深層学習入門

1. 機械学習

機械学習

機械学習

回帰分析

ロジスティック回帰

パーセプトロン・ADALINE

勾配降下法

ニューラルネットワーク

機械学習

機械学習

回帰分析

ロジスティック回帰

パーセプトロン・ADALINE

勾配降下法

ニューラルネットワーク

no free lunch theorem

We have dubbed the associated results NFL theorems because they demonstrate that if an algorithm performs well on a certain class of problems then it necessarily pays for that with degraded performance on the set of all remaining problems.

Wolpert and Macready,1995

機械学習

教師あり学習

教師あり学習は、訓練データとして特徴量と正解ラベルがペアとして与えられるときに行う学習方法である。機械が特徴量と正解ラベルの関係性を探索し、最適な関数で両者を結びつける。教師あり学習は、分類問題および回帰問題などに応用される。

- ニューラルネットワーク
- ロジスティック回帰
- サポートベクトルマシン（SVM）
- 決定木
- ランダムフォレスト
- k 近傍法
- 線形回帰
- スパース回帰

教師なし学習

教師なし学習は、訓練データとして特徴量のみが与えられるときに行う学習方法である。機械が大量なデータ（特徴量）を解釈し、データに隠されたパターンを抽出して、グループ分けしたりする。教師なし学習は、クラスタリングや次元削減・特徴抽出、外れ値検出などに適用される。

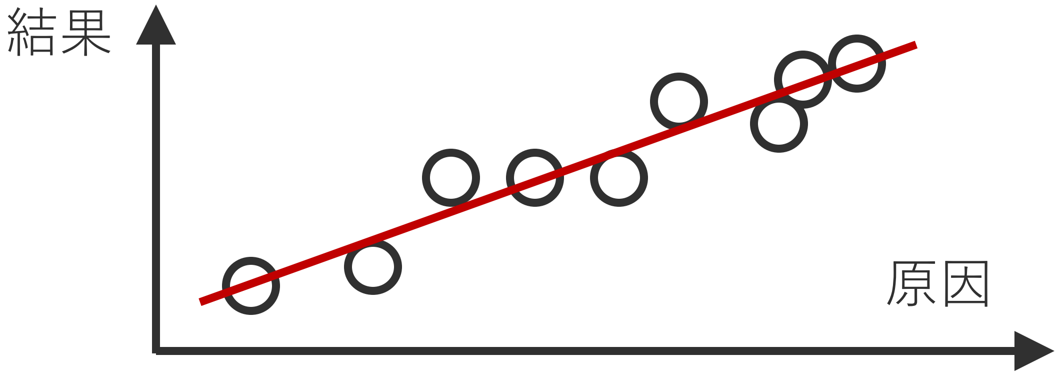
- 階層型クラスタリング
- k-means
- 自己組織化モデル（SOM）
- トピックモデル（pLSI, LDA）
- 主成分分析（PCA）
- 線形判別分析（LDA）

強化学習

最適化問題

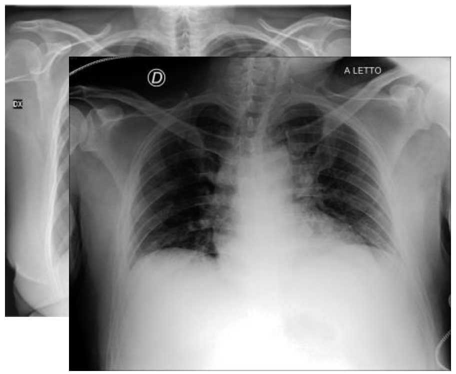
教師あり学習

回帰問題

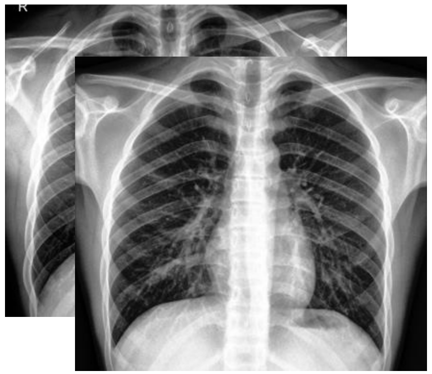


- 患者の年齢、血圧、コレステロール値、喫煙状況などのデータから心血管疾患の発症リスクを予測
- 患者の年齢、腫瘍の大きさ、病期、治療法などから治療開始後の生存期間を予測
- 患者の免疫系データやドナーとの適合性から臓器移植後の拒絶反応のリスクを予測

分類問題



COVID19



normal

- X 線画像から特定の疾患（肺結節、脳内出血、網膜疾患など）の有無を予測
- 血液検査の結果から特定の疾患の有無を予測
- 遺伝情報や臨床データを基に、患者が特定の薬剤に反応するか否かを予測

教師なし学習

クラスタリング

- 疾病データベースや電子カルテの記録から、症例の共通点をクラスタリングし、治療計画や治療法を改善に役立てる。
- 遺伝子発現データをクラスタリングして、似たような発現パターンを持つ遺伝子をグループ化し、バイオマーカーを見つける。

特徴抽出・次元削減

- 高次元データを低次元データに変換し、計算コストを下げたり、最適化過程を改善したり、可視化したりする際に利用する。

機械学習

機械学習

回帰分析

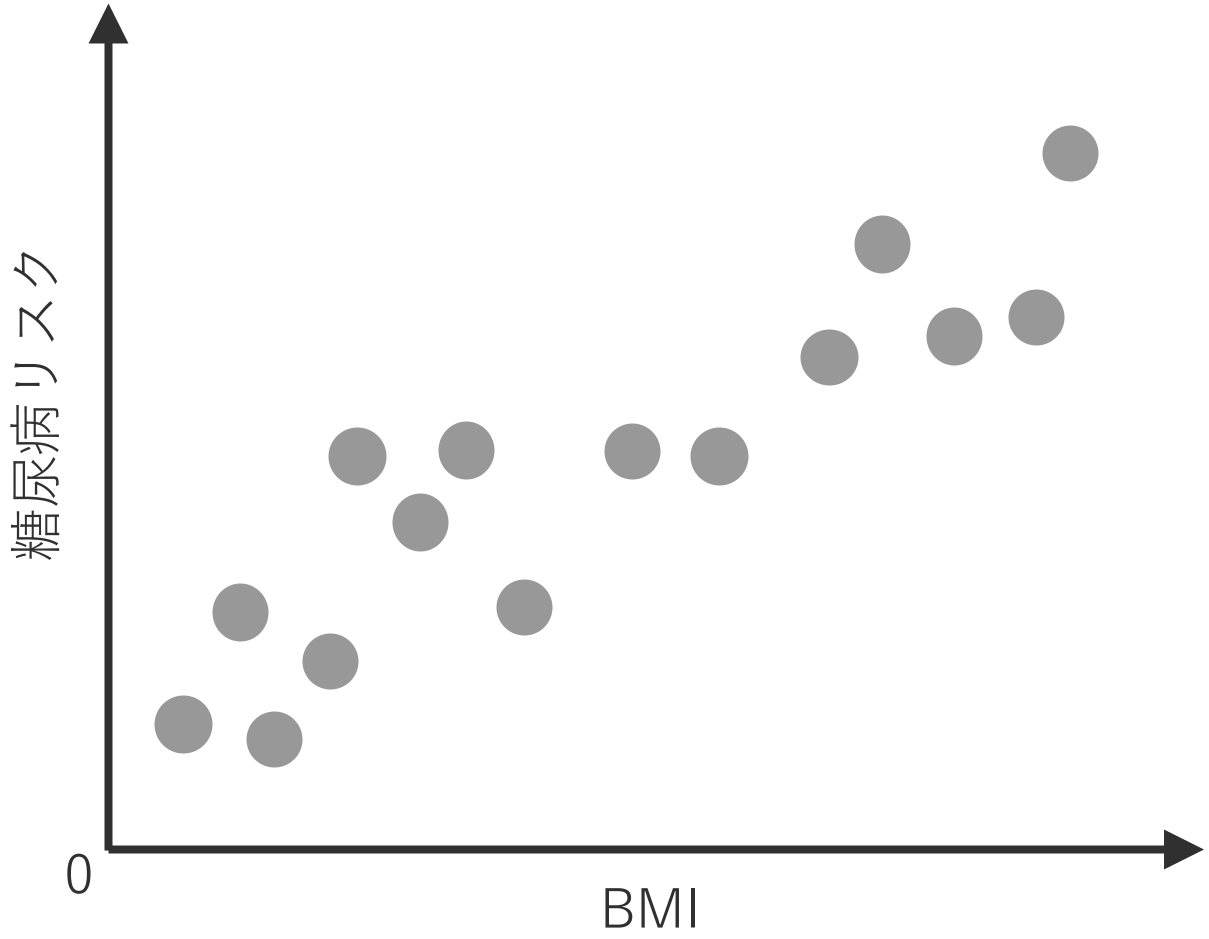
ロジスティック回帰

パーセプトロン・ADALINE

勾配降下法

ニューラルネットワーク

回帰分析

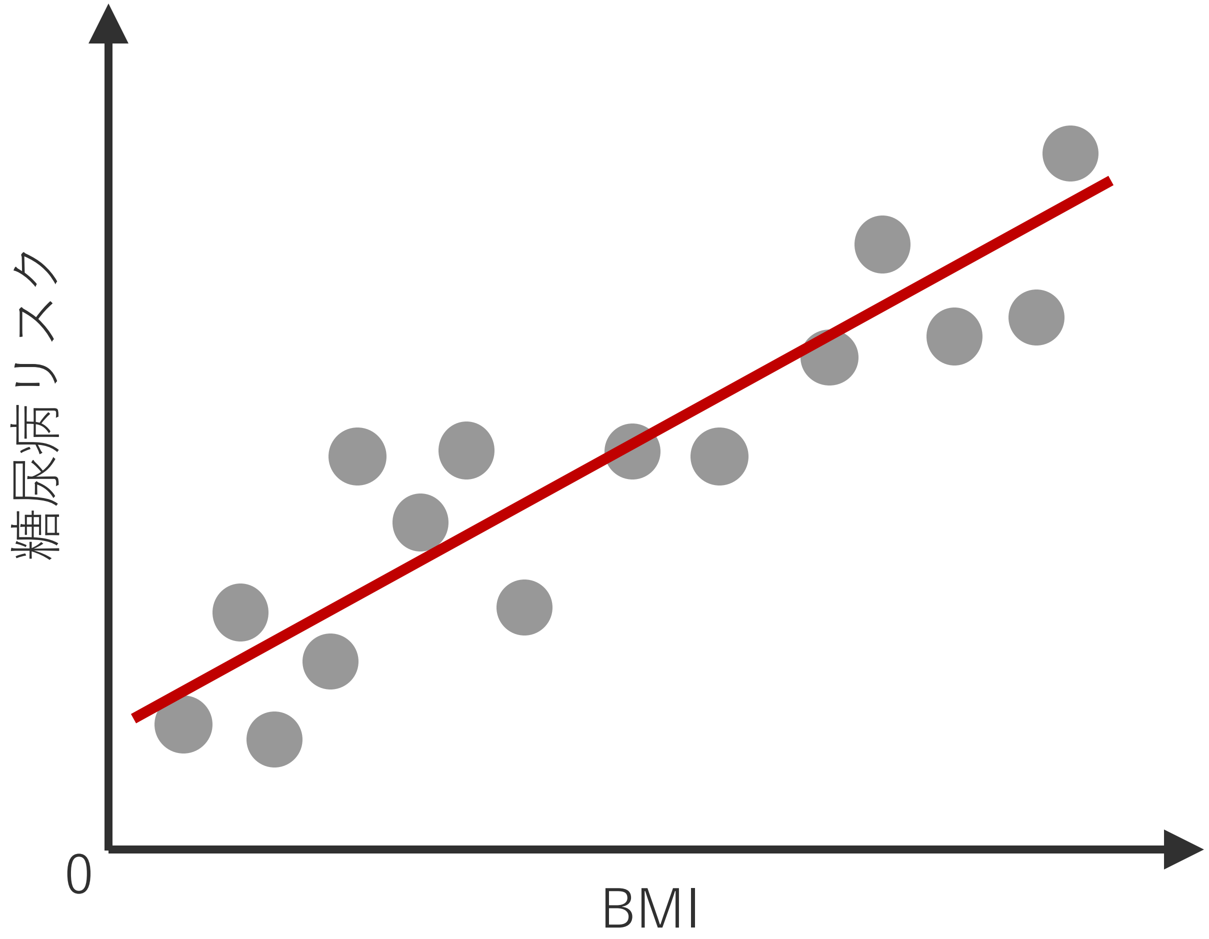


回帰分析

モデル

$$y = w_0 + w_1 x = \boldsymbol{x}\boldsymbol{w}$$

▲ 糖尿病リスク（目的変数）　▲ BMI（説明変数）



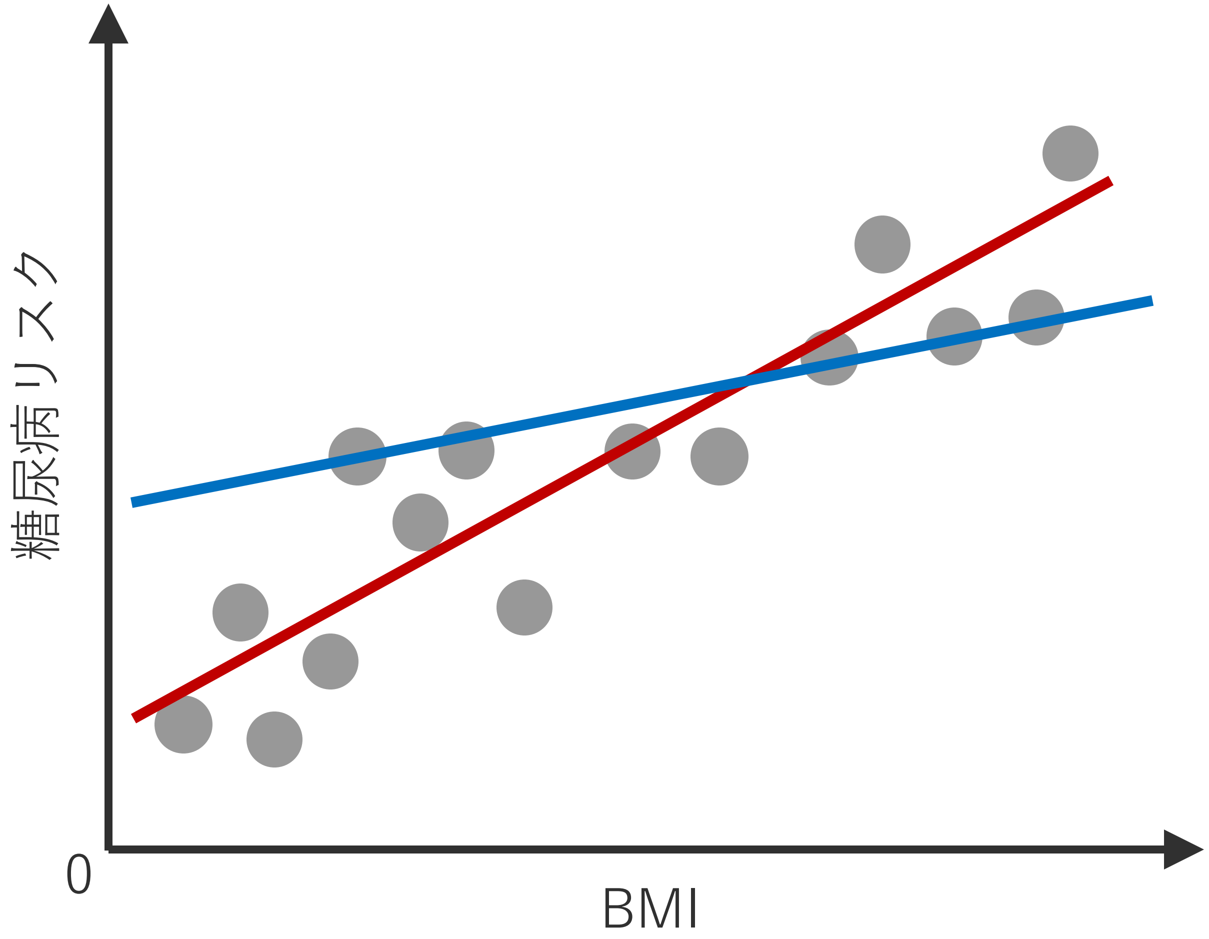
回帰分析

モデル

$$y = w_0 + w_1 x = \boldsymbol{x}\boldsymbol{w}$$

▲ 糖尿病リスク（目的変数）

▲ BMI（説明変数）



回帰分析

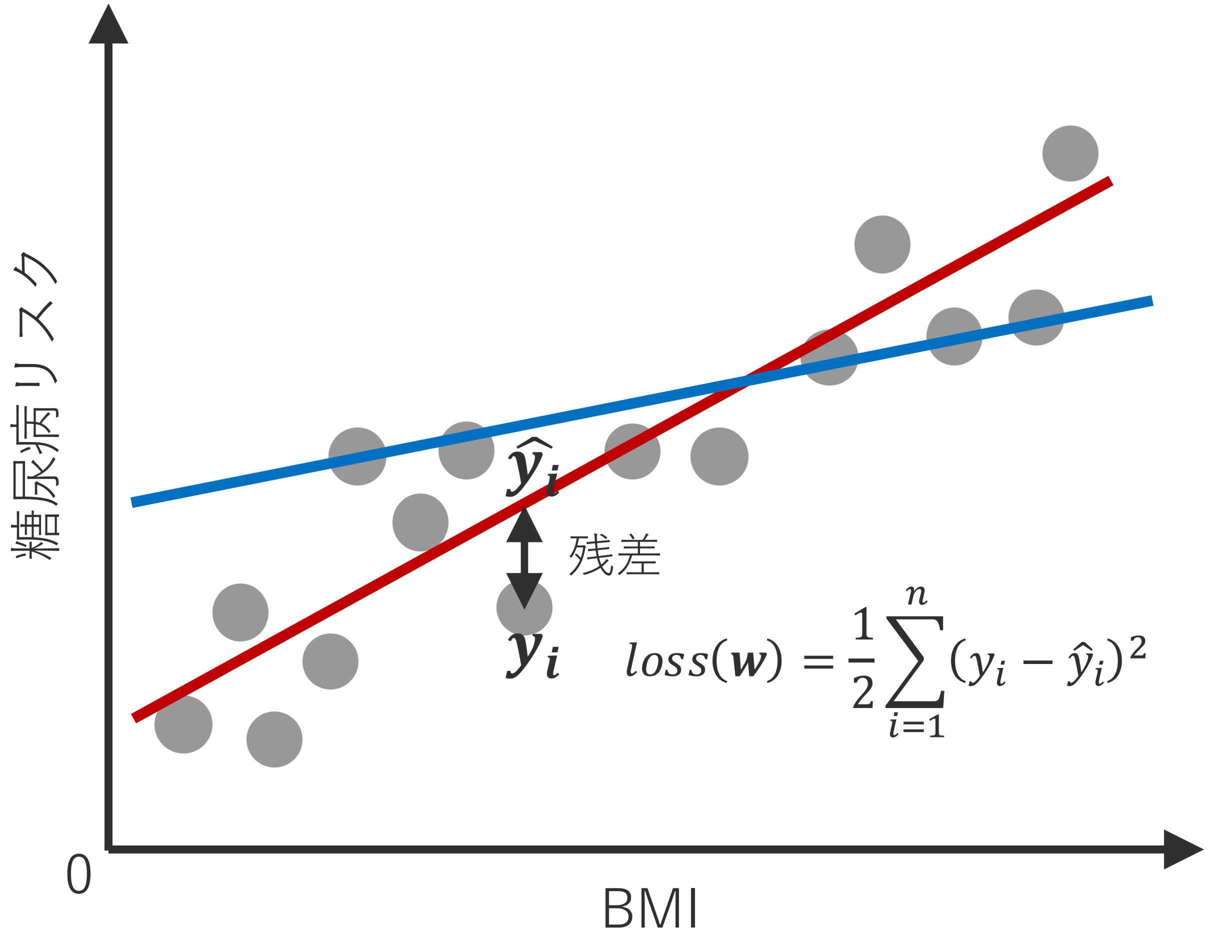
モデル

$$y = w_0 + w_1 x = \boldsymbol{x}\boldsymbol{w}$$

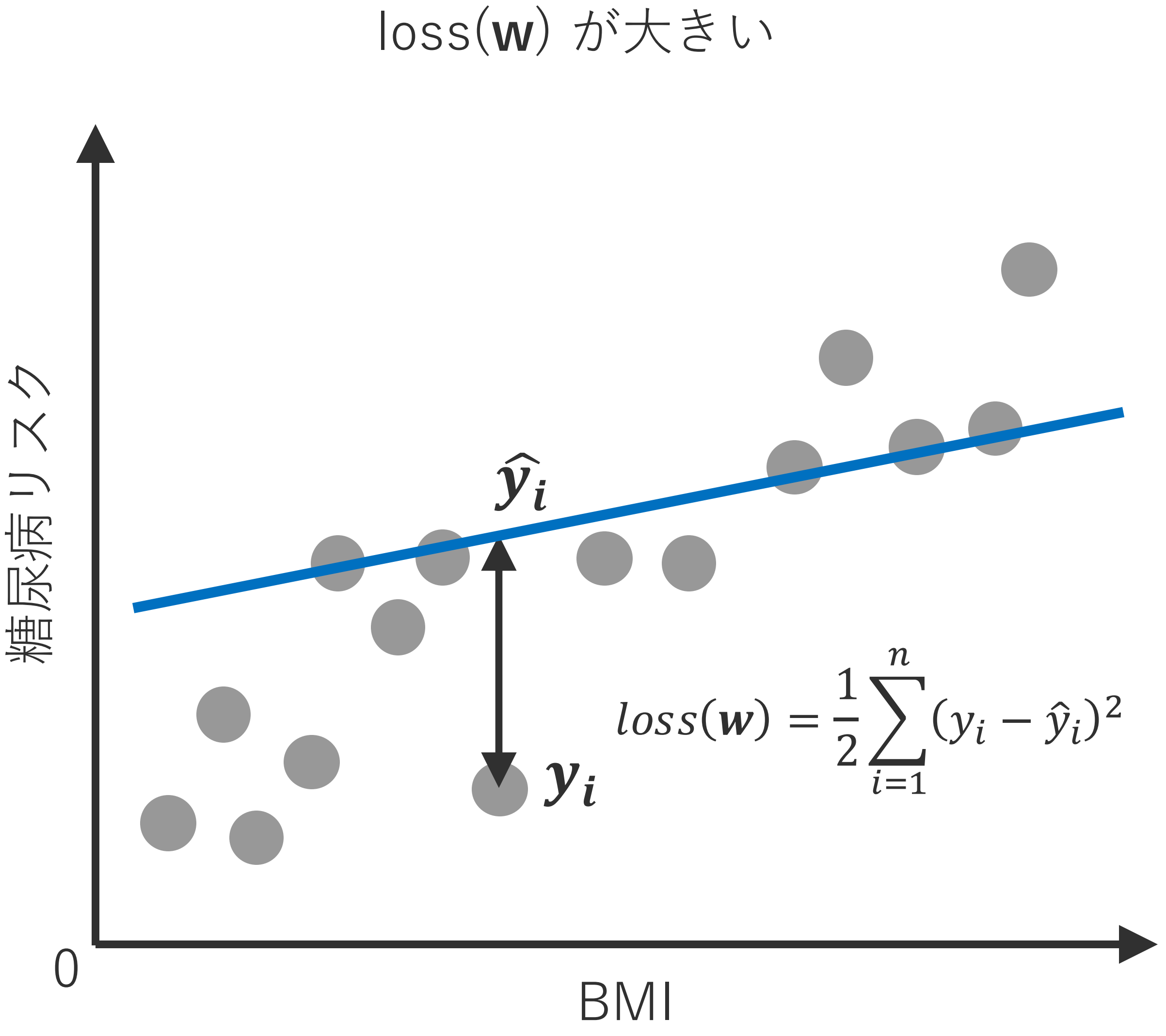
損失関数

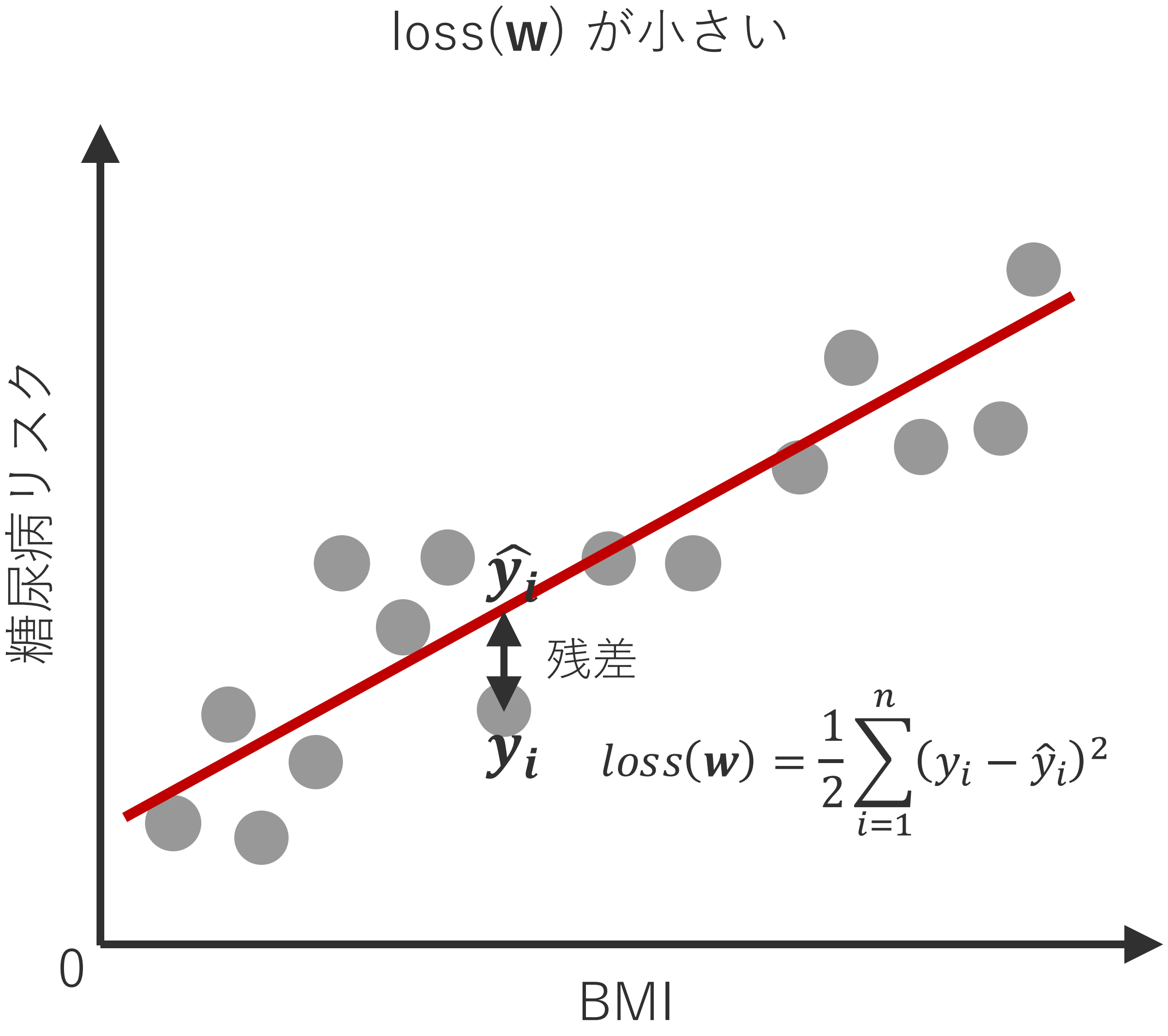
$$loss(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$= \frac{1}{2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i\boldsymbol{w})^2$$



