2　順伝播型ネットワーク

ニューラルネットワークを定式化

2.1　ユニットの出力

キーワード

順伝播型 (ニューラル) ネットワーク，多層パーセプトロン，重み，バイアス，活性化関数

ネットワークを構成する各ユニットは，入力 $x_1, ..., x_I$ を受けとり，総入力:

$$u = w_1 x_1 + \cdots + w_I x_I + b \tag{1}$$

を計算した上で，活性化関数 f を用いて

$$z = f(u) \tag{2}$$

を出力します．ここで $w_1, ..., w_I$ は**重み**，b は**バイアス**です．

1 つの層に複数のユニットが存在している場合，ベクトルを用いて次のように書けます．

$$\mathbf{u} = \mathbf{W}\mathbf{x} + \mathbf{b} \tag{3}$$

$$\mathbf{z} = \mathbf{f}(\mathbf{u}) \tag{4}$$

ただし，入力 (=第一層) のユニット数は I，第二層のユニット数は J とし，各記号は次のとおり：

$$\mathbf{u} = (u_1, ..., u_J)^\top, \mathbf{x} = (x_1, ..., x_I)^\top, \mathbf{b} = (b_1, ..., b_J)^\top, \mathbf{z} = (z_1, ..., z_J)^\top, \tag{5}$$

$$\mathbf{W} = \begin{pmatrix} w_{11} & \cdots & w_{1I} \\ \vdots & \ddots & \vdots \\ w_{J1} & \cdots & w_{JI} \end{pmatrix}, \mathbf{f}(\mathbf{u}) = (f(u_1), ..., f(u_J))^\top \tag{6}$$

2.2　活性化関数 (ユニット間，出力層ではない)

キーワード

ロジスティック (シグモイド) 関数，シグモイド関数，正規化線形関数，マックスアウト

- ロジスティック関数 ⊂ シグモイド関数

$$f(u) = \frac{1}{1 + e^{-u}} \tag{7}$$

- 双曲線正接関数 (hyperbolic tangent) ⊂ シグモイド関数

$$f(u) = \tanh(u) \tag{8}$$

- 正規化線形関数

$$f(u) = \max(u, 0) \tag{9}$$

  この関数をもつユニットを **ReLU** と言います．

- 線形写像・恒等写像

- ロジスティック関数を区分線形近似したもの

$$f(u) = \begin{cases} -1 & u < -1 \\ u & -1 \leq u < 1 \\ 1 & u \geq 1 \end{cases} \tag{10}$$

- マックスアウト (K 個のユニットをまとめたようなもの)

$$u_{jk} = \sum_i w_{jik} z_i + b_{jk} \tag{11}$$

$$f(u_j) = \max_{k+1,\ldots,K} u_{jk} \tag{12}$$

2.3 多層ネットワーク

キーワード

入力層，中間層 (=隠れ層)，出力層

各層のユニットの入出力の計算を区別するために，層が $l = 1, 2, ..., L$ となっているときに各変数の肩に (l) と記すことにします．すると，層 2 のユニットの出力 $\mathbf{z}^{(2)}$ は，入力層 $\mathbf{x}$ から

$$\mathbf{u}^{(2)} = \mathbf{W}^{(2)}\mathbf{x} + \mathbf{b}^{(2)} \tag{13}$$

$$\mathbf{z}^{(2)} = \mathbf{f}(\mathbf{u}^{(2)}) \tag{14}$$

と計算され，その後の層 $l+1$ のユニットの出力 $\mathbf{z}^{(l+1)}$ は，1 つ下の層の出力 $\mathbf{z}^{(l)}$ から

$$\mathbf{u}^{(l+1)} = \mathbf{W}^{(l+1)}\mathbf{z}^{(l)} + \mathbf{b}^{(l+1)} \tag{15}$$

$$\mathbf{z}^{(l+1)} = \mathbf{f}(\mathbf{u}^{(l+1)}) \tag{16}$$

と計算される．最後に層 L まで計算が終わったら，ネットワークの最終的な出力を

$$\mathbf{y} = \mathbf{z}^{(L)} \tag{17}$$

と表記する．各層の重み $\mathbf{W}^{(l)}$ とバイアス $\mathbf{b}^{(l)}$ をユニットのパラメータと呼び，出力 $\mathbf{y}$ は入力とパラメータによって決まることから，

$$\mathbf{y}\left(\mathbf{x}; \mathbf{W}^{(2)}, ..., \mathbf{W}^{(L)}, \mathbf{b}^{(2)}, ..., \mathbf{b}^{(L)}\right) \tag{18}$$

あるいは簡単に

$$\mathbf{y}(\mathbf{x}; \mathbf{w}) \tag{19}$$

と書きます．

2.4 出力層の設計と誤算関数

キーワード

訓練サンプル，訓練データ，誤算関数，回帰，二値分類，多クラス分類，最尤推定，ソフトマックス関数，交差エントロピー

学習に使用する入出力データのペア

$$\{(\mathbf{x}_1, \mathbf{d}_1), (\mathbf{x}_2, \mathbf{d}_2), ..., (\mathbf{x}_N, \mathbf{d}_N)\} \tag{20}$$

を訓練データといいます．

2.4.1　回帰

出力層の活性化関数には恒等写像を用い，誤算関数には二乗誤算の和

$$E(\mathbf{w} = \frac{1}{2}\sum_{n=1}^{N}||\mathbf{d}_n - \mathbf{y}(\mathbf{x}_n;\mathbf{w})||^2 \tag{21}$$

を使用し，$E(\mathbf{w})$ がもっとも小さくなる $\mathbf{w}$ を求める．

2.4.2　二値分類

ここでは $\mathbf{x}$ を指定したときに $d = 1$ となる事後確率 $p(d = 1|\mathbf{x})$ を，ネットワーク全体の出力 $y(\mathbf{x};\mathbf{w})$ でモデル化し，出力層の活性化関数にロジスティック関数を使うことで，

$$p(d = 1|\mathbf{x}) \approx y(\mathbf{x};\mathbf{w}) \tag{22}$$

とします．パラメータ $\mathbf{w}$ は最尤推定で求めることにします．$p(d = 1|\mathbf{x}) = y(\mathbf{x};\mathbf{w})$ より $p(d = 0|\mathbf{x}) = 1 - y(\mathbf{x};\mathbf{w})$ であることと，

$$p(d|\mathbf{x}) = p(d = 1|\mathbf{x})^d p(d = 0|\mathbf{x})^{1-d} \tag{23}$$

より，N 個の入力に対する尤度は

$$L(\mathbf{w}) = \prod_{n=1}^{N} p(d_n|\mathbf{x}_n;\mathbf{w}) = \prod_{n=1}^{N} y(\mathbf{x}_n;\mathbf{w})^{d_n}\,\{1 - y(\mathbf{x}_n;\mathbf{w})\}^{1-d_n} \tag{24}$$

となります．誤差関数としてはこれに対数をとって符号を反転させた

$$E(\mathbf{w}) = -\sum_{n=1}^{N}\left[d_n \log y(\mathbf{x}_n;\mathbf{w}) + (1 - d_n)\log\{1 - y(\mathbf{x}_n;\mathbf{w})\}\right] \tag{25}$$

を用て，パラメータ $\mathbf{w}$ を求めます．

2.4.3　多クラス分類

ネトワークの出力層にクラス数 K と同じ K 個のユニットを並べ，出力層の各ユニット $k(= 1, ..., K)$ の出力を

$$y_k = z_k^{(L)} = \frac{\exp(u_k^{(L)})}{\sum_{j=1}^{K}\exp(u_j^{(L)})} \tag{26}$$

とします．