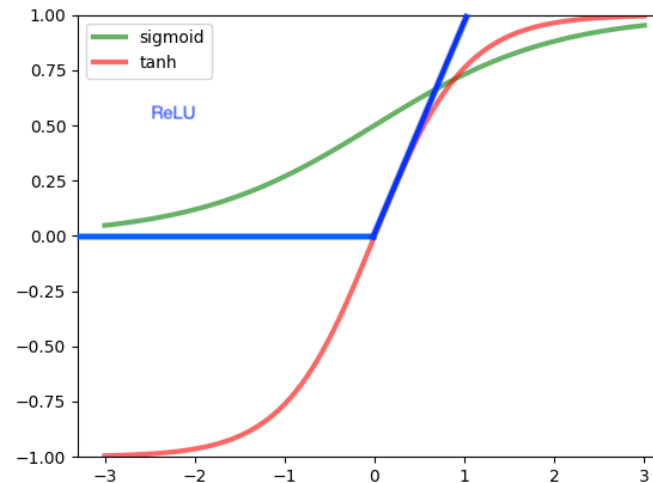


出典: <https://becominghuman.ai/deep-learning-an-overview-5f40feb15d1e>

- 順伝播: 入力層から出力層への情報の流れ
- 入力層: 複数ニューロンで構成。**結合係数** または **重み** (w_i) と呼ばれる数値で重みづけて次層に接続
- 初期段階では、重みはランダムに割り当てられる。後に更新される **学習**
- 隠れ層のすべてのニューロンには、**バイアス** (b_i) という値が関連付けられており、これが入力和に加えられる
- この重み付けされた和は、**活性化関数** (σ) と呼ばれる非線形関数に渡される
- 最後に出力層で確率を求め、最も高い確率のニューロンが **最終出力** \hat{y} となる
- この処理は、数学的には次のように表される

$$\hat{y} = \sigma \left(\sum_{i=1}^n x_i w_i + b_i \right)$$

活性化関数



シグモイド関数, tanh, ReLU

- ロジスティックシグモイド関数 logistic sigmoid function (Rumelhart, Hinton, and Williams 1986)

$$y = \sigma(x) = (1 + e^{-x})^{-1}, f'(x) = y(1 - y) \quad (4)$$

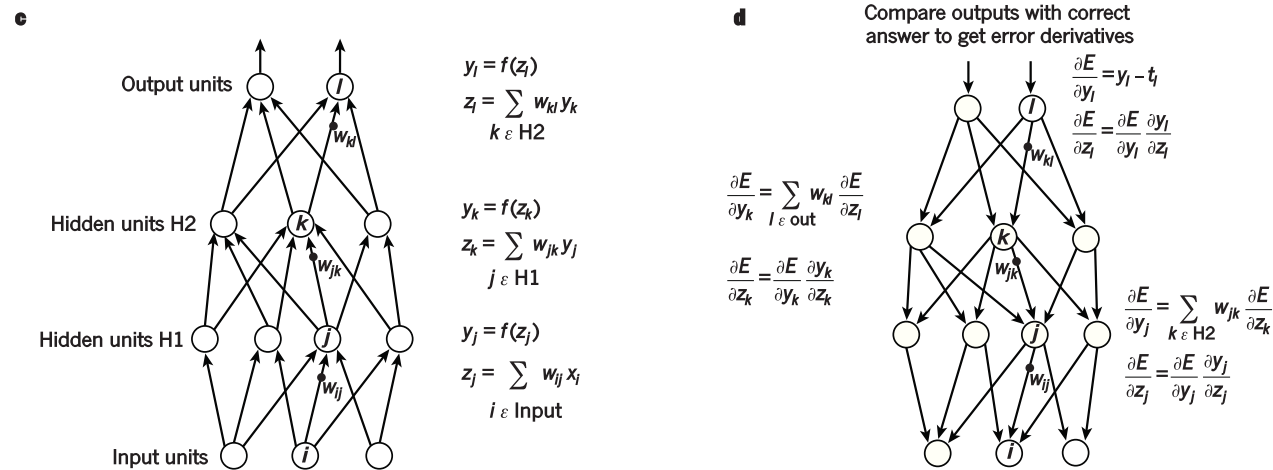
- ハイパータンジェント (LeCun et al. 1998)

$$y = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, f'(x) = 1 - y^2 \quad (5)$$

- 整流線形ユニットReLU (Nair and Hinton 2010a, 2010b)

$$\text{ReLU}(x) = \max [0, x] \quad (6)$$

誤差逆伝播法



出典: (LeCun, Bengio, and Hinton 2015)