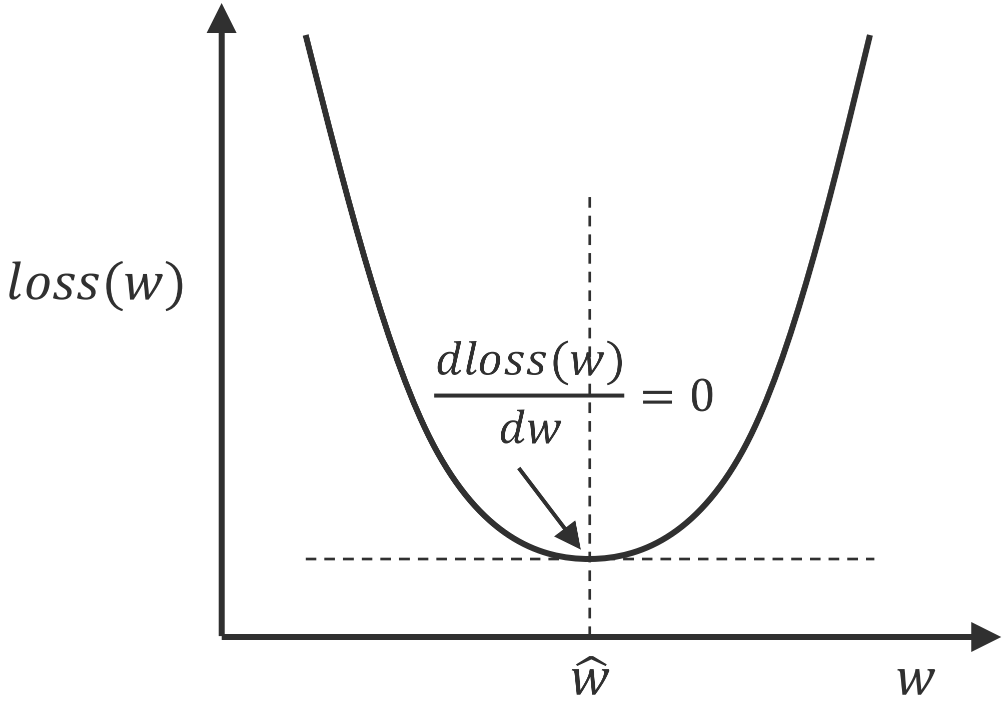
回帰分析


loss(**w**) が大きい
糖尿病リスク
0
BMI
$\hat{\boldsymbol{y}}_{\boldsymbol{i}}$
$\boldsymbol{y}_{\boldsymbol{i}}$
$loss(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$


loss(**w**) が小さい
糖尿病リスク
0
BMI
$\hat{\boldsymbol{y}}_{\boldsymbol{i}}$
残差
$\boldsymbol{y}_{\boldsymbol{i}}$
$loss(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

回帰分析

$$loss(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i\boldsymbol{w})^2$$



最小二乗法

$$\frac{\partial loss(\boldsymbol{w})}{\partial \boldsymbol{w}} = -\boldsymbol{X}^t\boldsymbol{y} + (\boldsymbol{X}^t\boldsymbol{X})^t\boldsymbol{w} = 0$$

$$\boldsymbol{w} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y}$$

- ある説明変数の傾向が他の説明変数の傾向と似ている場合、**X** の逆行列を求めることができない。
  - Ridge 回帰、スパース回帰

$$\boldsymbol{w} = (\boldsymbol{X}^t\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^t\boldsymbol{y}$$

- 標本数（**y**, **x**）が大量にあるとき、ハードウェアの制限で逆行列を求めることができない場合がある。
  - 勾配降下法

機械学習

機械学習

回帰分析

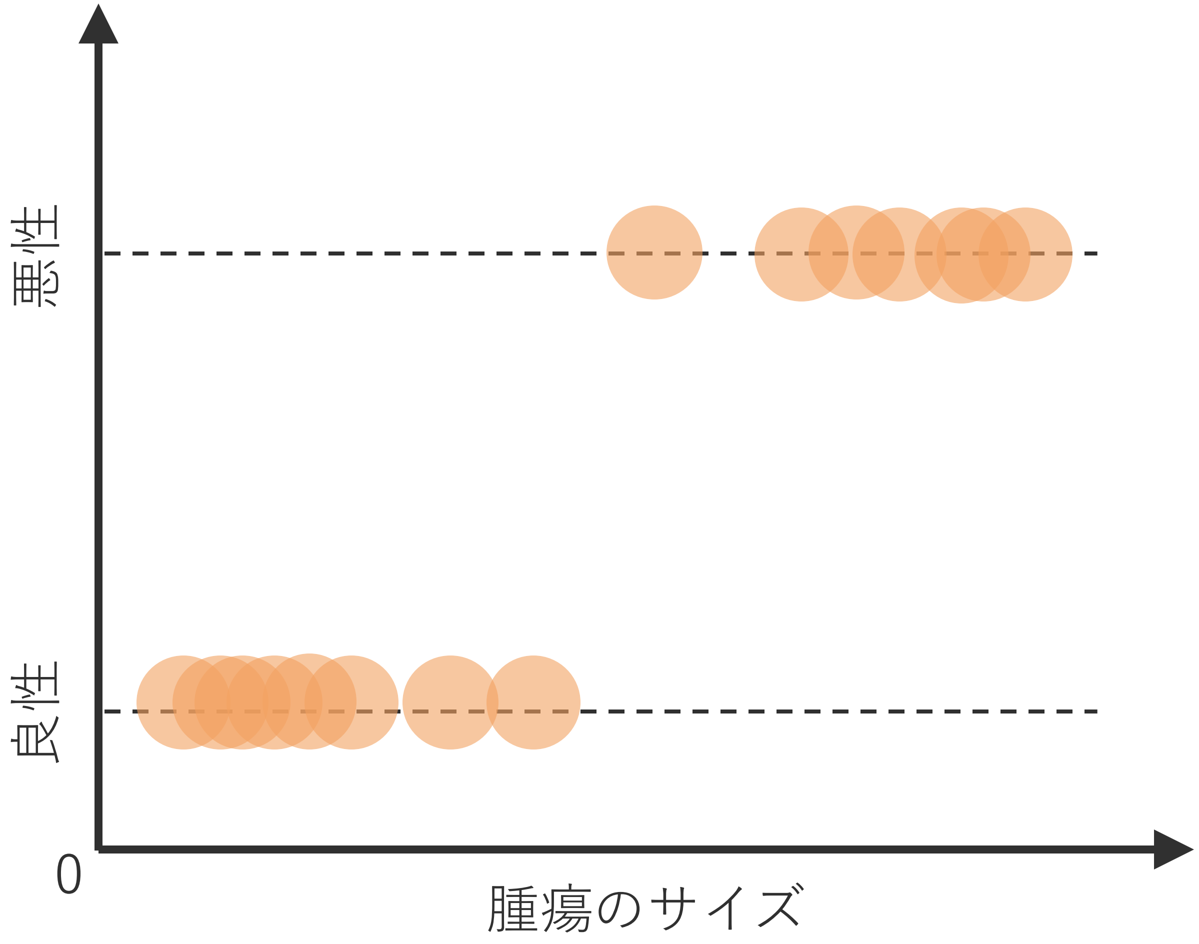
ロジスティック回帰

パーセプトロン・ADALINE

勾配降下法

ニューラルネットワーク

ロジスティック回帰



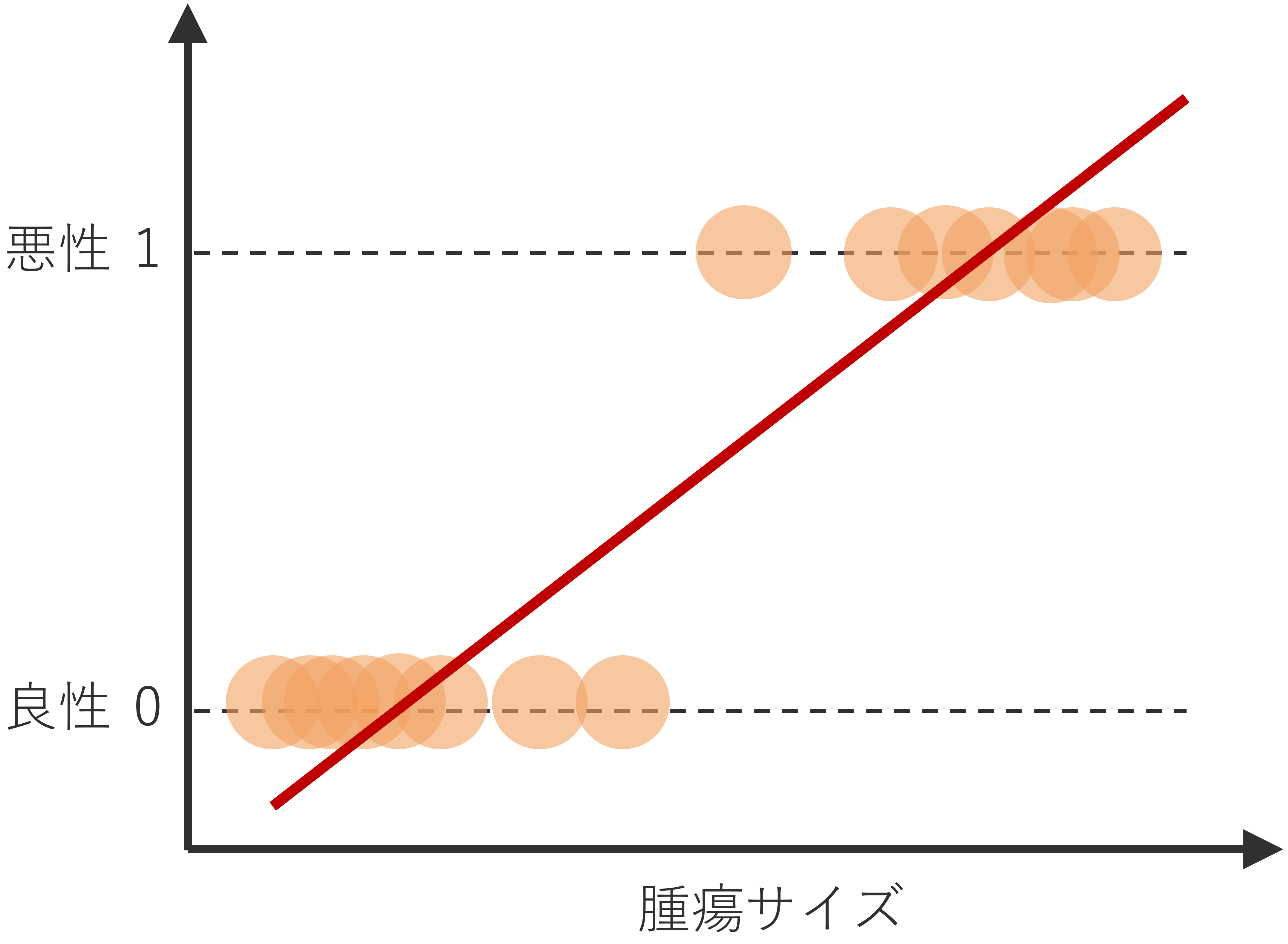
ロジスティック回帰

モデル（線形回帰）

$$y = w_0 + w_1 x = \boldsymbol{xw}$$

▲ 腫瘍種類（目的変数）

▲ 腫瘍サイズ（説明変数）



ロジスティック回帰

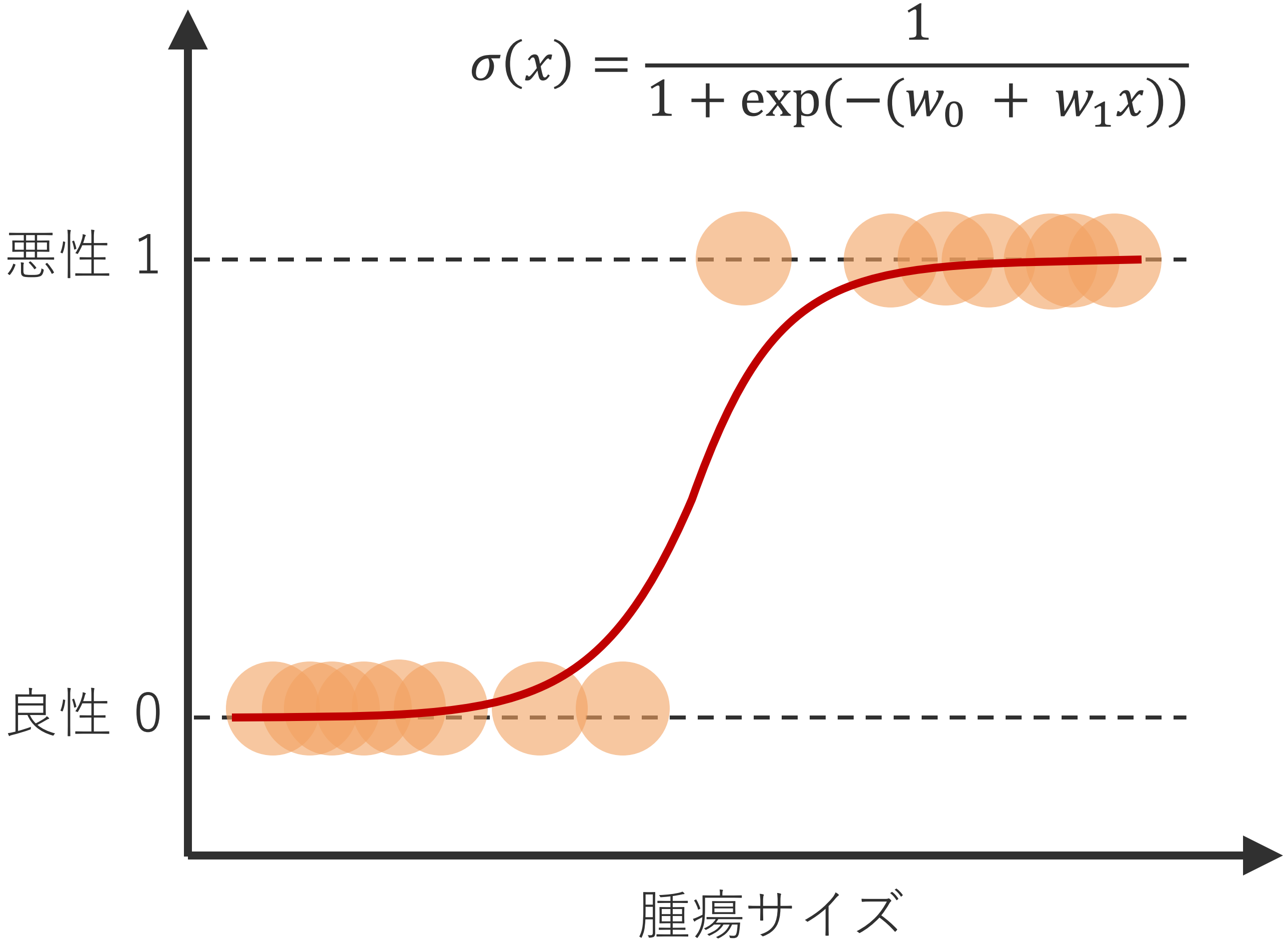
モデル（ロジスティック回帰）

腫瘍種類
（目的変数）

$$y = \frac{1}{1 + \exp(-(w_0 + w_1 x))}$$

$$= \frac{1}{1 + \exp(-(\boldsymbol{w}\boldsymbol{x}))}$$

腫瘍サイズ
（説明変数）



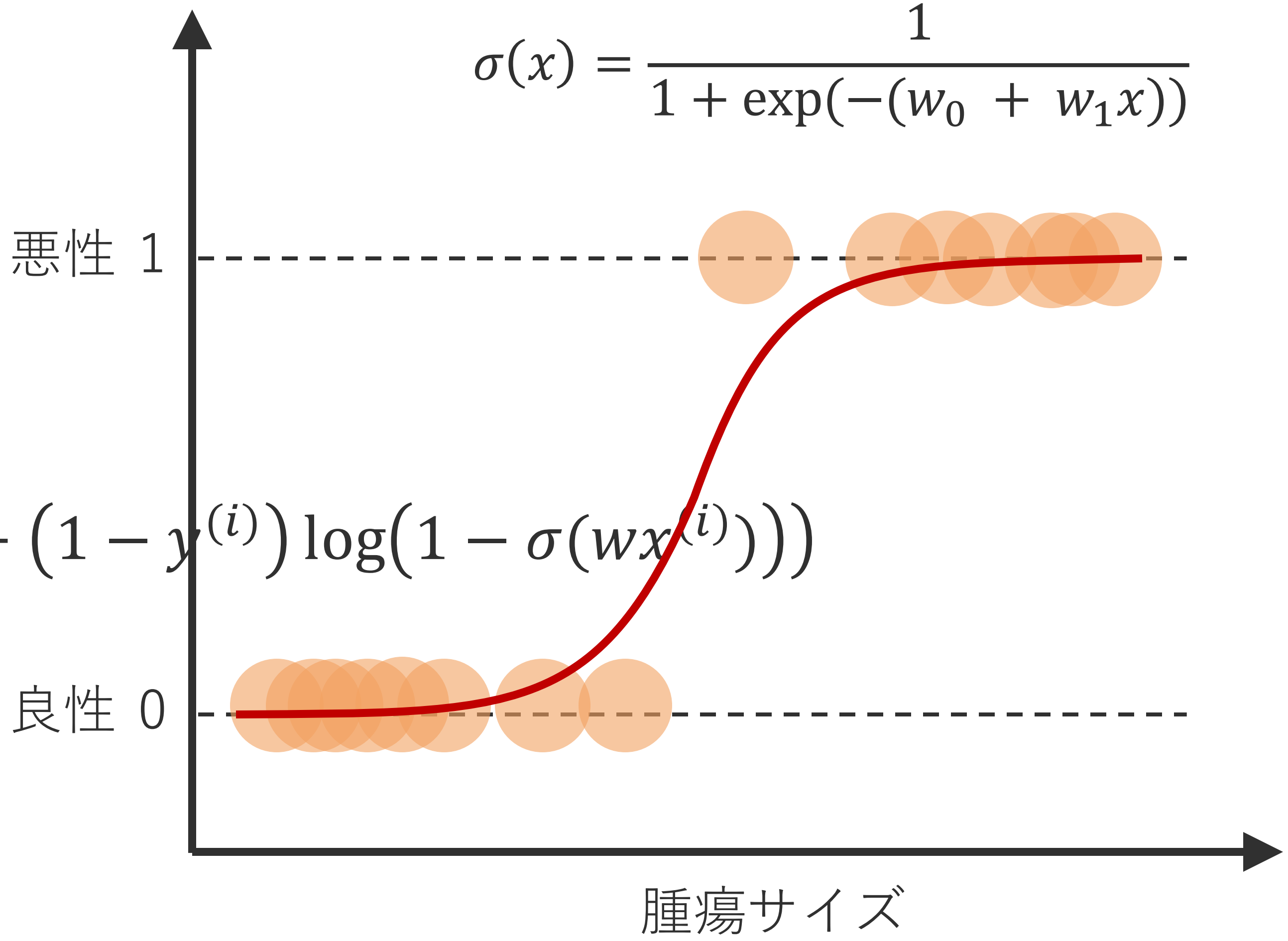
ロジスティック回帰

モデル（ロジスティック回帰）

$$y = \frac{1}{1 + \exp(-(\boldsymbol{wx}))}$$

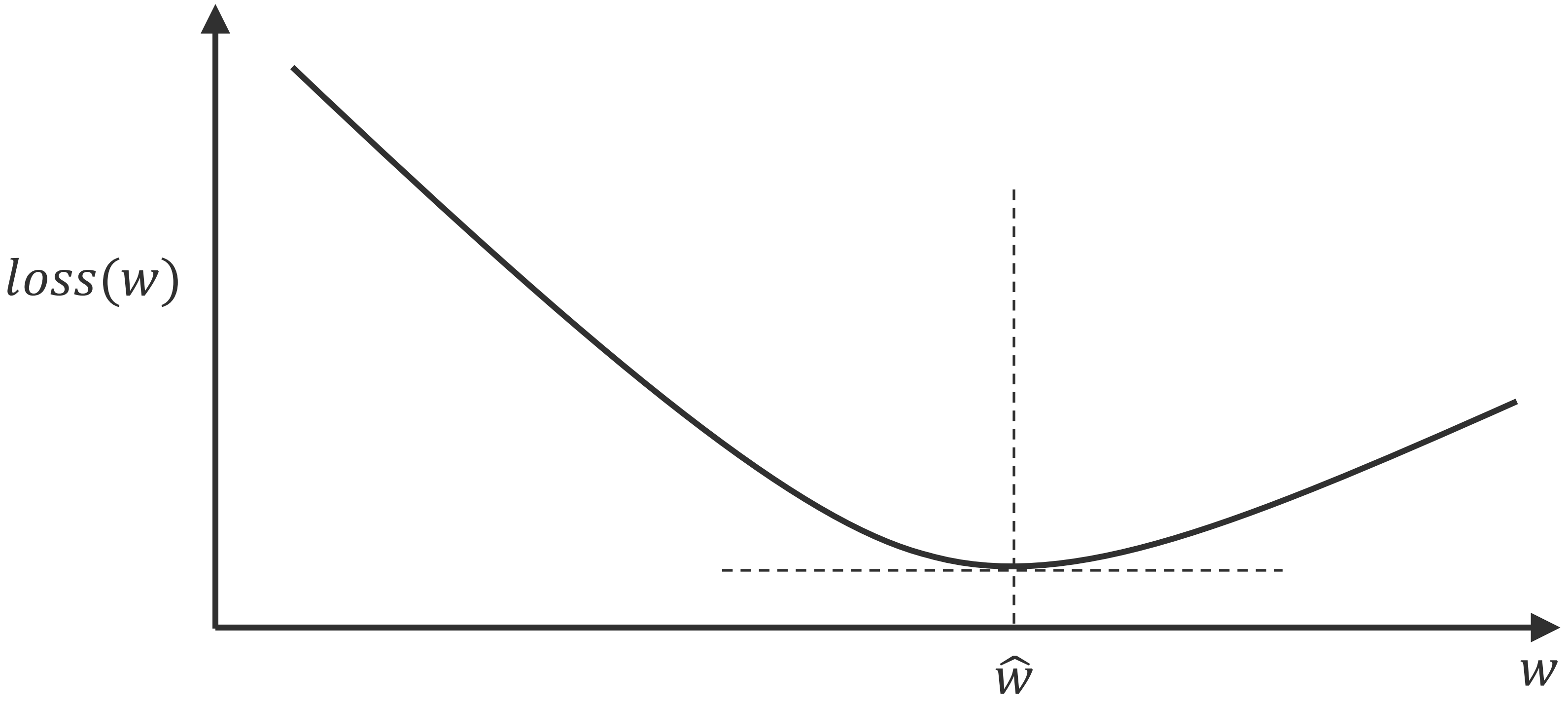
損失関数

$$loss(w) = -\sum_{i=1}^{n}\left(y^{(i)}\log\left(\sigma(wx^{(i)})\right) + \left(1 - y^{(i)}\right)\log\left(1 - \sigma(wx^{(i)})\right)\right)$$



ロジスティック回帰

$$loss(w) = -\sum_{i=1}^{n}\left(y^{(i)}\log\left(\sigma\left(wx^{(i)}\right)\right) + \left(1-y^{(i)}\right)\log\left(1-\sigma\left(wx^{(i)}\right)\right)\right)$$



機械学習

機械学習

回帰分析

ロジスティック回帰

パーセプトロン・ADALINE

勾配降下法

ニューラルネットワーク

パーセプトロン

入力 $x_1 \quad x_2$

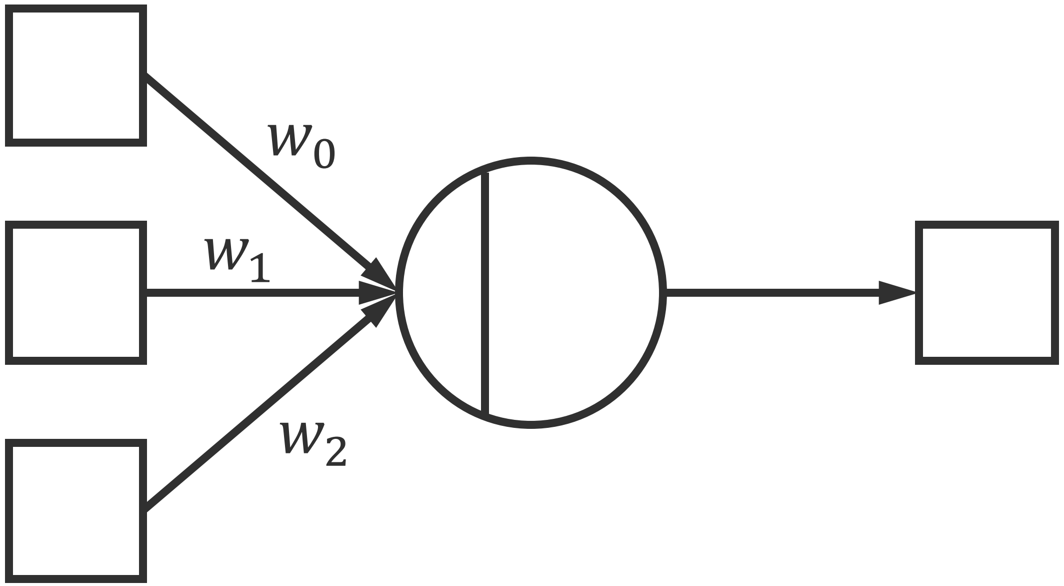
処理

$$z = w_0 + w_1 x_1 + w_2 x_2$$

処理

$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

出力 1



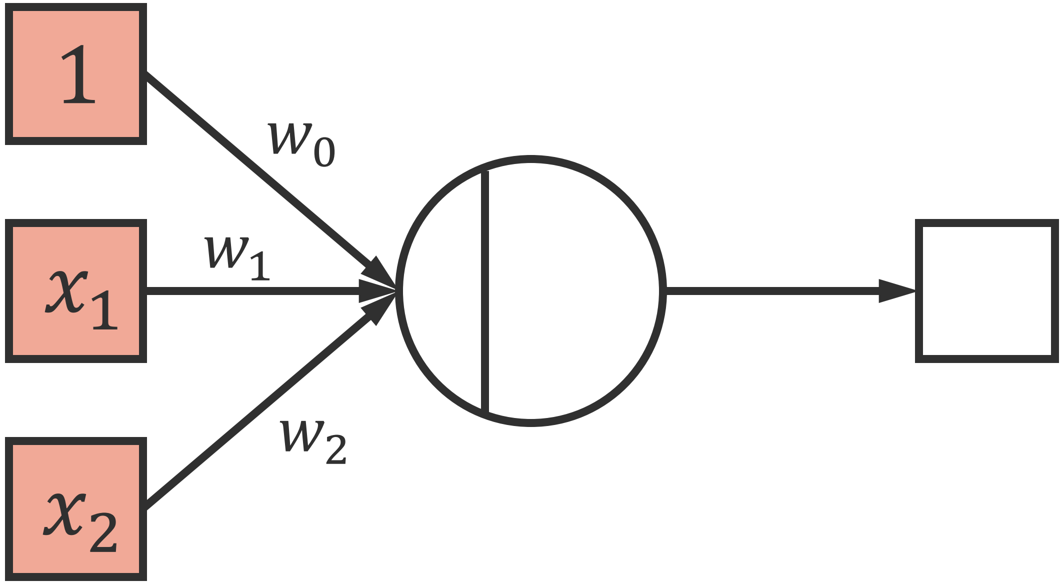
パーセプトロン

入力 $x_1 \quad x_2$

処理 $z = w_0 + w_1 x_1 + w_2 x_2$

処理 $\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$

出力 1



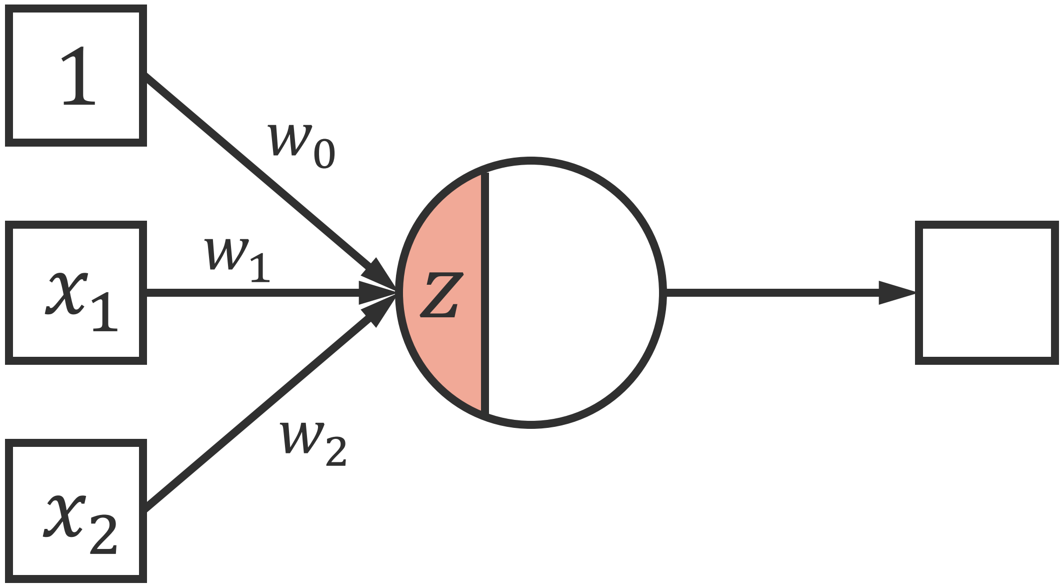
パーセプトロン

入力 $x_1 \quad x_2$

処理 $z = w_0 + w_1 x_1 + w_2 x_2$

処理 $\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$

出力 1



パーセプトロン

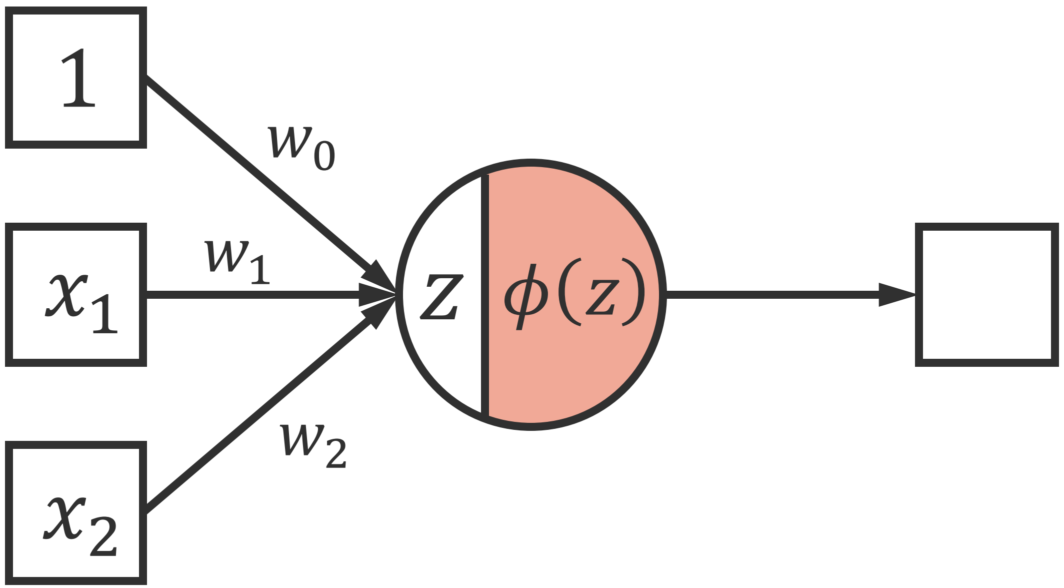
入力　$x_1 \quad x_2$

処理　$z = w_0 + w_1 x_1 + w_2 x_2$

処理

$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

出力　1



パーセプトロン

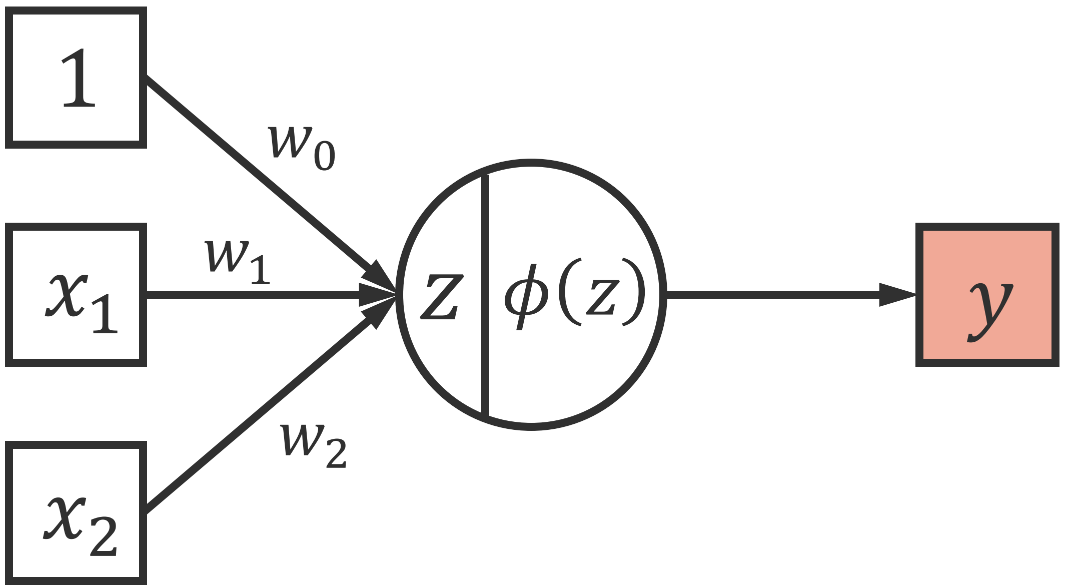
入力 $x_1 \quad x_2$

処理 $z = w_0 + w_1 x_1 + w_2 x_2$

処理

$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

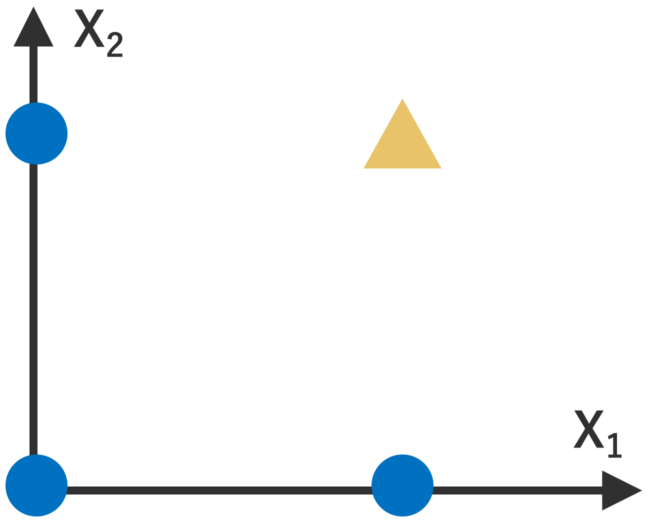
出力 y

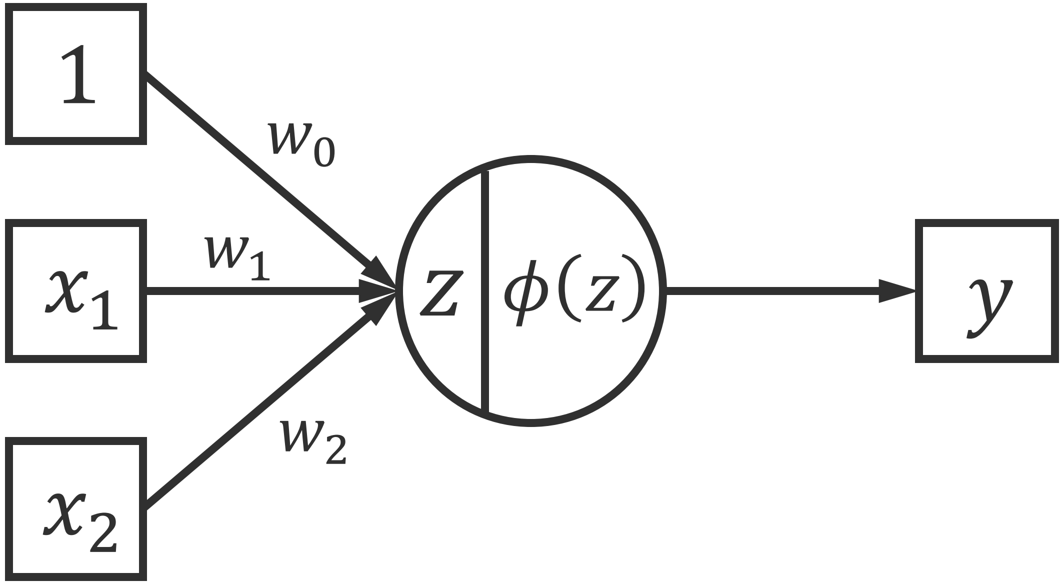


パーセプトロン学習則

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性

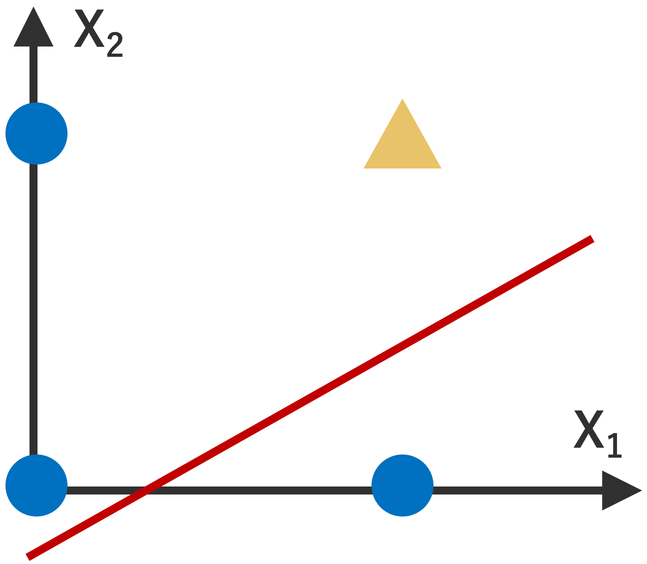




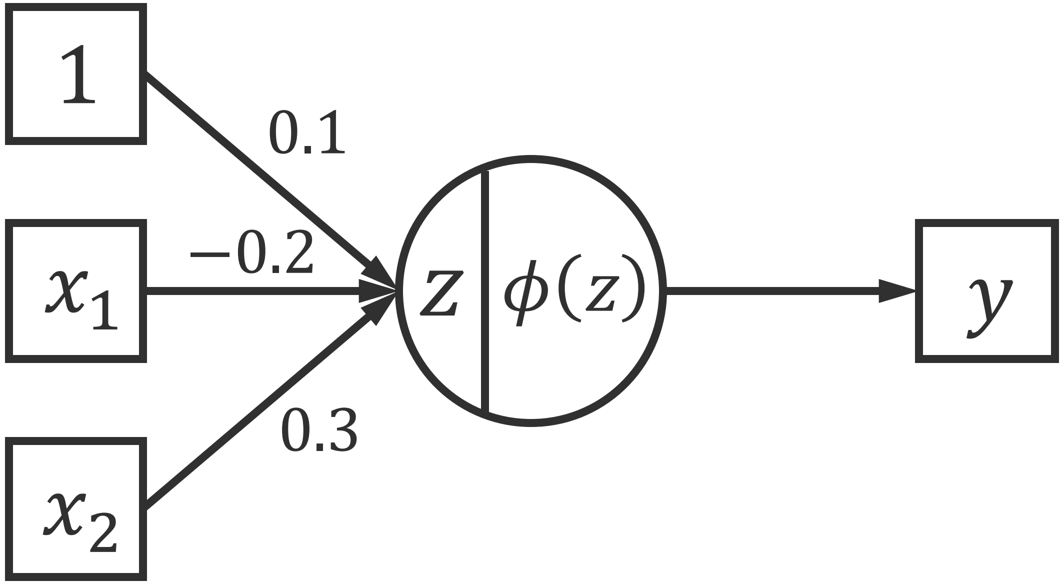
パーセプトロン学習則

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性



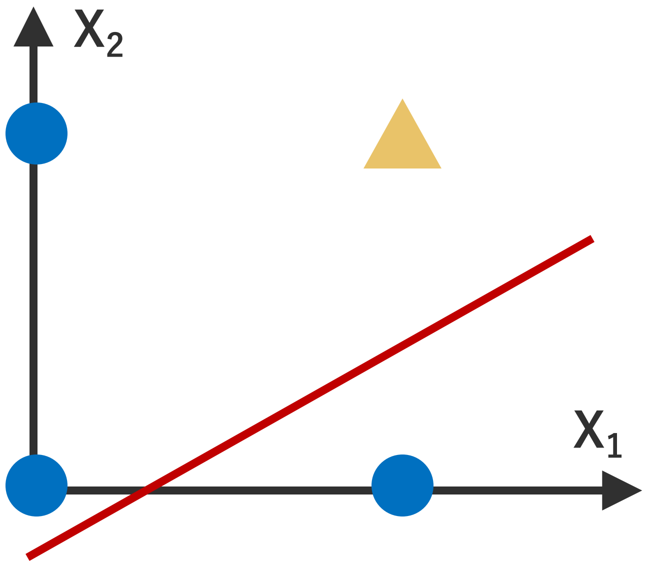
初期化

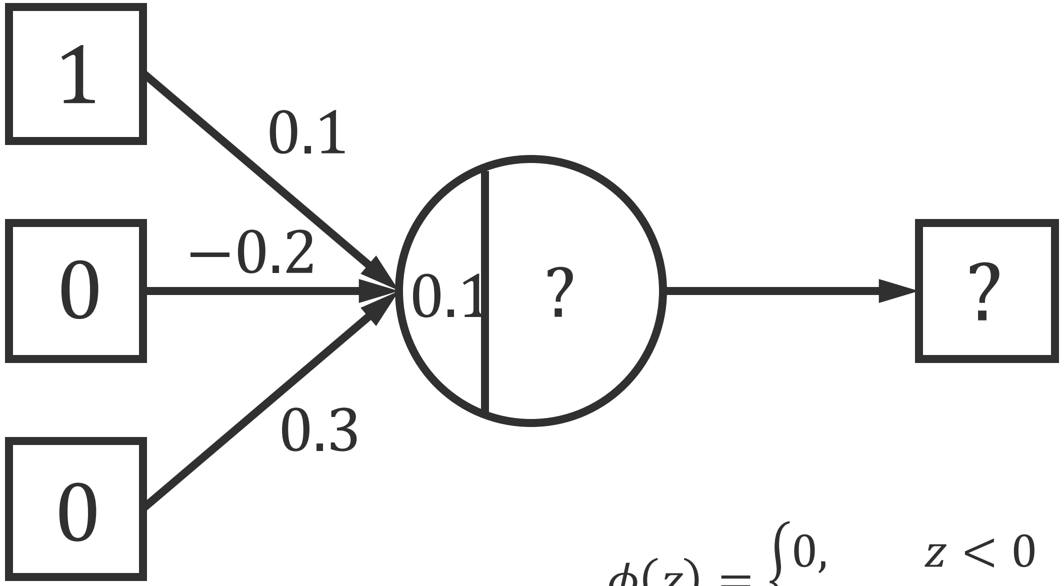


パーセプトロン学習則

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性



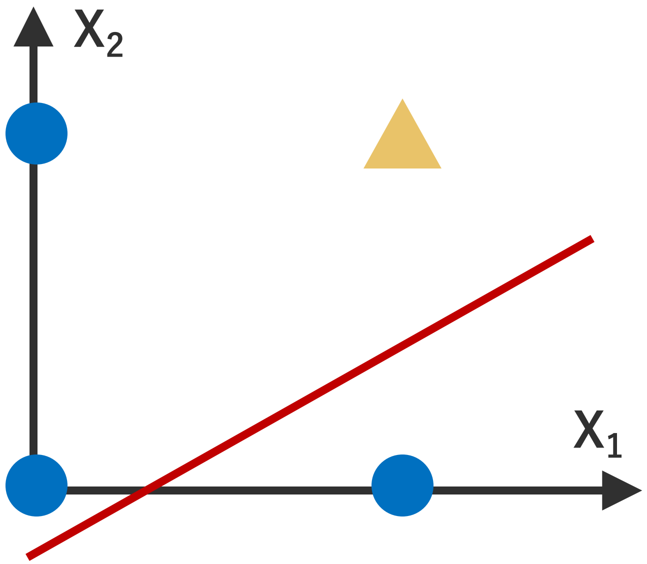


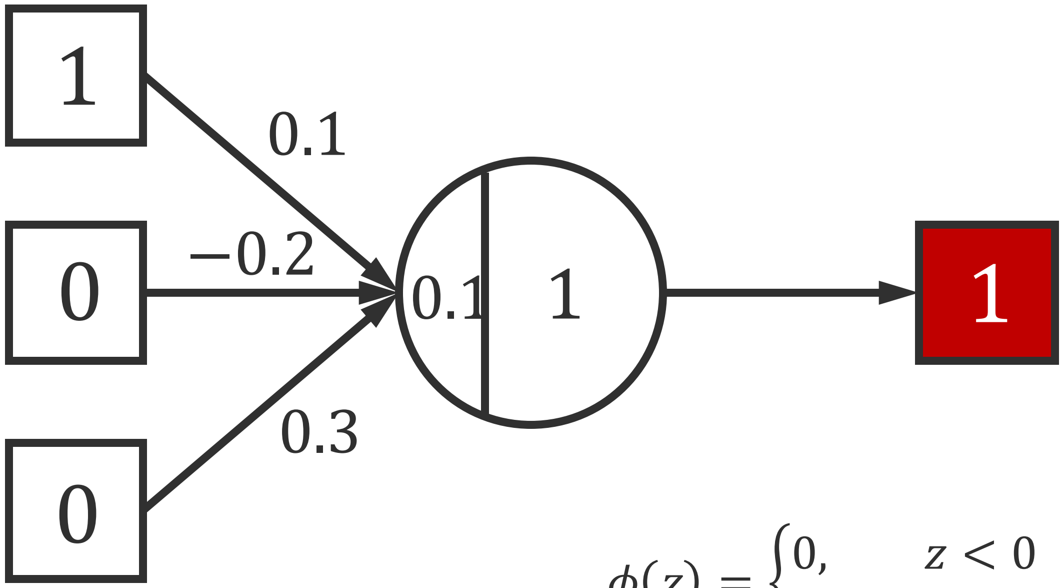
$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