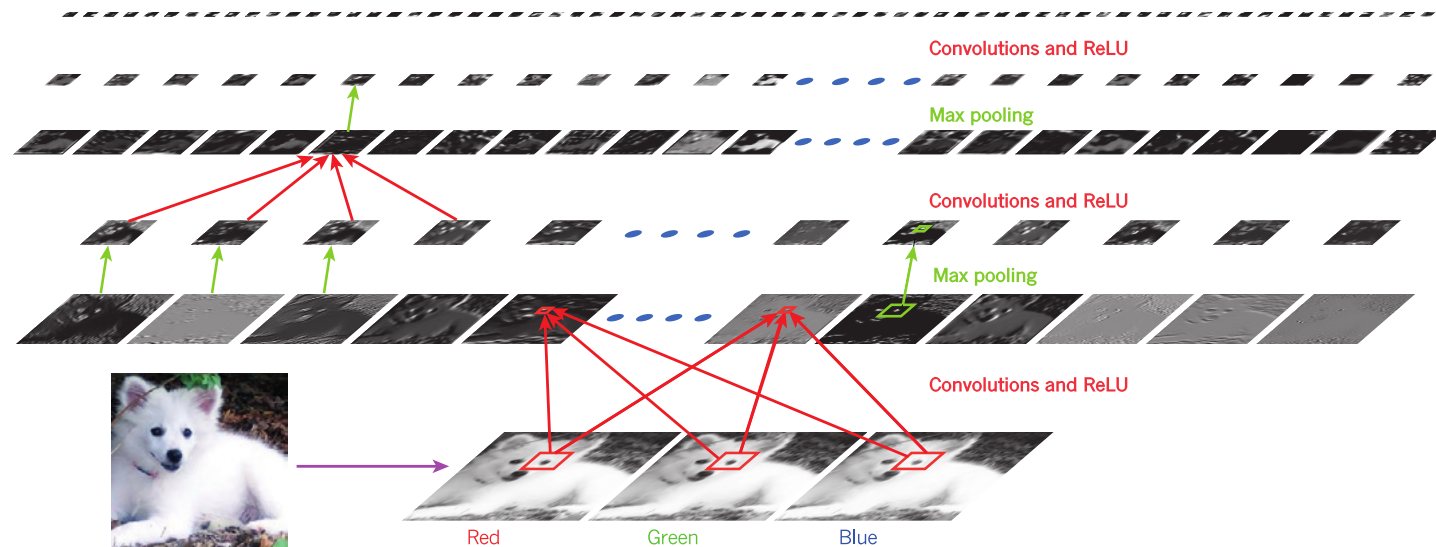
- 逆伝播は、順伝播と同じですが、方向が逆
- 誤差逆伝播では、情報は出力層から入力層に渡される
- 最後層が予測を行った後、**損失関数**によって評価される
- 損失関数は、予想される出力からの逸脱を定量化するのに役立ちます。
- 予測された出力と実際の出力との乖離を表す値を与える
- この情報は隠れ層に戻され、より正確な予測を得るために重みとバイアスを調整する
- 重みとバイアスは、勾配降下法（オプティマイザー）を用いて更新される

注釈： ウェイトとバイアスはモデルのパラメータと呼び、学習率はモデルのハイパーパラメータと呼ぶ。

既存のニューラルネットワークにおける問題

1. 勾配消失問題 the gradient vanishing problem (\Rightarrow 勾配爆発問題 the gradient exploding problem)
2. 信用割当問題 credit assignment problems (強化学習の意味での信用割当問題とは異なる)

Samoyed (16); Papillon (5.7); Pomeranian (2.7); Arctic fox (1.0); Eskimo dog (0.6); white wolf (0.4); Siberian husky (0.4)



(LeCun, Bengio, and Hinton 2015) より

まとめ

- 勾配消失問題: 活性化関数の工夫
- 信用割当問題: 畳み込みニューラルネットワーク

クイズ

活性化関数の中で、シグモイド関数と ReLU では、どちらが勾配消失問題が起こりにくいでしょうか？

クイズの答え

活性化関数の中で、シグモイド関数と ReLU では、どちらが勾配消失問題が起こりにくいでしょうか？

整流線形関数 ReLU

文献

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* 521: 436–44.

LeCun, Yann, Leon Bottou, Genevieve B. Orr, and Klaus-Robert Müller. 1998. "Efficient BackProps." In *Neural Networks: Tricks and the Trade*, edited by Grégoire Montavon, Geneviève B. Orr, and Klaus-Robert Müller, 9–48. Berlin Heidelberg, Germany: Springer-Verlag. https://doi.org/10.1007/978-3-642-35289-8_3.

Nair, Vinod, and Geoffrey E. Hinton. 2010a. "Rectified Linear Units Improve Restricted Boltzmann Machines." In *In Proceedings the 27th International Conference on Machine Learning (ICML)*, edited by Johannes Fürnkranz and Thorsten Joachims. Haifa, Israel: Omnipress. <http://www.icml2010.org/>.

—. 2010b. "Rectified Linear Units Improve Restricted Boltzmann Machines." In *Proceedings of the 27th International Conference on Machine Learning (ICML)*, edited by Johannes Fürnkranz and Thorsten Joachims, 807–14. Haifa, Israel: Omnipress.

Rosenblatt, Frank. 1958. "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain." *Psychological Review* 65: 386–408.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. 1986. "Learning Representations by Back-Propagating Errors." *Nature* 323 (6088): 533–36. <https://doi.org/doi:10.1038/323533a0>.