パーセプトロン学習則

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性





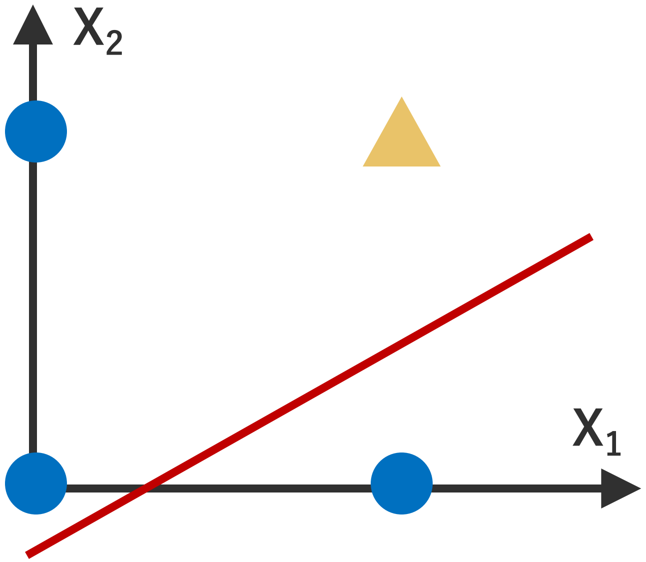
$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

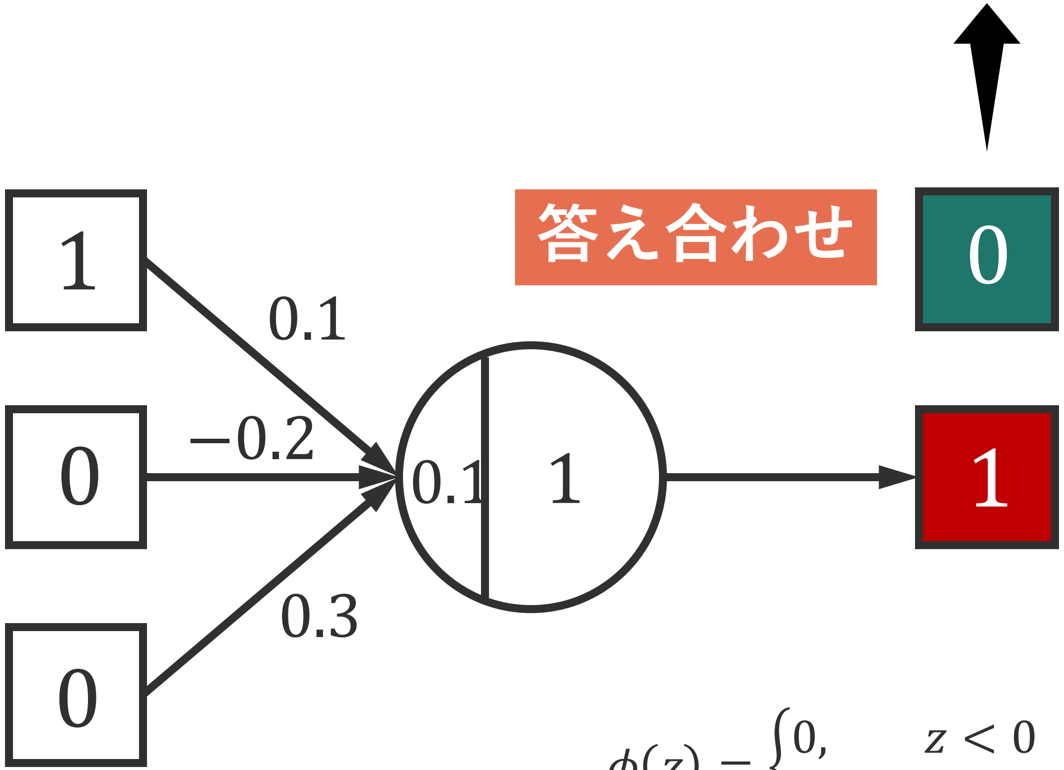
パーセプトロン学習則

$$w_0^{(new)} = 0.1 + 1(0 - 1)0.1 = 0.0$$
$$w_1^{(new)} = -0.2 + 0(0 - 1)0.1 = -0.2$$
$$w_2^{(new)} = 0.3 + 0(0 - 1)0.1 = 0.3$$

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性





$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

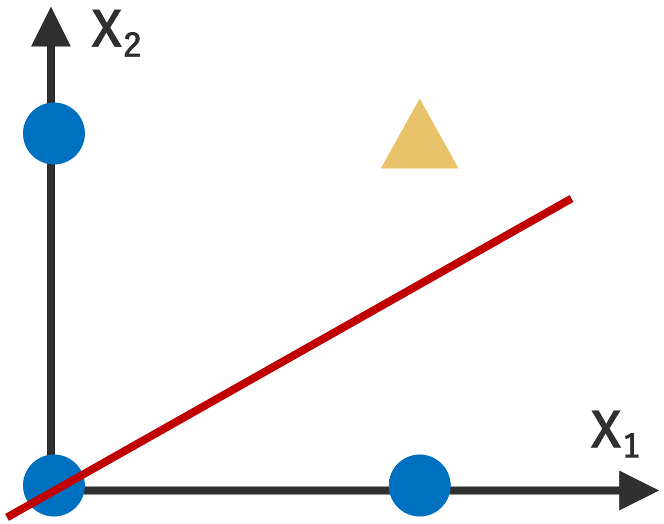
パーセプトロン学習則

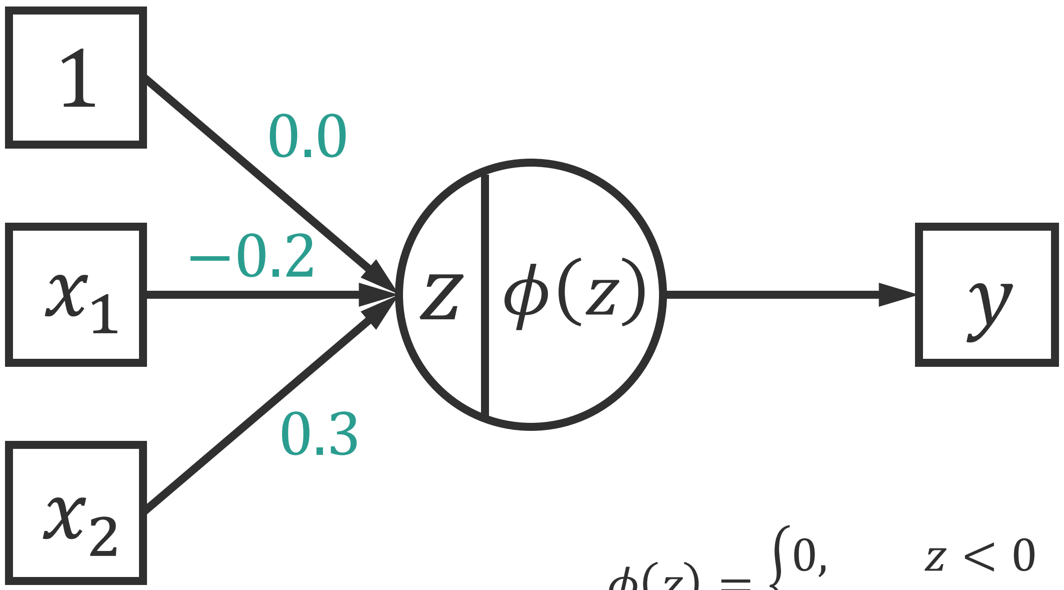
$$w_0^{(new)} = 0.1 + 1(\mathbf{0} - \mathbf{1})0.1 = 0.0$$
$$w_1^{(new)} = -0.2 + 0(\mathbf{0} - \mathbf{1})0.1 = -0.2$$
$$w_2^{(new)} = 0.3 + 0(\mathbf{0} - \mathbf{1})0.1 = 0.3$$

パラメータ更新

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性



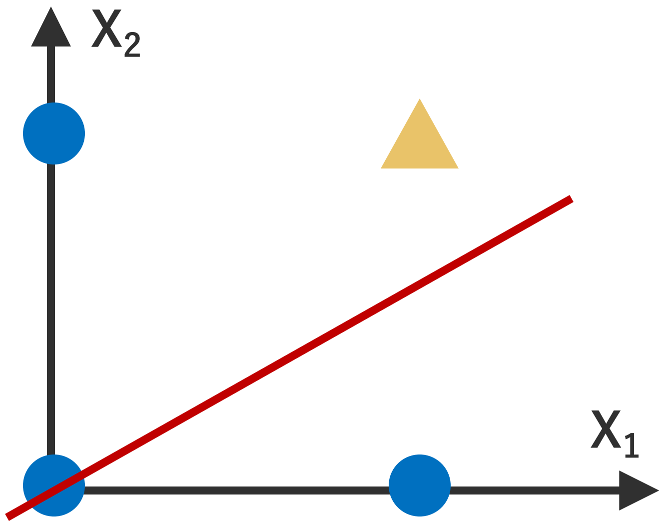


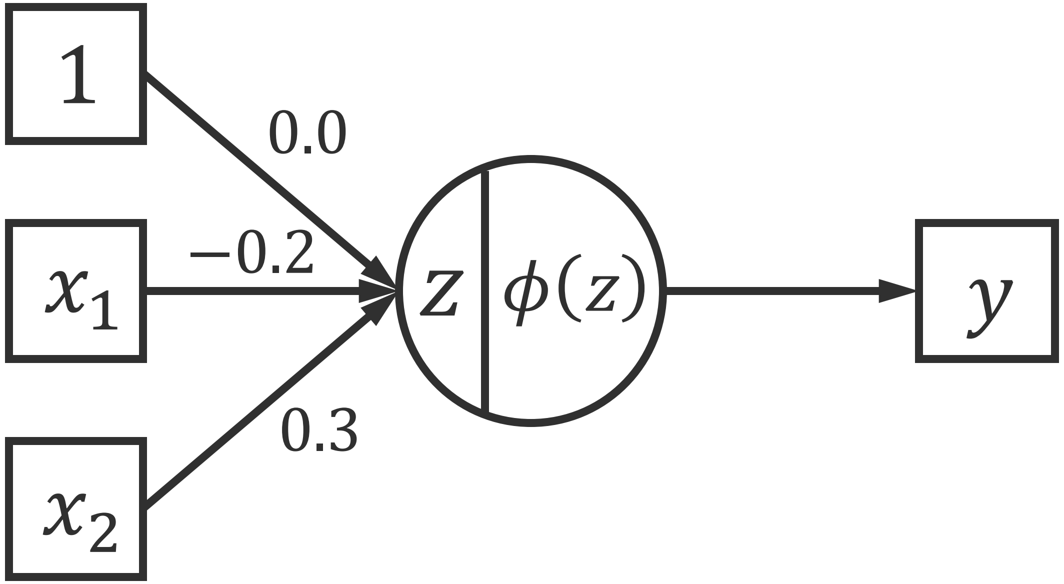
$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

パーセプトロン学習則

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性



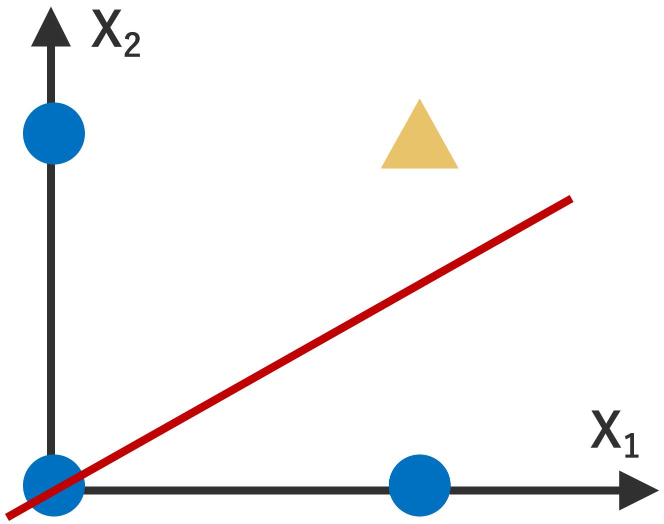


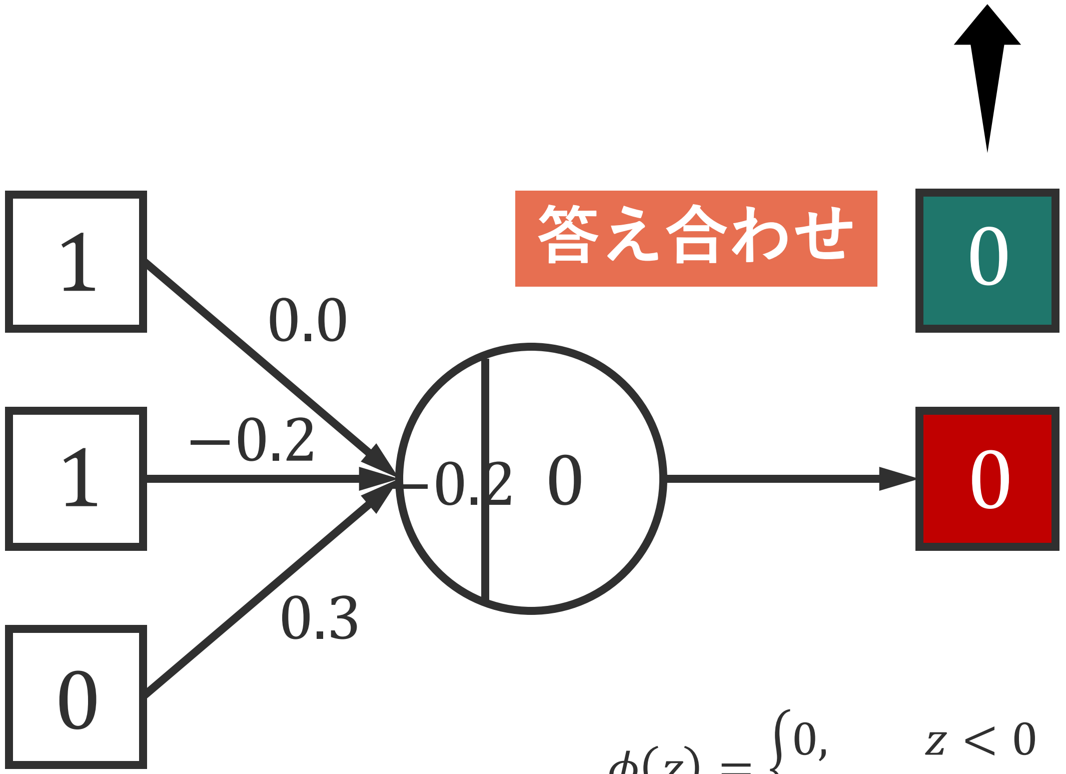
パーセプトロン学習則

$$w_0^{(new)} = 0.0 + 1(\mathbf{0} - \mathbf{0})0.1 = 0.0$$
$$w_1^{(new)} = -0.2 + 1(\mathbf{0} - \mathbf{0})0.1 = -0.2$$
$$w_2^{(new)} = 0.3 + 0(\mathbf{0} - \mathbf{0})0.1 = 0.3$$

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性





$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

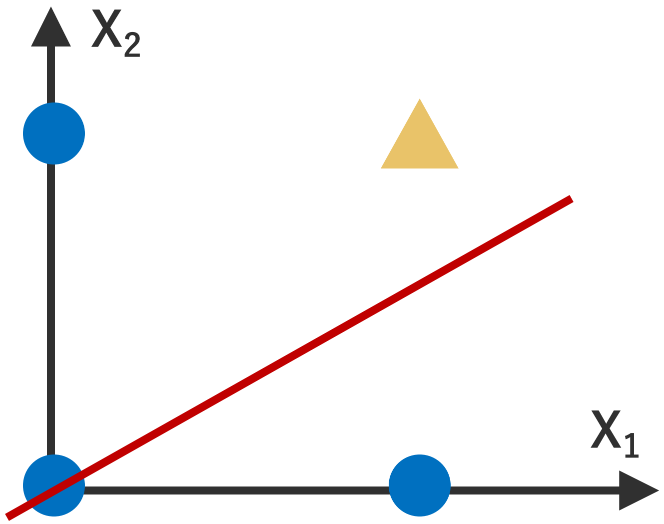
パーセプトロン学習則

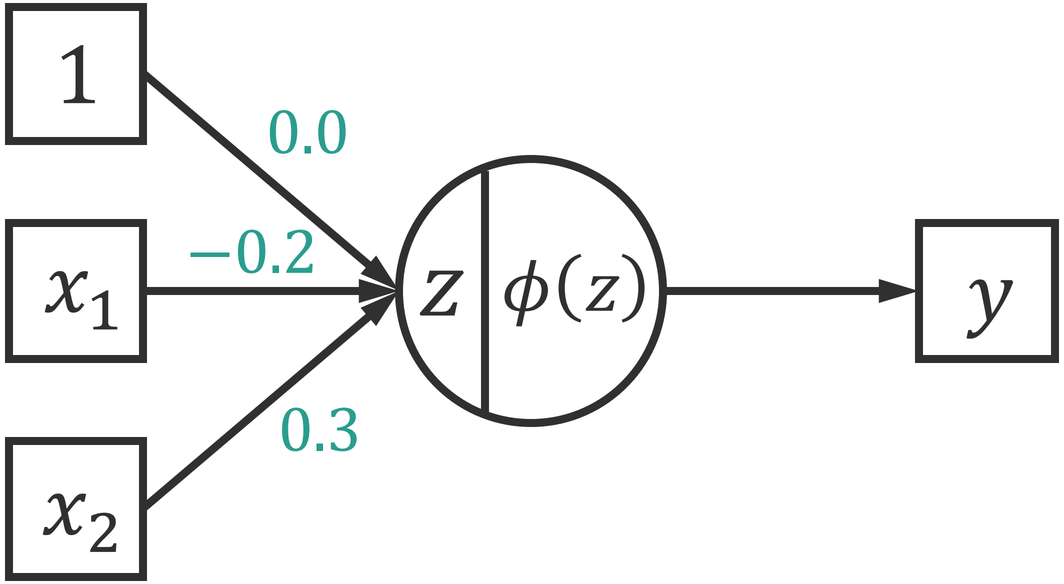
$$w_0^{(new)} = 0.0 + 1(\mathbf{0} - \mathbf{0})0.1 = 0.0$$
$$w_1^{(new)} = -0.2 + 1(\mathbf{0} - \mathbf{0})0.1 = -0.2$$
$$w_2^{(new)} = 0.3 + 0(\mathbf{0} - \mathbf{0})0.1 = 0.3$$

パラメータ更新

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性

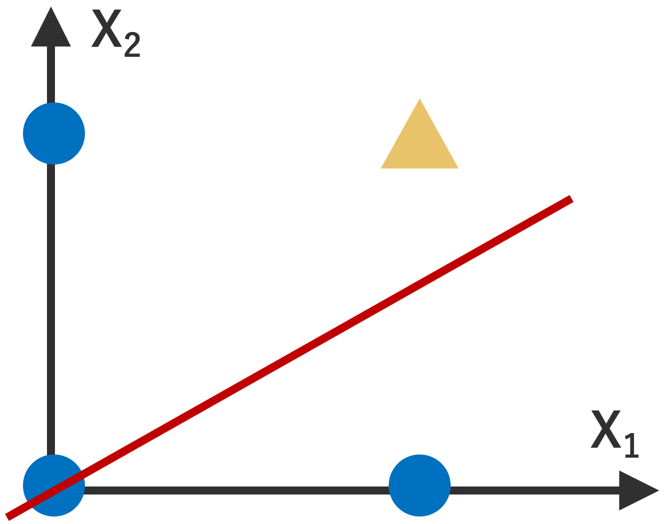


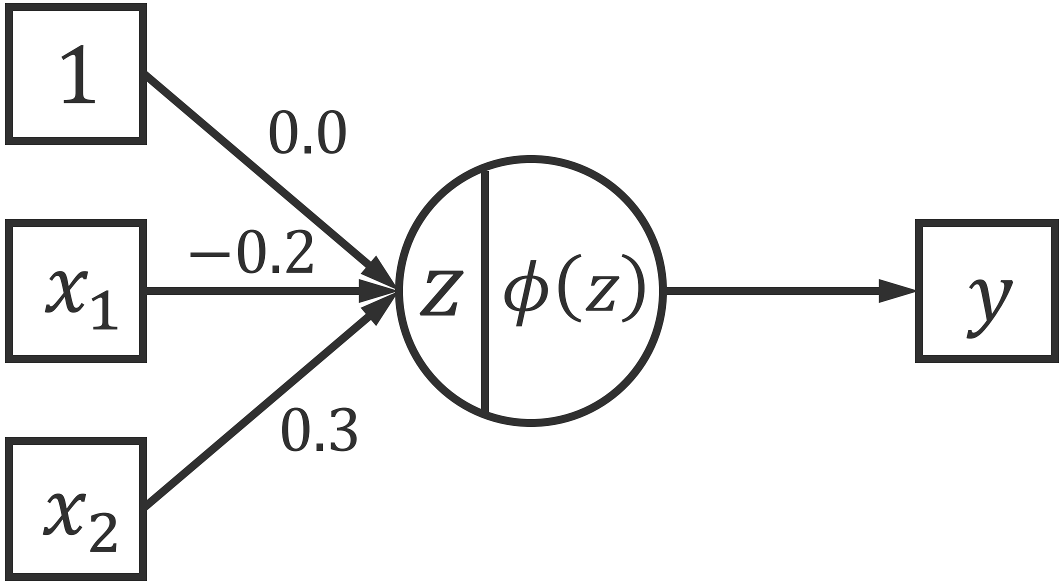


パーセプトロン学習則

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性





パーセプトロン学習則

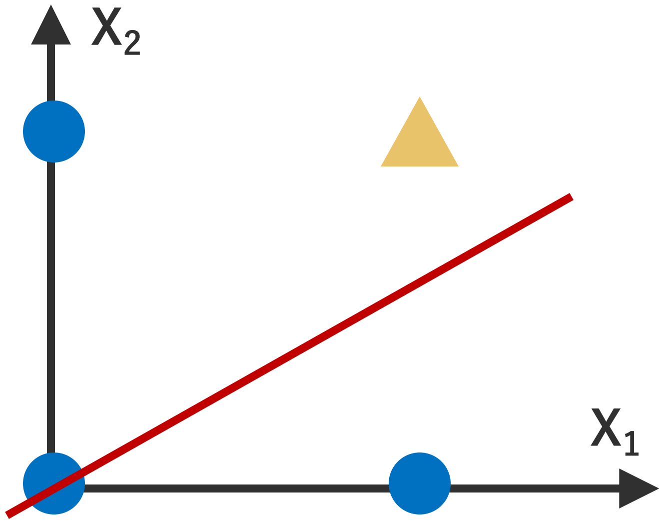
$$w_0^{(new)} = 0.0 + 1(\mathbf{0} - \mathbf{1})0.1 = -0.1$$

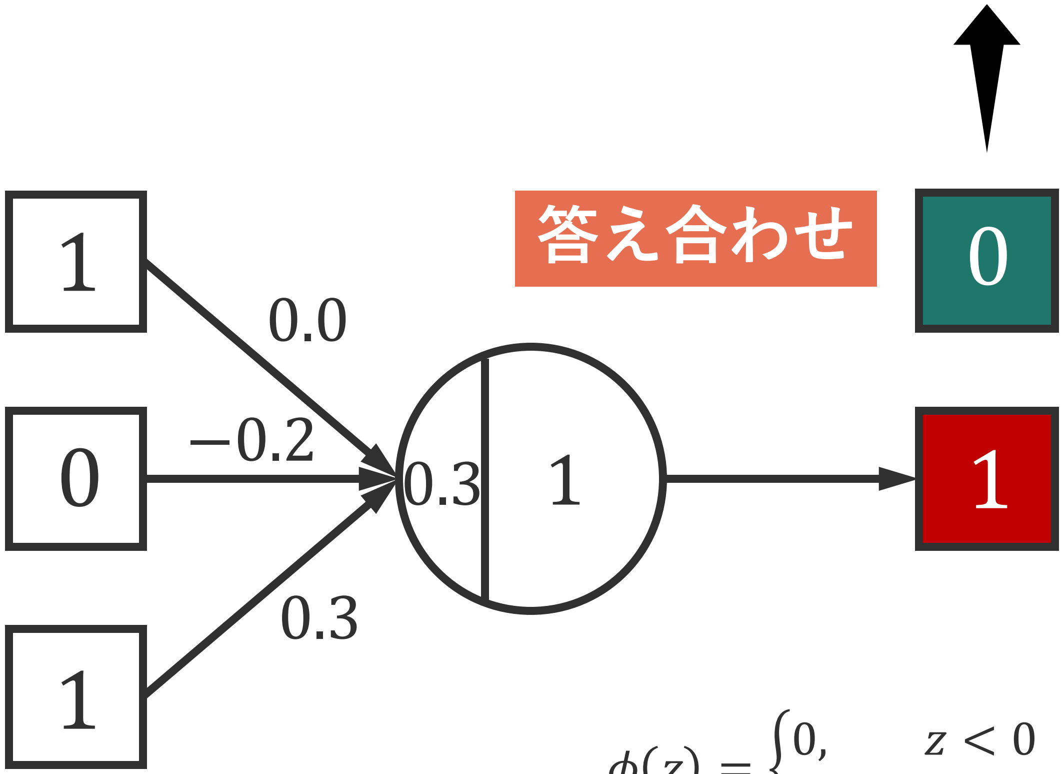
$$w_1^{(new)} = -0.2 + 0(\mathbf{0} - \mathbf{1})0.1 = -0.2$$

$$w_2^{(new)} = 0.3 + 1(\mathbf{0} - \mathbf{1})0.1 = 0.2$$

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性





$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

パーセプトロン学習則

$$w_0^{(new)} = 0.0 + 1(\mathbf{0} - \mathbf{1})0.1 = -0.1$$

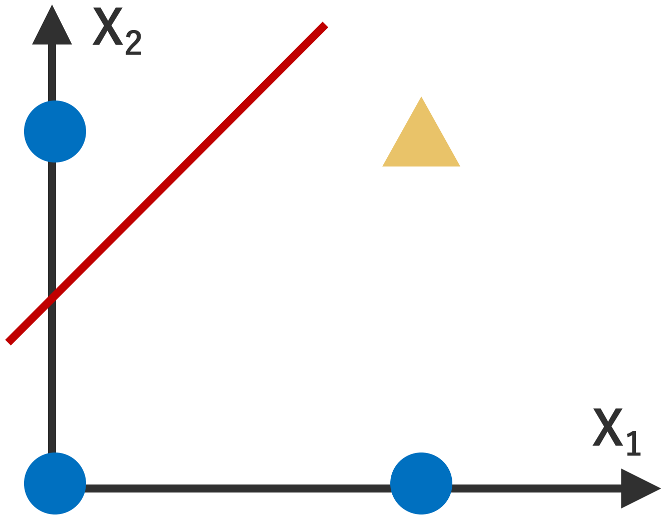
$$w_1^{(new)} = -0.2 + 0(\mathbf{0} - \mathbf{1})0.1 = -0.2$$

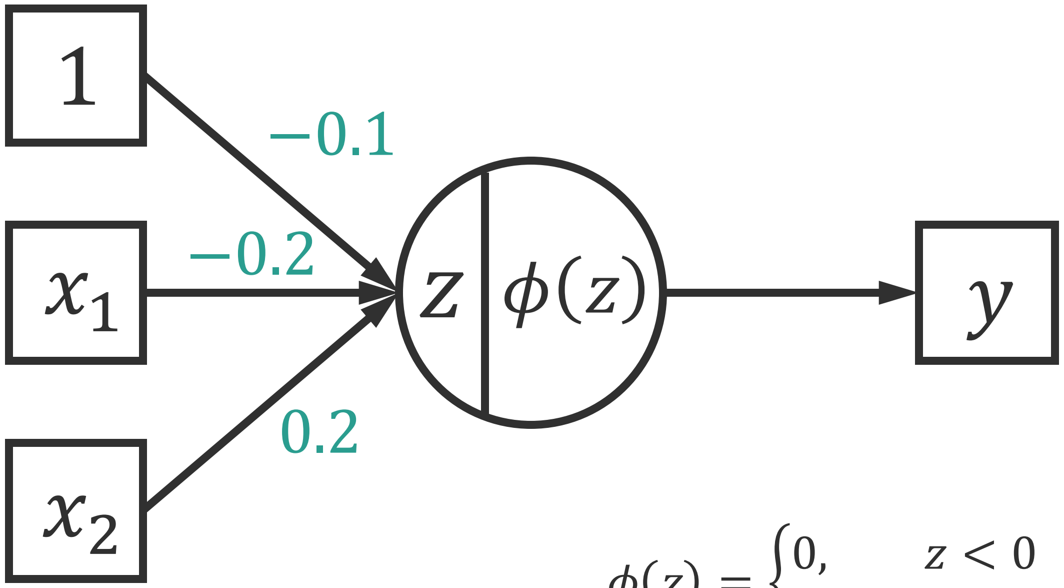
$$w_2^{(new)} = 0.3 + 1(\mathbf{0} - \mathbf{1})0.1 = 0.2$$

パラメータ更新

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性



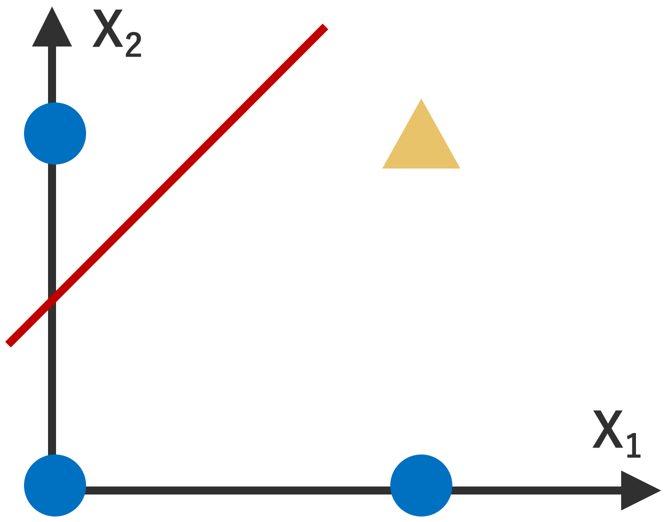


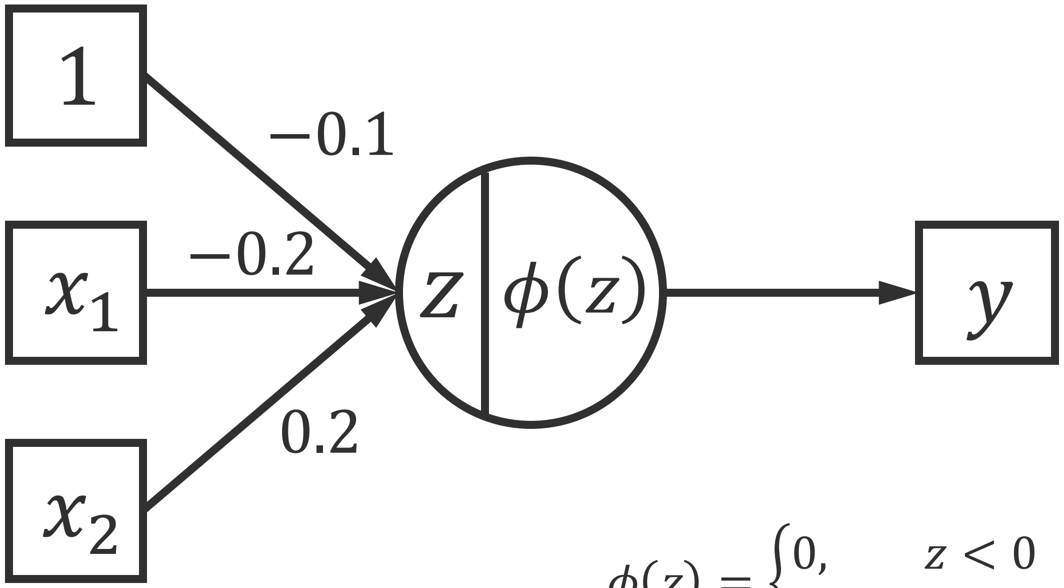
$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

パーセプトロン学習則

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性





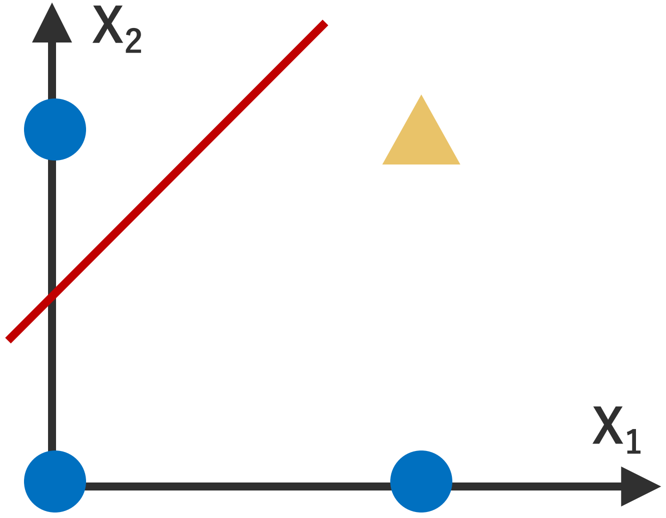
$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

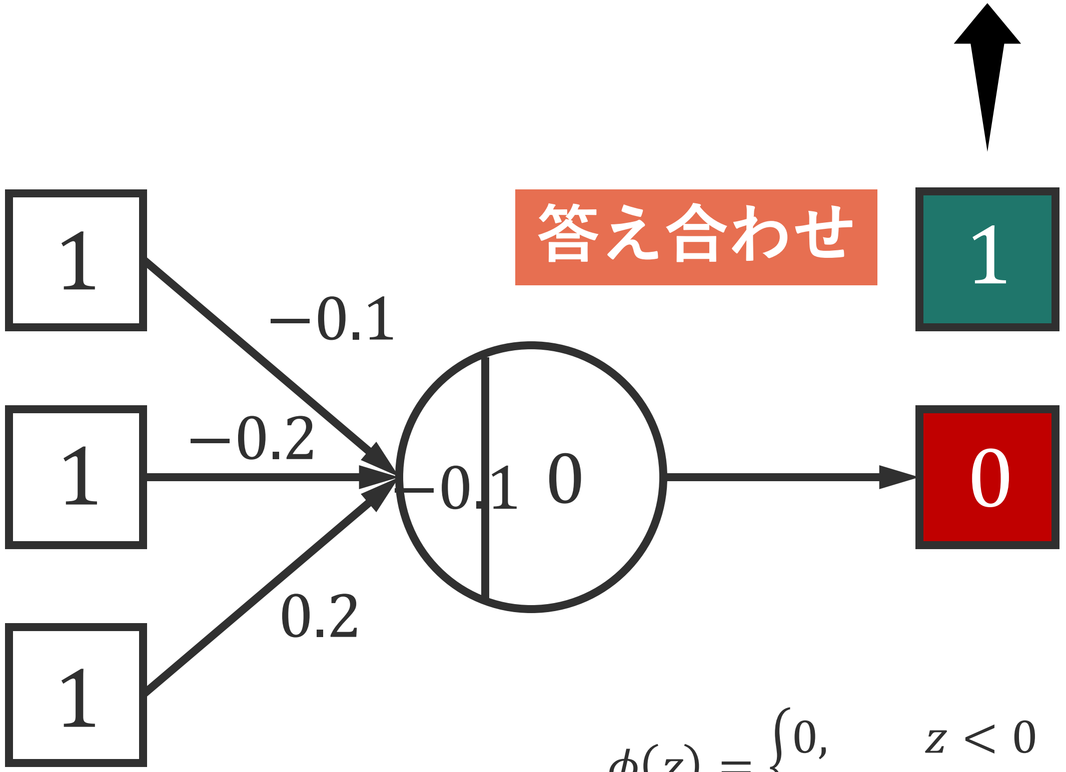
パーセプトロン学習則

$$w_0^{(new)} = -0.1 + 1(1 - \mathbf{0})0.1 = 0$$
$$w_1^{(new)} = -0.2 + 1(1 - \mathbf{0})0.1 = -0.1$$
$$w_2^{(new)} = 0.2 + 1(1 - \mathbf{0})0.1 = 0.3$$

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性





$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

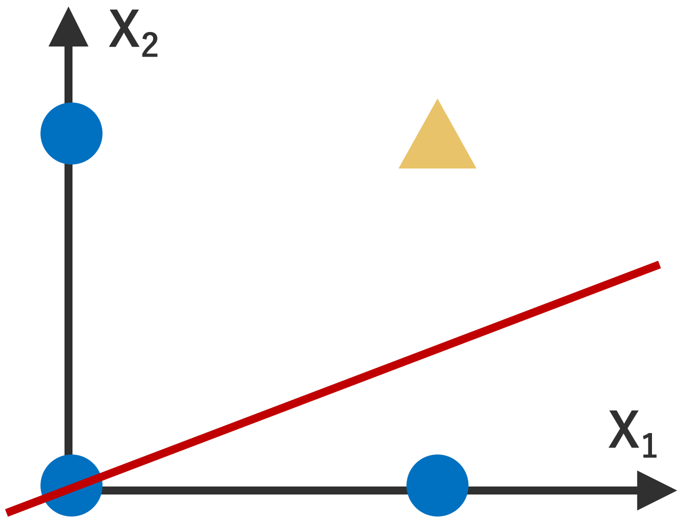
パーセプトロン学習則

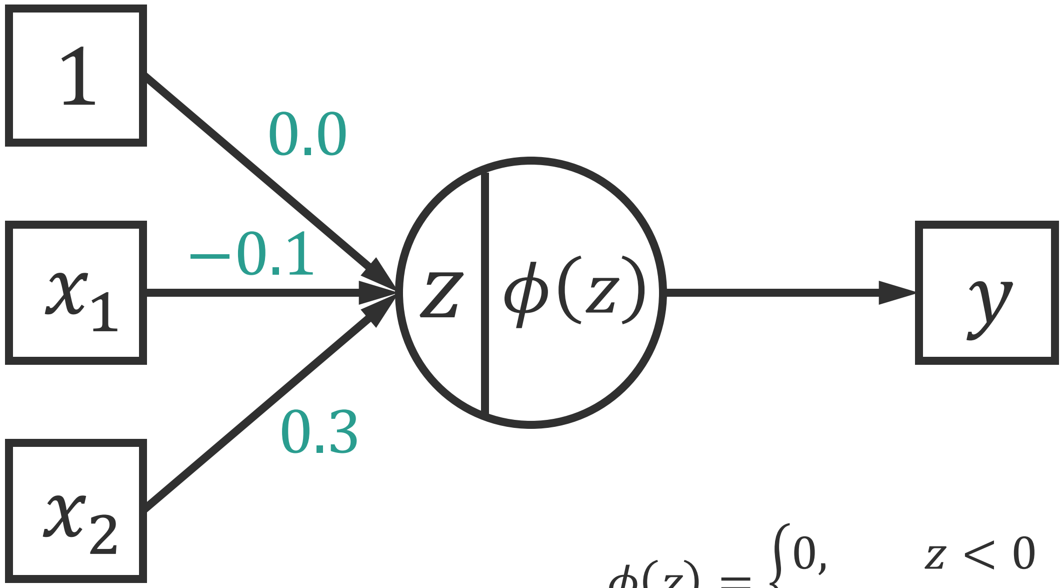
$$w_0^{(new)} = -0.1 + 1(1 - 0)0.1 = 0$$
$$w_1^{(new)} = -0.2 + 1(1 - 0)0.1 = -0.1$$
$$w_2^{(new)} = 0.2 + 1(1 - 0)0.1 = 0.3$$

パラメータ更新

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性



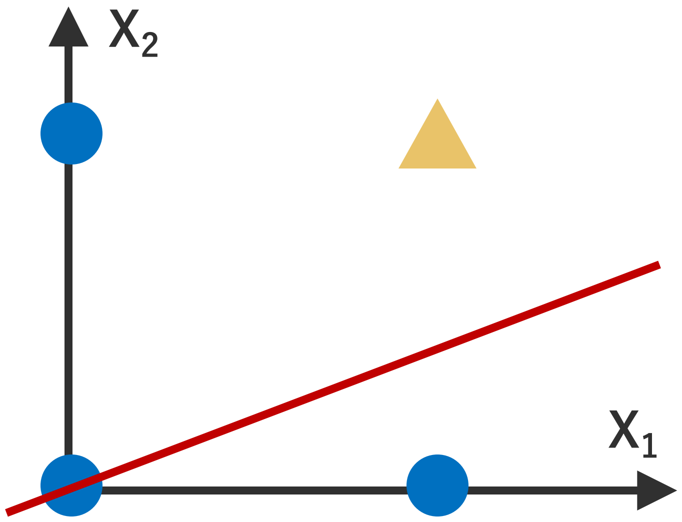


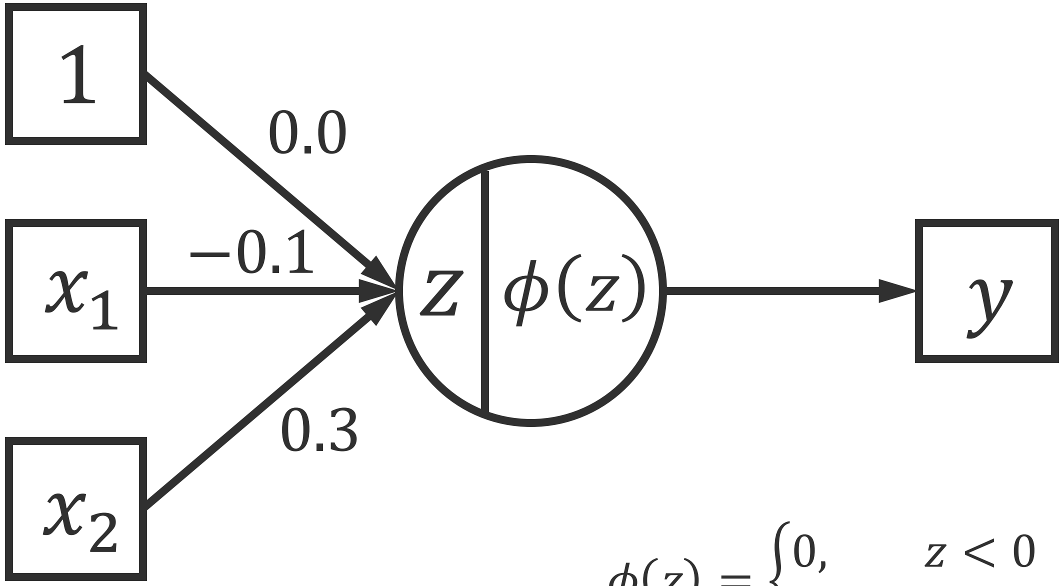
$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

パーセプトロン学習則

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性





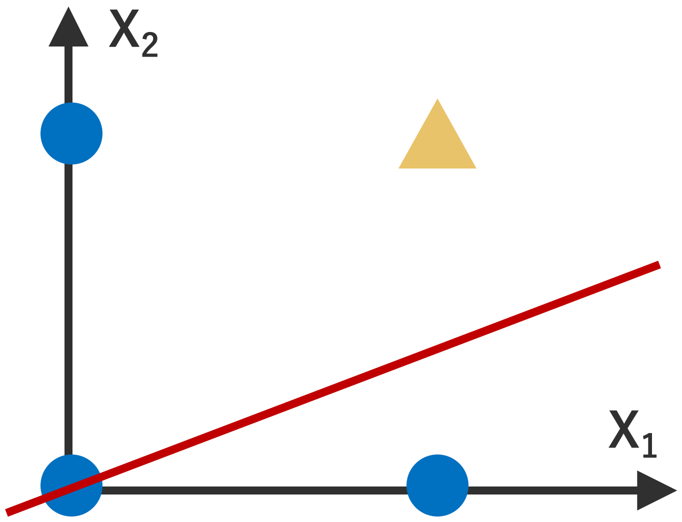
$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

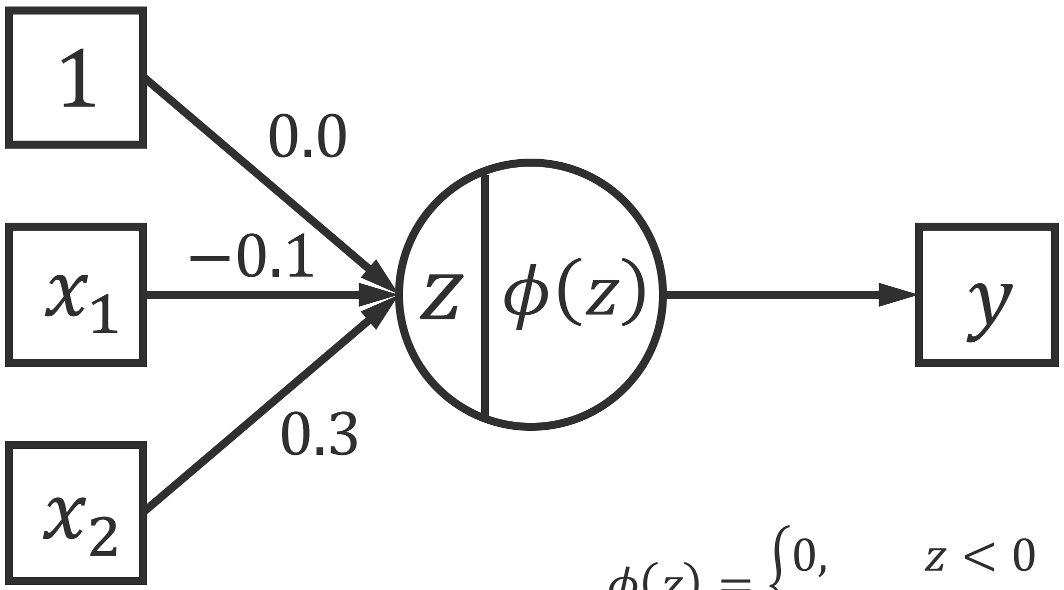
パーセプトロン学習則

シャッフル

X_1	X_2	Y
0	0	0
1	1	1
1	0	0
0	1	0

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性





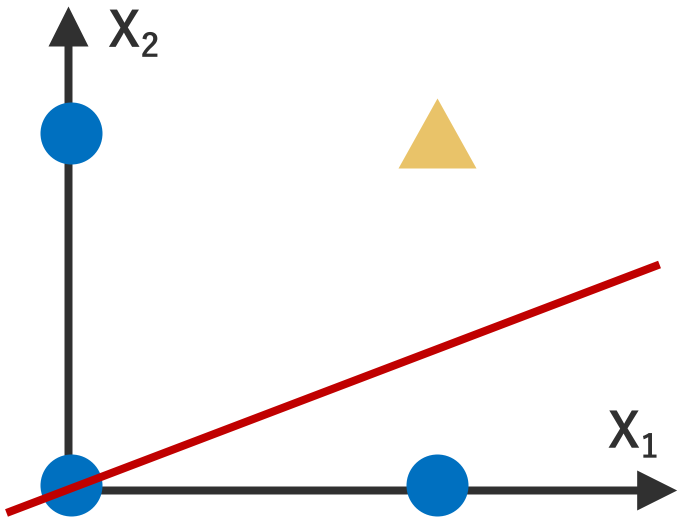
$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

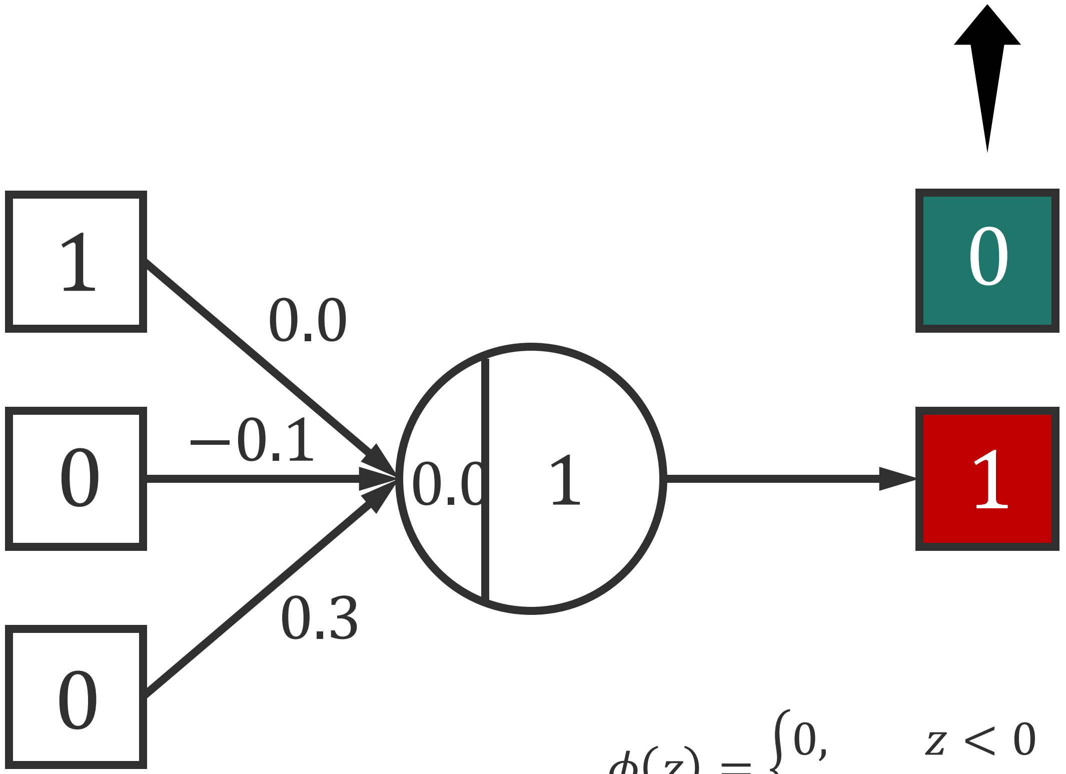
パーセプトロン学習則

$$w_0^{(new)} = 0.0 + 1(\mathbf{0} - \mathbf{1})0.1 = -0.1$$
$$w_1^{(new)} = -0.1 + 0(\mathbf{0} - \mathbf{1})0.1 = -0.1$$
$$w_2^{(new)} = 0.3 + 0(\mathbf{0} - \mathbf{1})0.1 = 0.3$$

X_1	X_2	Y
0	0	0
1	1	1
1	0	0
0	1	0

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性





$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

パーセプトロン学習則

$$w_0^{(new)} = 0.0 + 1(0 - 1)0.1 = -0.1$$

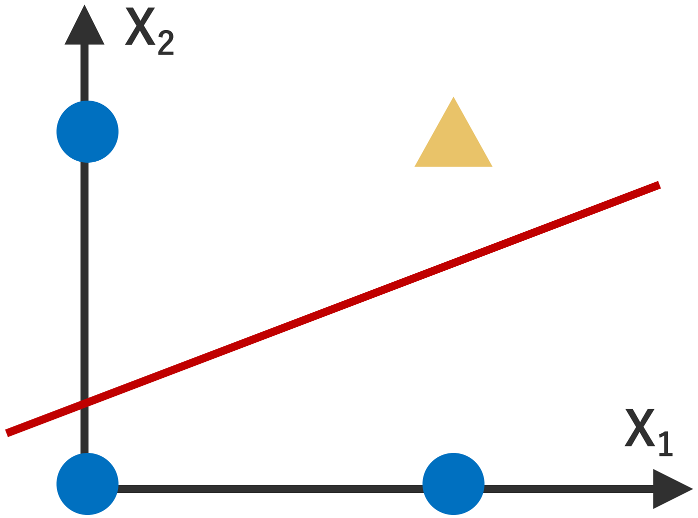
$$w_1^{(new)} = -0.1 + 0(0 - 1)0.1 = -0.1$$

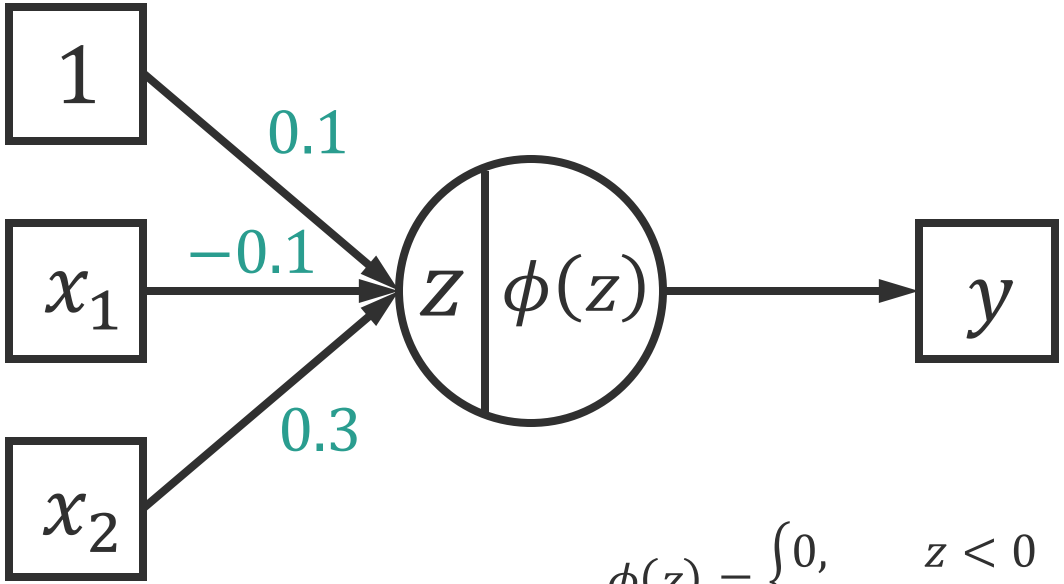
$$w_2^{(new)} = 0.3 + 0(0 - 1)0.1 = 0.3$$

パラメータ更新

X_1	X_2	Y
0	0	0
1	1	1
1	0	0
0	1	0

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性





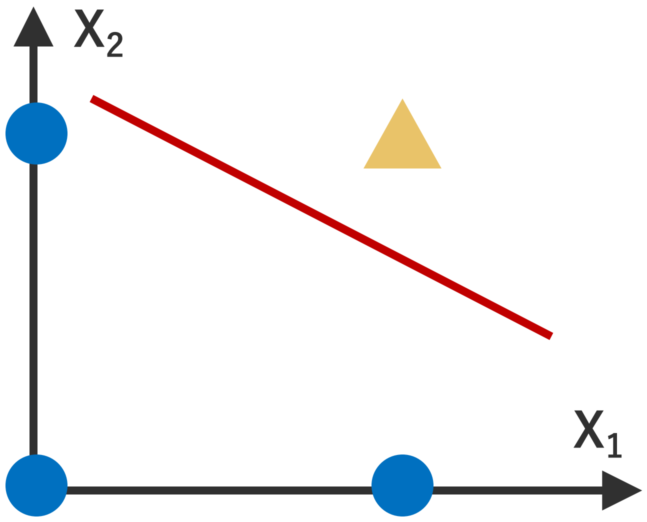
$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

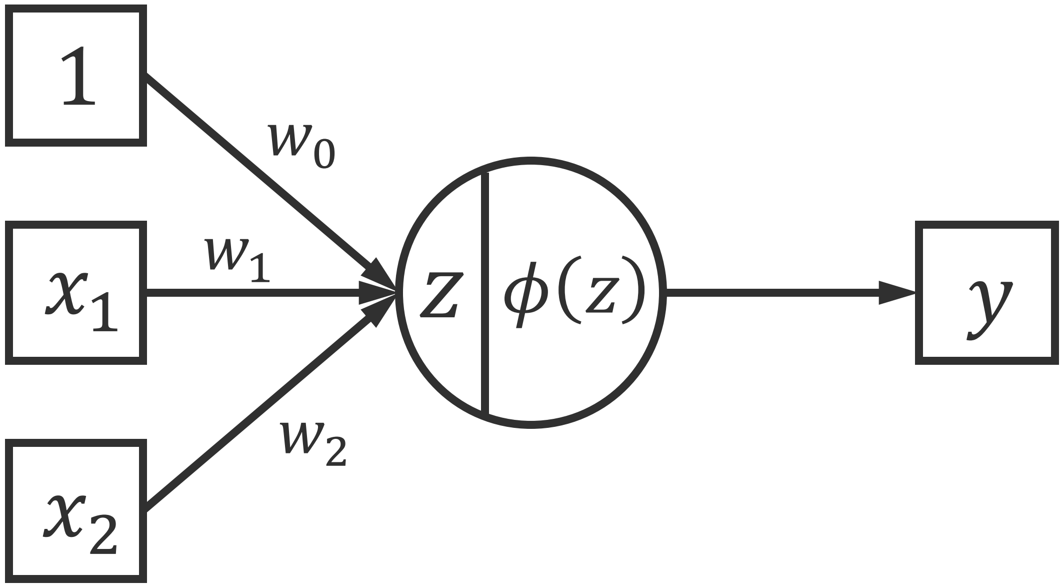
パーセプトロン学習則

すべてのサンプルを正しく分類できるまで学習を繰り返す

X_1	X_2	Y
0	0	0
1	1	1
1	0	0
0	1	0

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性

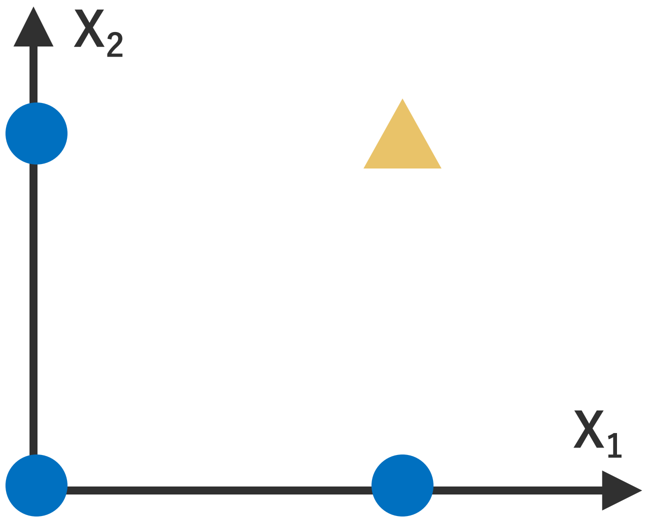


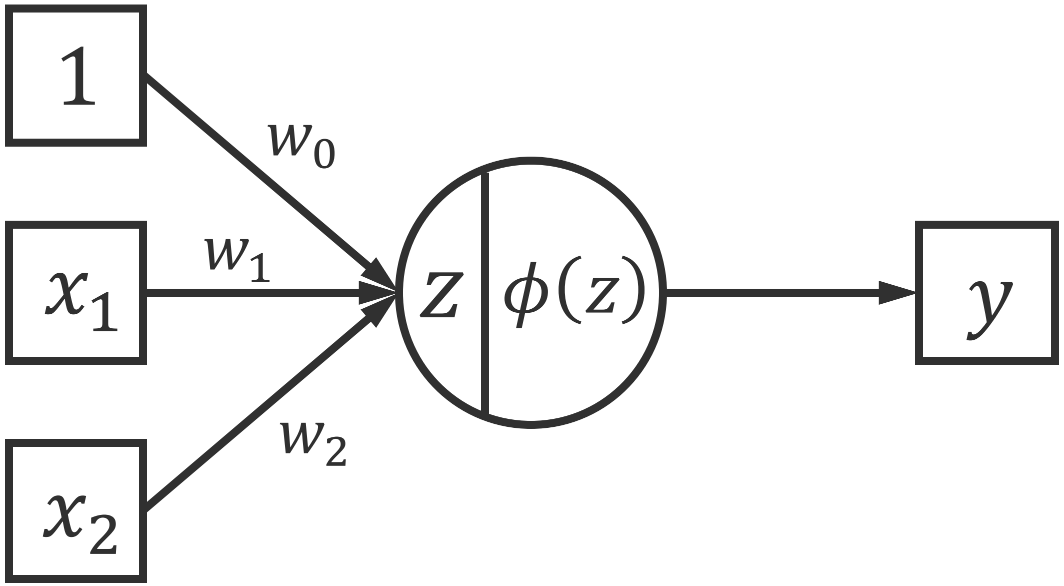


ADALINE

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性

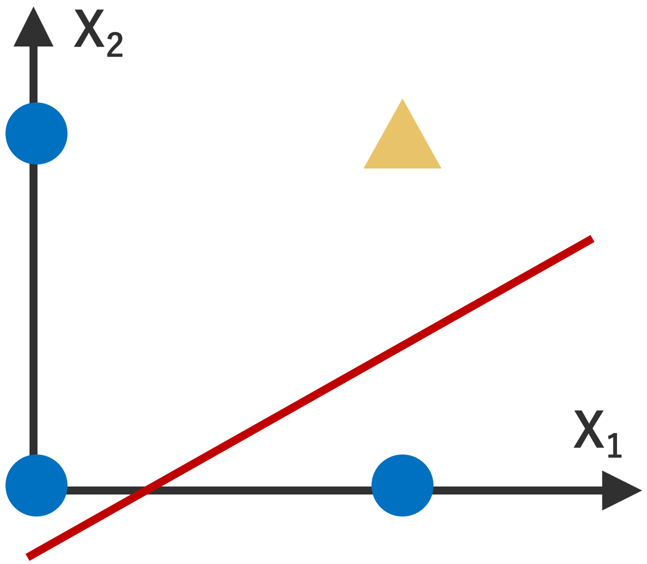




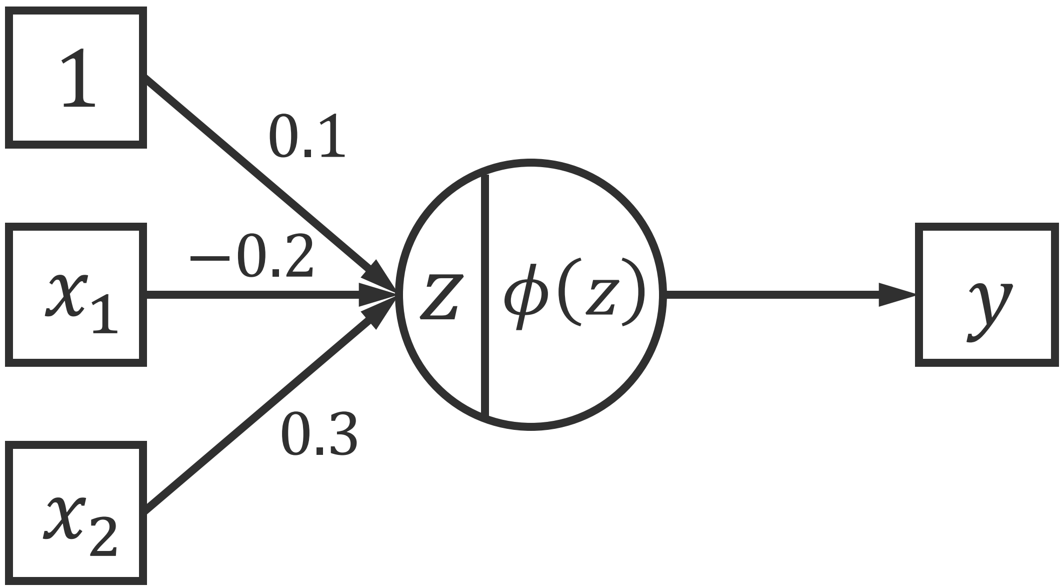
ADALINE

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性



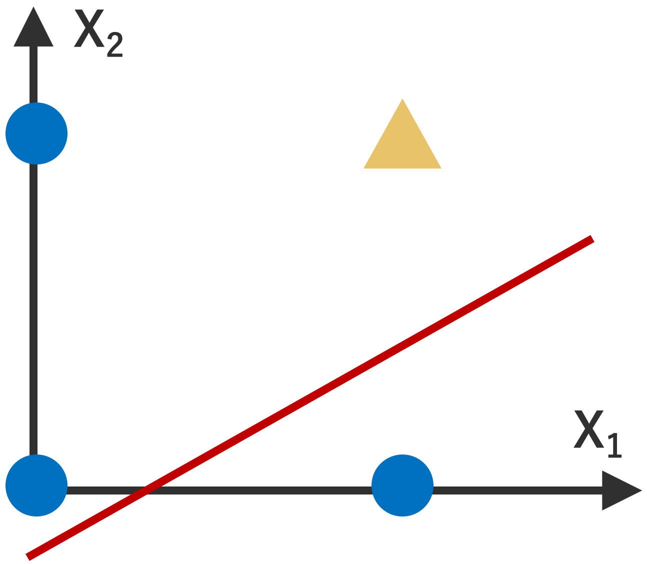
初期化

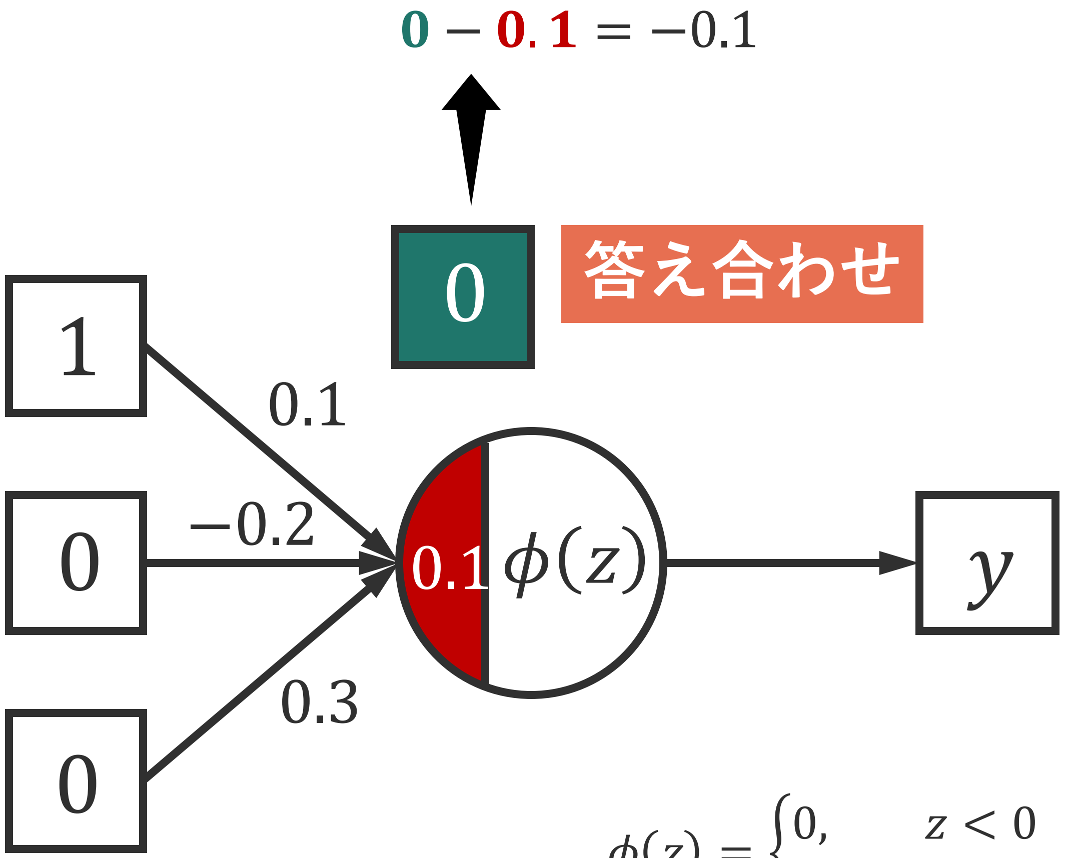


ADALINE

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性



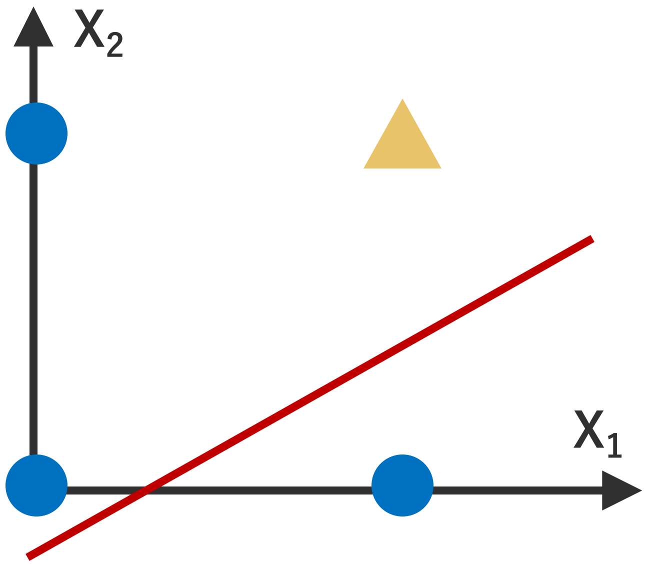


$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

ADALINE

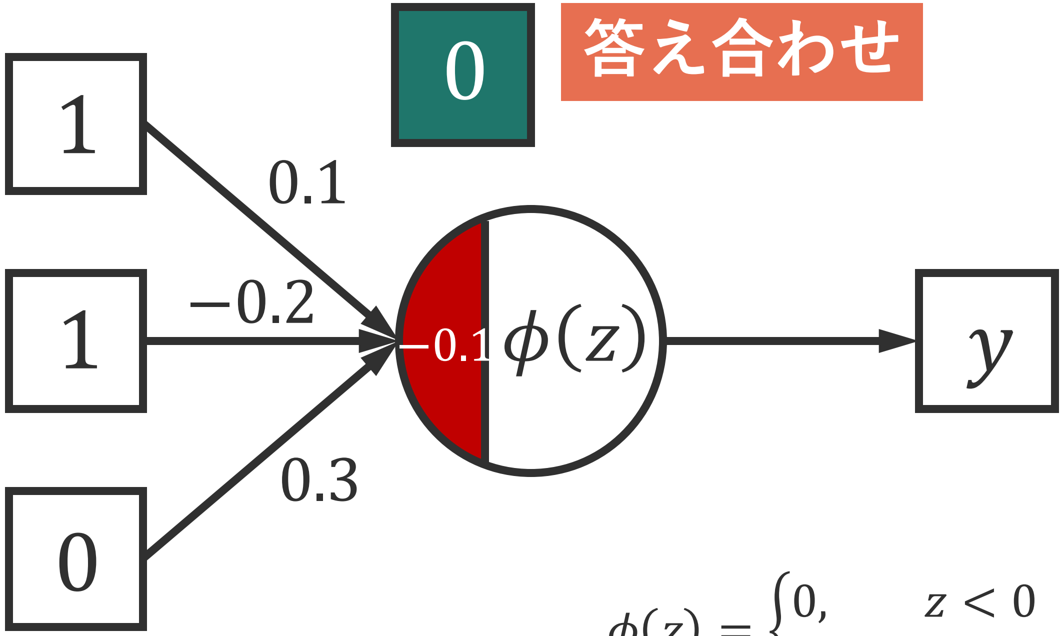
X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性



$\mathbf{0} - (\mathbf{-0.1}) = 0.1$

$\mathbf{0} - \mathbf{0.1} = -0.1$



$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

ADALINE

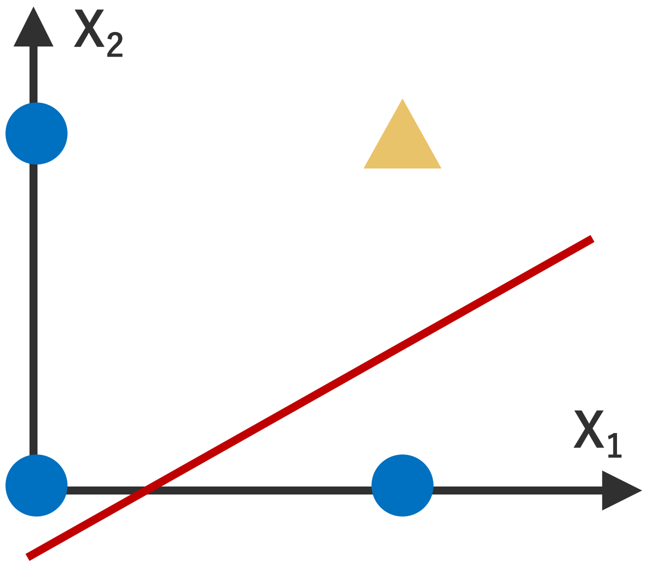
$\mathbf{0} - \mathbf{0.4} = -0.4$

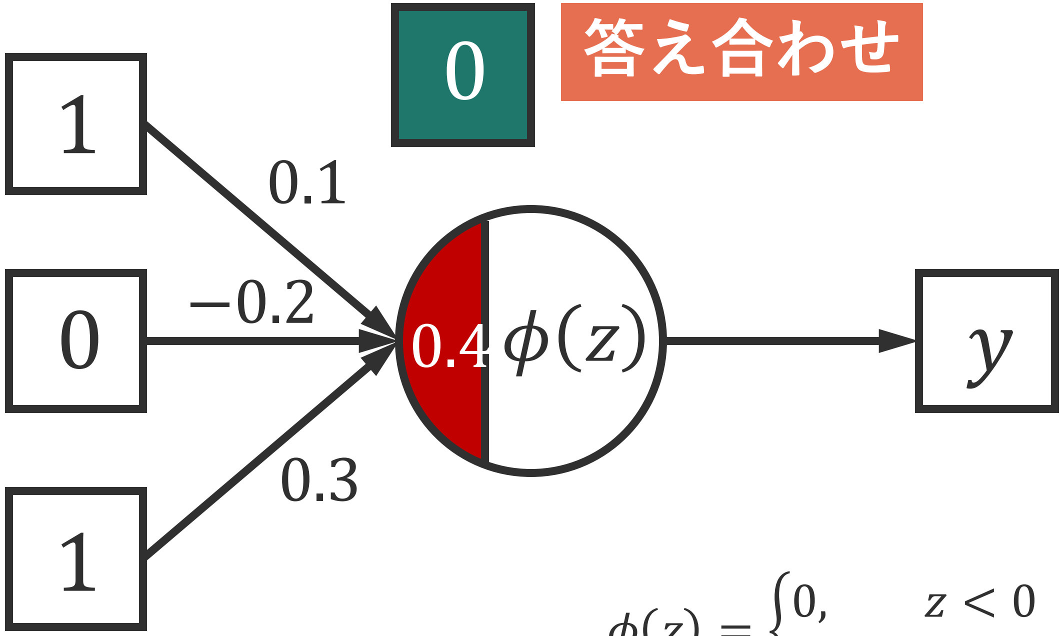
$\mathbf{0} - (\mathbf{-0.1}) = 0.1$

$\mathbf{0} - \mathbf{0.1} = -0.1$

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性





$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

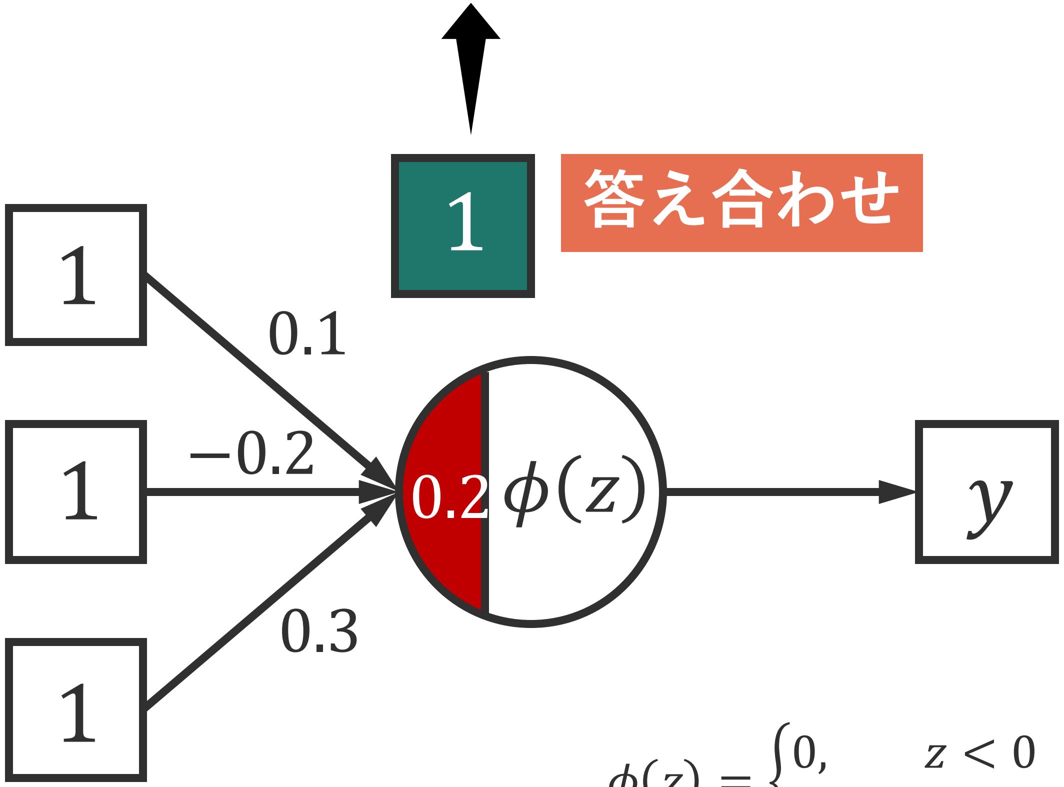
ADALINE

$\mathbf{1} - \mathbf{0.2} = 0.8$

$\mathbf{0} - \mathbf{0.4} = -0.4$

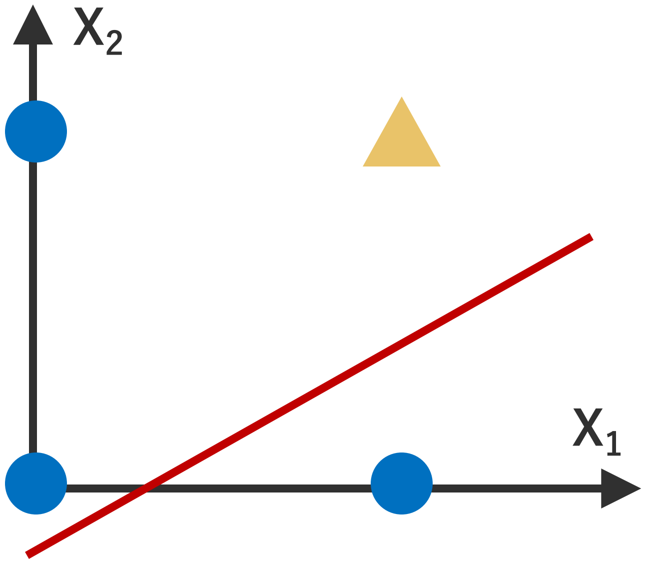
$\mathbf{0} - (\mathbf{-0.1}) = 0.1$

$\mathbf{0} - \mathbf{0.1} = -0.1$



X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性



$$\phi(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

ADALINE

$\mathbf{1} - \mathbf{0.2} = 0.8$

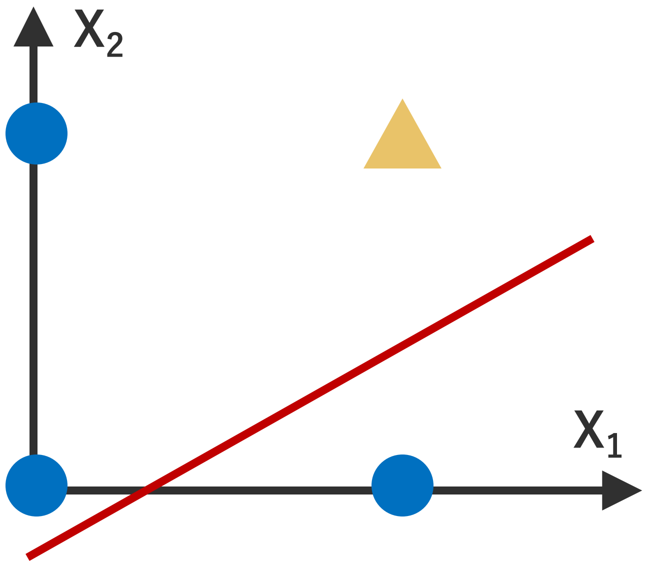
$\mathbf{0} - \mathbf{0.4} = -0.4$

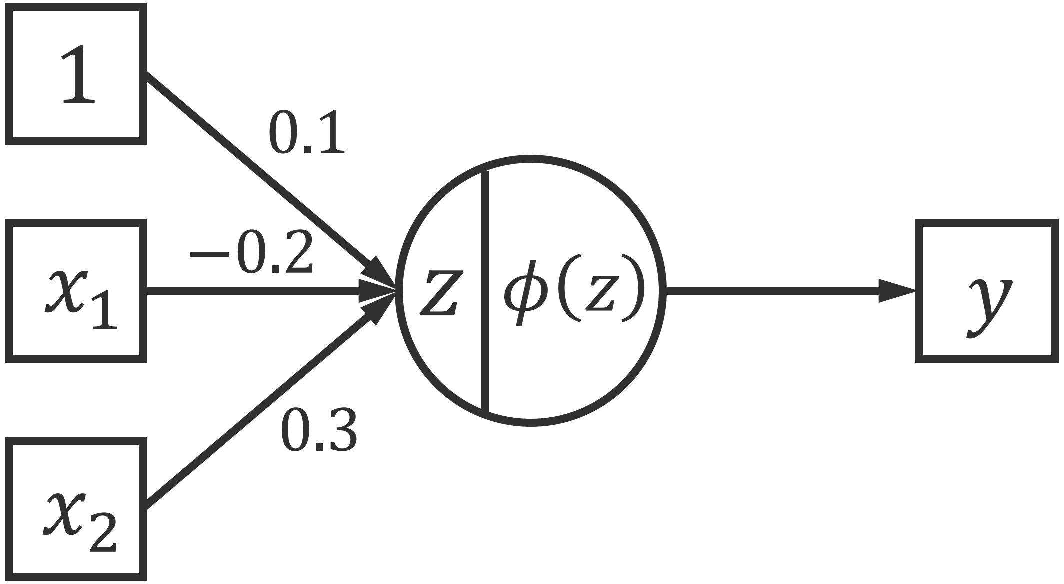
$\mathbf{0} - (\mathbf{-0.1}) = 0.1$

$\mathbf{0} - \mathbf{0.1} = -0.1$

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性





ADALINE

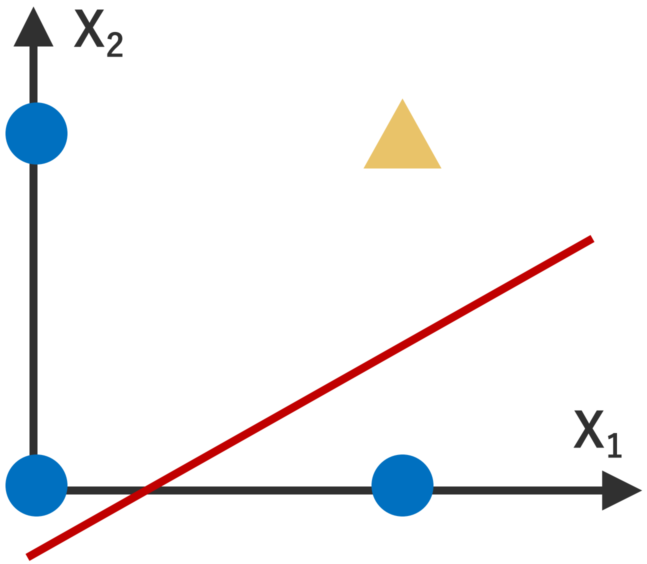
$$w_0^{(new)} = 0.14$$

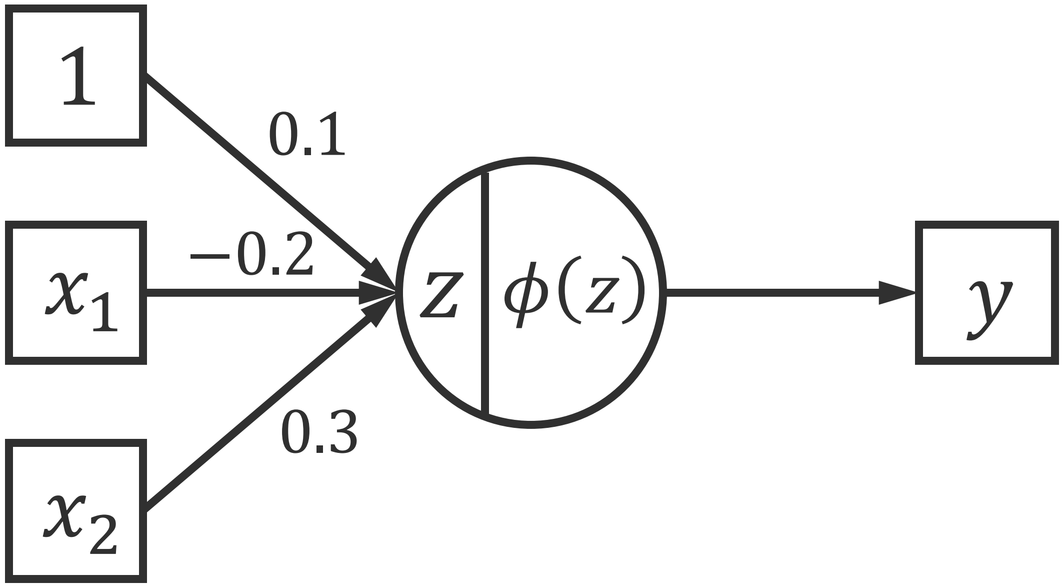
$$w_0^{(new)} = w_0^{(old)} + \Delta w_0$$
$$= w_0^{(old)} + \eta \sum (y - \boldsymbol{wx})\, x_j$$
$$= 0.1 + 0.1 \times (-0.1 \times 1 + 0.1 \times 1 - 0.4 \times 1 + 0.8 \times 1)$$
$$= 0.14$$

$$\mathbf{1} - \mathbf{0.2} = 0.8$$
$$\mathbf{0} - \mathbf{0.4} = -0.4$$
$$\mathbf{0} - (\mathbf{-0.1}) = 0.1$$
$$\mathbf{0} - \mathbf{0.1} = -0.1$$

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性





ADALINE

$$w_0^{(new)} = 0.14$$

$$w_1^{(new)} = -0.16$$

$$w_1^{(new)} = w_1^{(old)} + \Delta w_1$$

$$= w_1^{(old)} + \eta \sum (y - \boldsymbol{wx})\, x_j$$

$$= -0.2 + 0.1 \times (-0.1 \times 0 + 0.1 \times 1 - 0.4 \times 0 + 0.8 \times 1)$$

$$= -0.16$$

$$\mathbf{1} - \mathbf{0.2} = 0.8$$

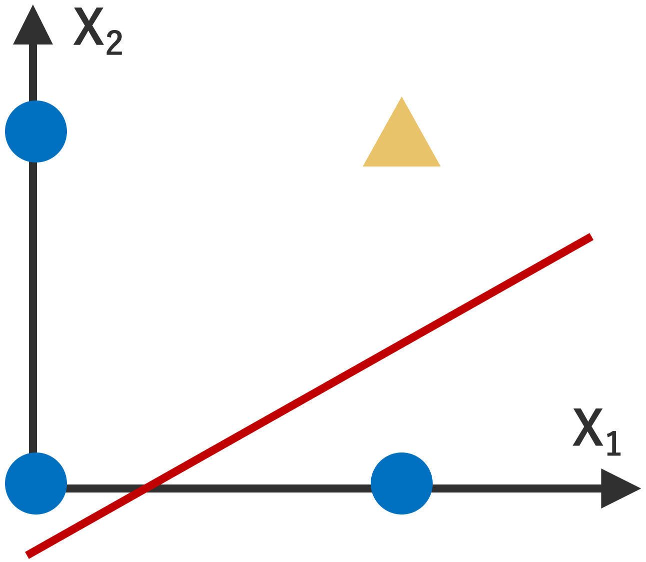
$$\mathbf{0} - \mathbf{0.4} = -0.4$$

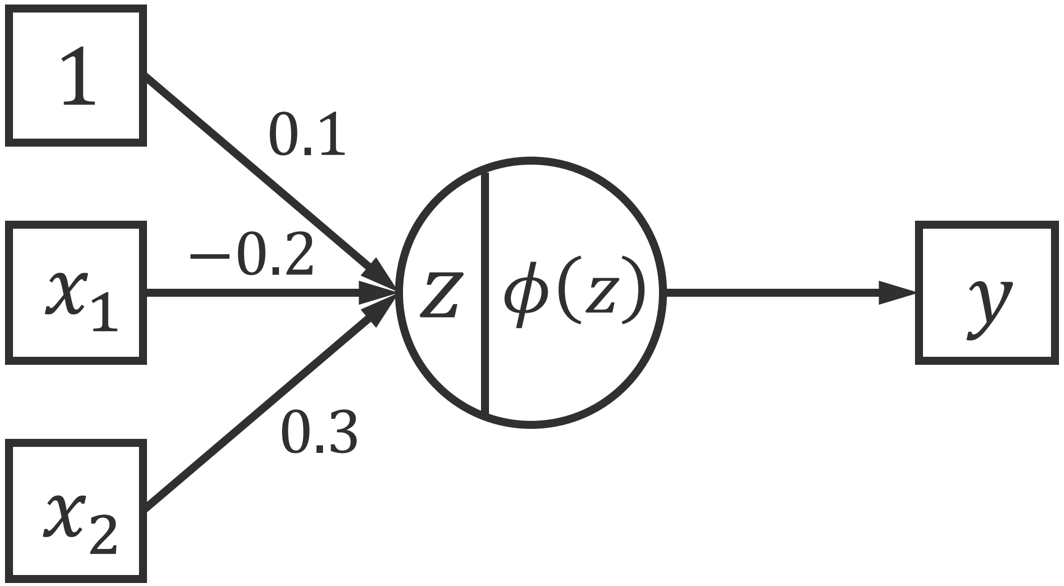
$$\mathbf{0} - (\mathbf{-0.1}) = 0.1$$

$$\mathbf{0} - \mathbf{0.1} = -0.1$$

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性





ADALINE

$$w_0^{(new)} = 0.14$$

$$w_1^{(new)} = -0.16$$

$$w_2^{(new)} = 0.34$$

$$
\begin{aligned}
w_1^{(new)} &= w_1^{(old)} + \Delta w_1 \\
&= w_1^{(old)} + \eta \sum (y - \boldsymbol{wx})\, x_j \\
&= 0.3 + 0.1 \times (-0.1 \times 0 + 0.1 \times 0 - 0.4 \times 1 + 0.8 \times 1) \\
&= 0.34
\end{aligned}
$$

$$\mathbf{1} - \mathbf{0.2} = 0.8$$

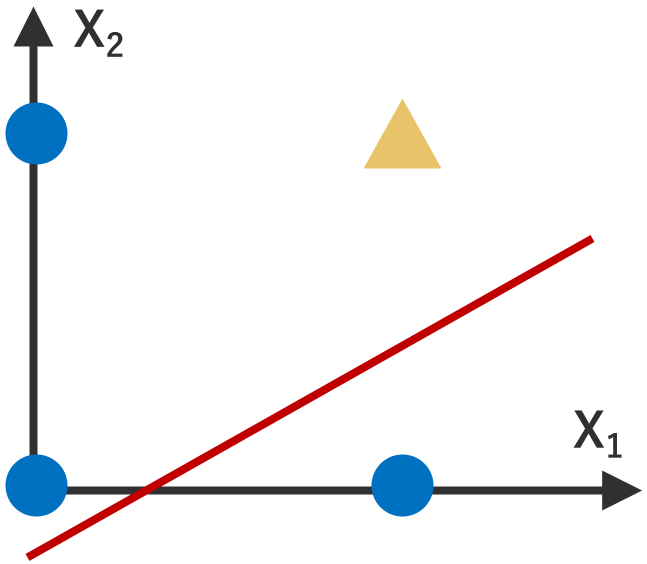
$$\mathbf{0} - \mathbf{0.4} = -0.4$$

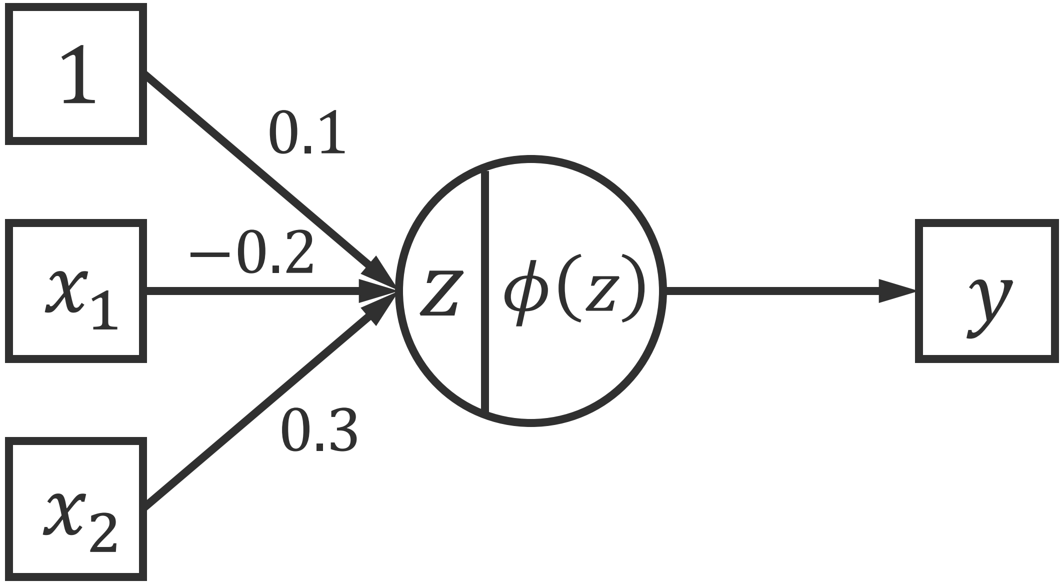
$$\mathbf{0} - (\mathbf{-0.1}) = 0.1$$

$$\mathbf{0} - \mathbf{0.1} = -0.1$$

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性





ADALINE

$$w_0^{(new)} = 0.14$$

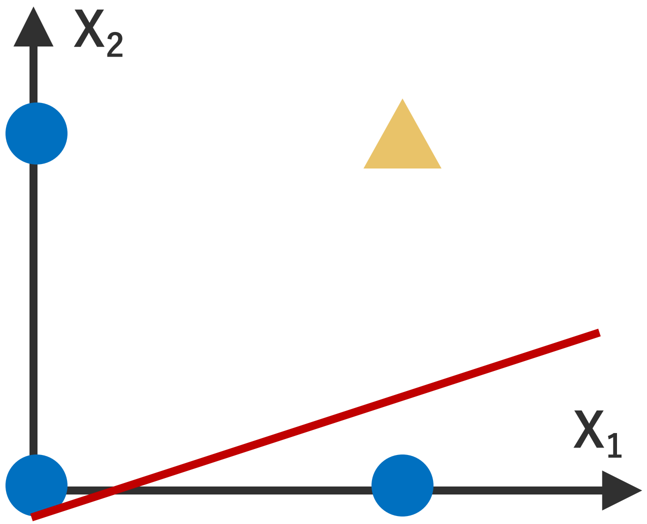
$$w_1^{(new)} = -0.16$$

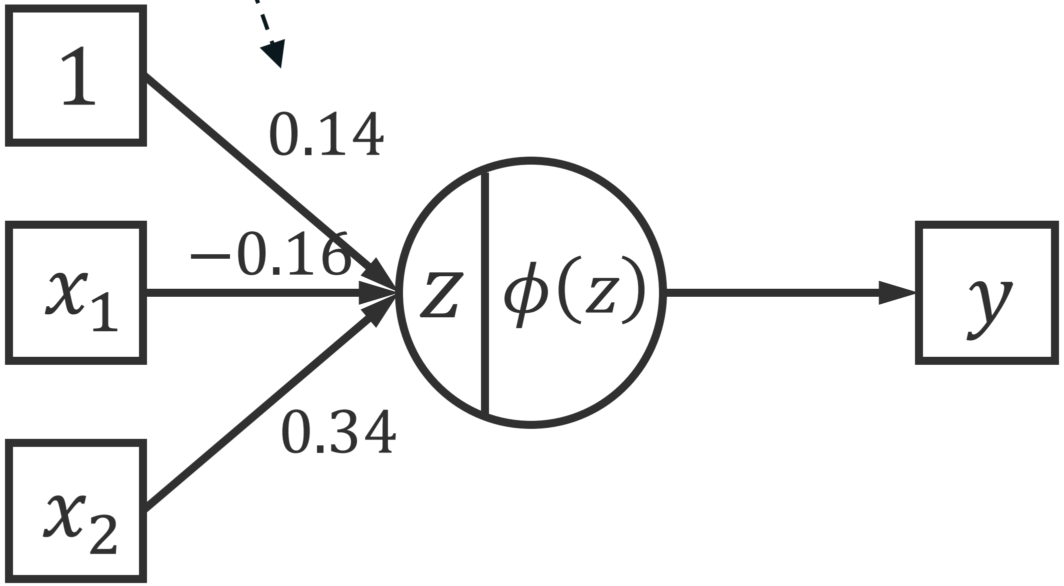
$$w_2^{(new)} = 0.34$$

パラメータ更新

X_1	X_2	Y
0	0	0
1	0	0
0	1	0
1	1	1

X1: 腫瘍面積（大小）
X2: 腫瘍形状（滑らかかどうか）
Y: 悪性・良性





ADALINE

$$\boldsymbol{w}^{(new)} = \boldsymbol{w}^{(old)} + \Delta\boldsymbol{w}$$

$$\Delta\boldsymbol{w} = -\eta\boldsymbol{\nabla} loss(\boldsymbol{w})$$

$$\boldsymbol{\nabla} loss(\boldsymbol{w}) = \frac{\partial loss(\boldsymbol{w})}{\partial\boldsymbol{w}}$$

$loss(w) = \frac{1}{2}\sum_{i}^{n}\left(y^{(i)} - w^{t}x^{(i)}\right)^2$　より、

損失関数 loss を w で偏微分すると、

$$\frac{\partial loss(w)}{\partial w} = \sum_{i}^{n}\left(y^{(i)} - w^{t}x^{(i)}\right)x^{(i)}$$

が得られる。よって w_j は次の式に基づいて更新される。

$$\begin{aligned} w_j^{(new)} &= w_j^{(old)} - \eta\frac{\partial loss(\boldsymbol{w})}{\partial w_j} \\ &= w_j^{(old)} - \eta\sum_{i}^{n}\left(y^{(i)} - w^{t}x^{(i)}\right)x_j^{(i)} \end{aligned}$$

機械学習

機械学習

回帰分析

ロジスティック回帰

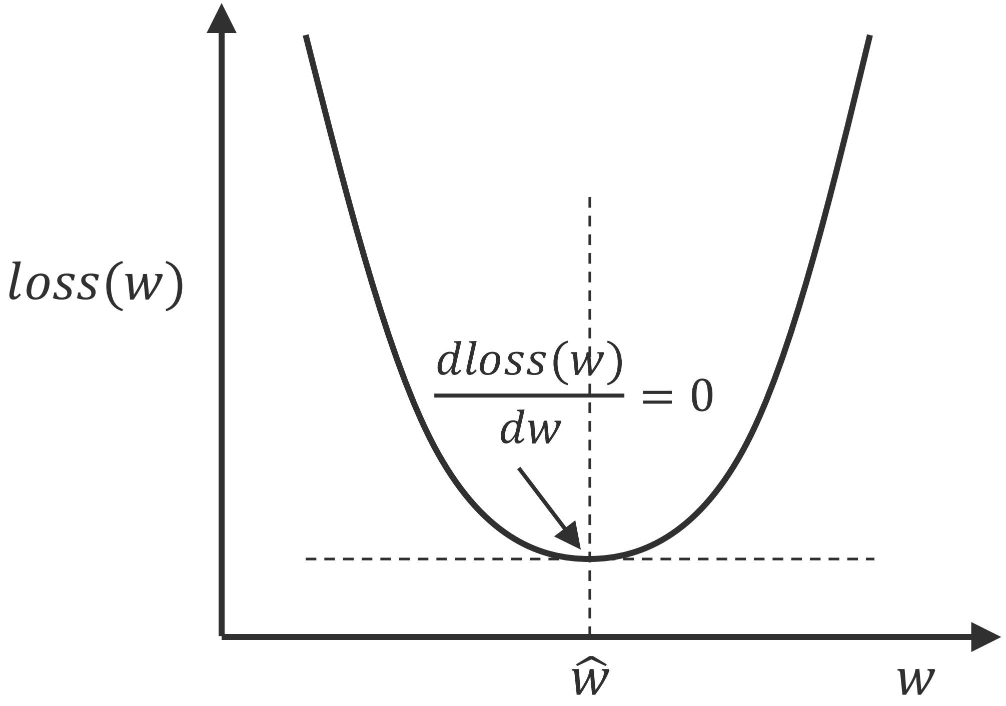
パーセプトロン・ADALINE

勾配降下法

ニューラルネットワーク

損失関数

$$loss(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i\boldsymbol{w})^2$$



最小二乗法

$$\frac{\partial loss(\boldsymbol{w})}{\partial \boldsymbol{w}} = -\boldsymbol{X}^t\boldsymbol{y} + (\boldsymbol{X}^t\boldsymbol{X})^t\boldsymbol{w} = 0$$

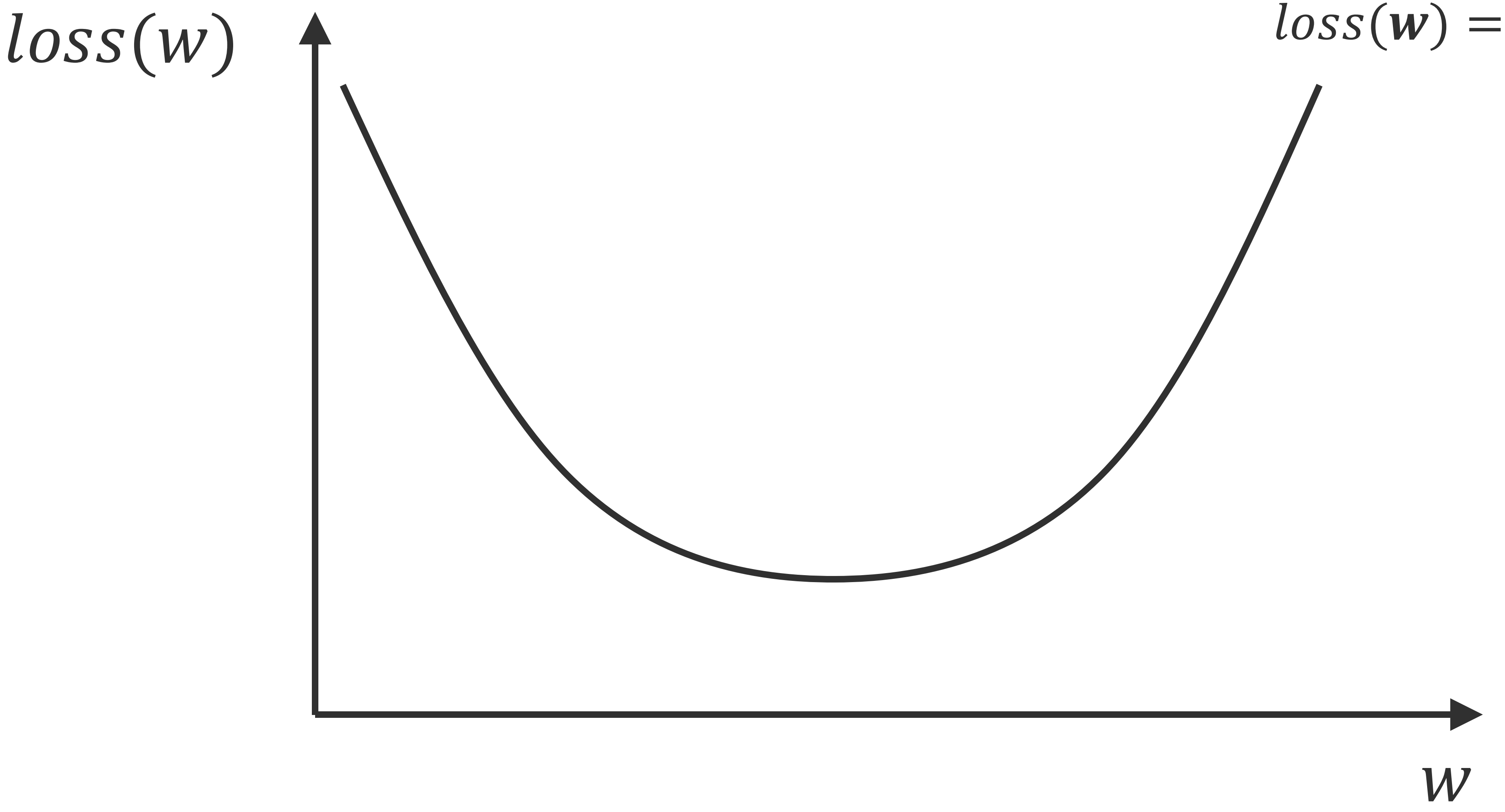
$$\boldsymbol{w} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y}$$

- ある説明変数の傾向が他の説明変数の傾向と似ている場合、**X** の逆行列を求めることができない。
  - Ridge 回帰、スパース回帰

$$\boldsymbol{w} = (\boldsymbol{X}^t\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^t\boldsymbol{y}$$

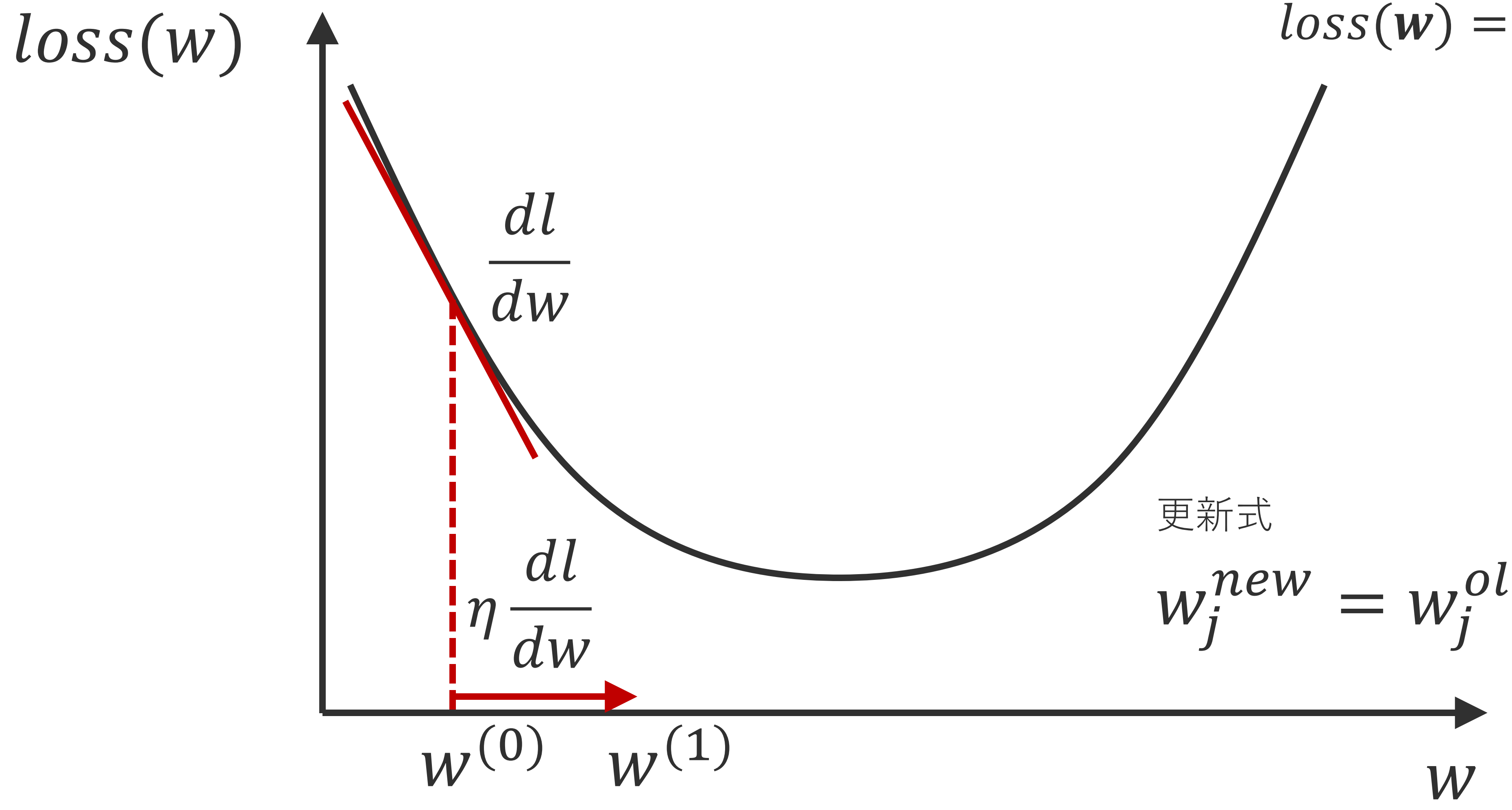
- 標本数（**y**, **x**）が大量にあるとき、ハードウェアの制限で逆行列を求めることができない場合がある。
  - 勾配降下法

勾配降下法



$$loss(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i\boldsymbol{w})^2$$

勾配降下法

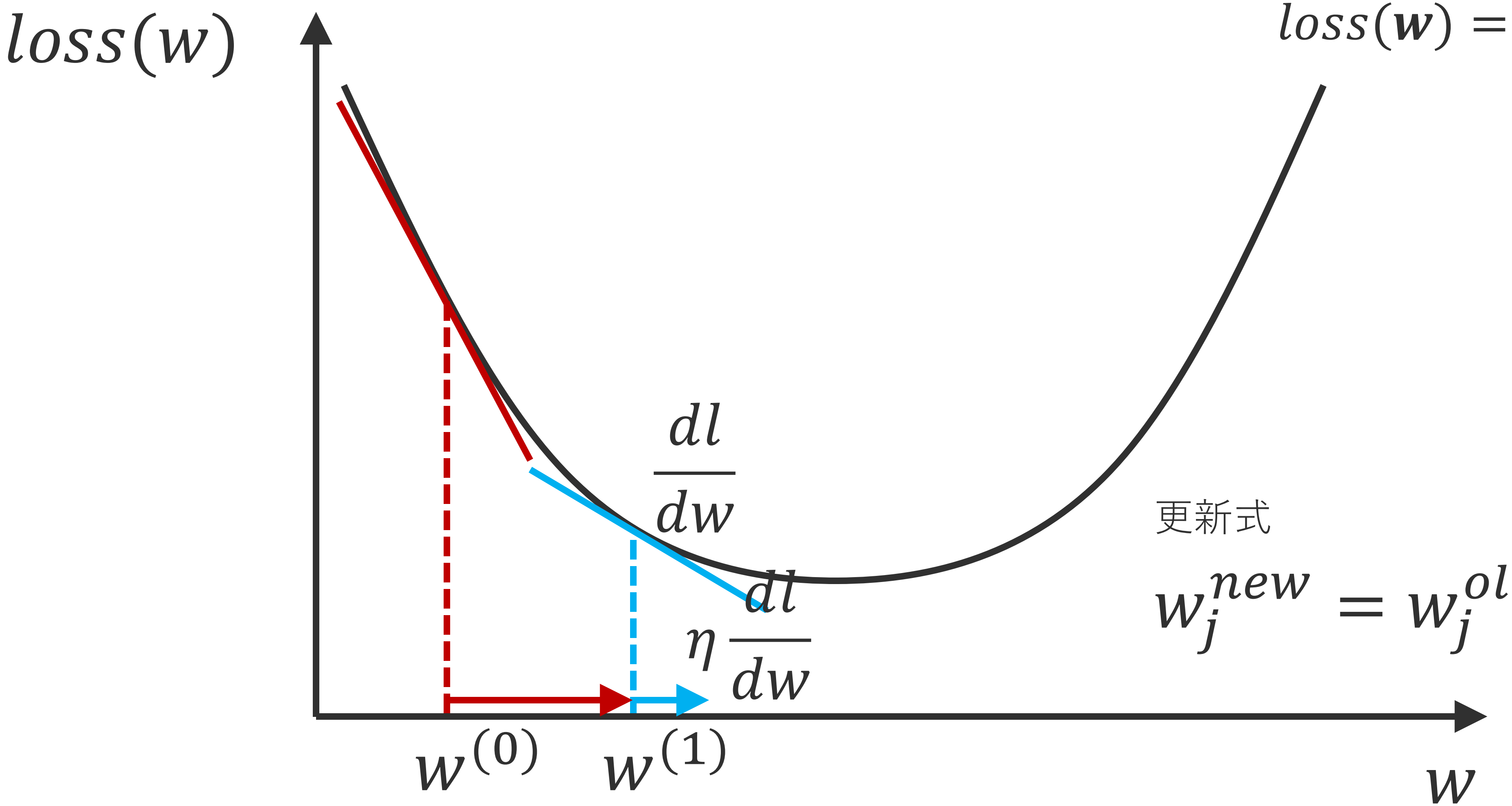


$$loss(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i\boldsymbol{w})^2$$

更新式

$$w_j^{new} = w_j^{old} - \eta\frac{\partial l}{\partial w}\bigg|_{w=w_j^{old}}$$

勾配降下法



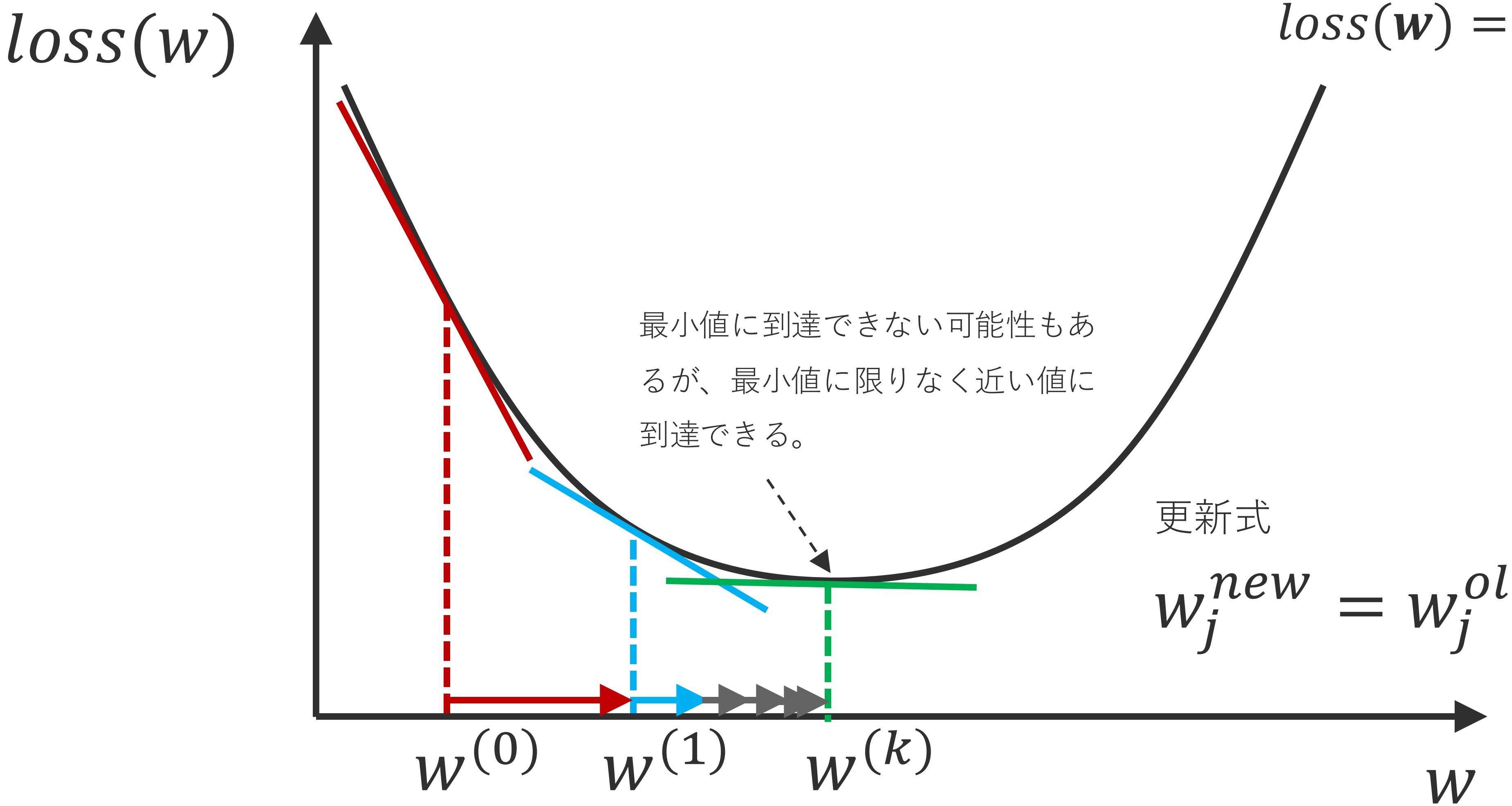
$$loss(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i\boldsymbol{w})^2$$

更新式

$$w_j^{new} = w_j^{old} - \eta \frac{\partial l}{\partial w}\bigg|_{w=w_j^{old}}$$

勾配降下法

$$loss(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i\boldsymbol{w})^2$$



更新式

$$w_j^{new} = w_j^{old} - \eta \left.\frac{\partial l}{\partial w}\right|_{w=w_j^{old}}$$

勾配降下法

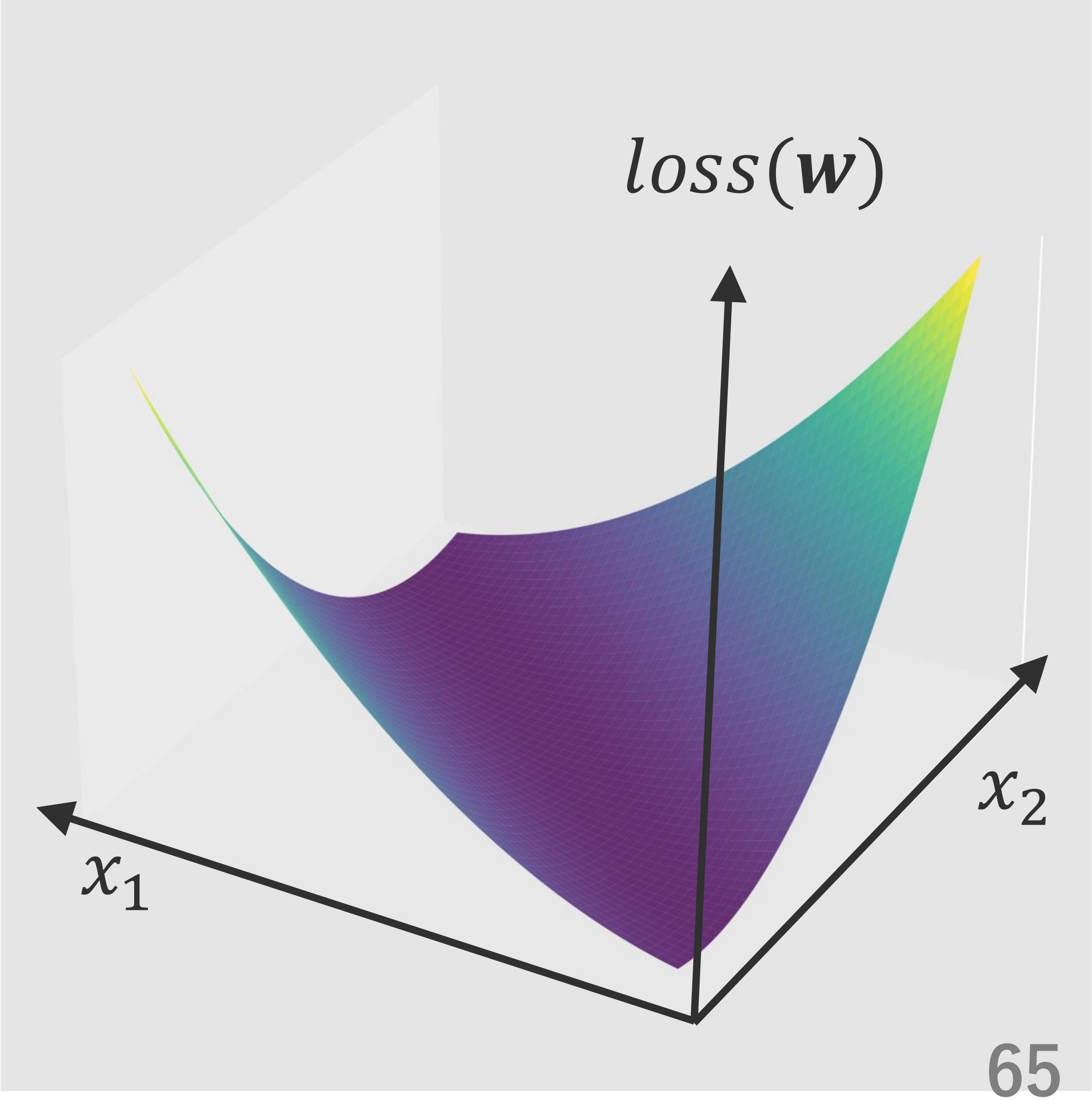
モデル

$$y = w_0 + w_1 x_1 + w_2 x_2$$

▲ 糖尿病リスク　▲ BMI　▲ 中性脂肪

損失関数

$$loss(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{2}(y_i - \boldsymbol{x}_i\boldsymbol{w})^2$$



勾配降下法

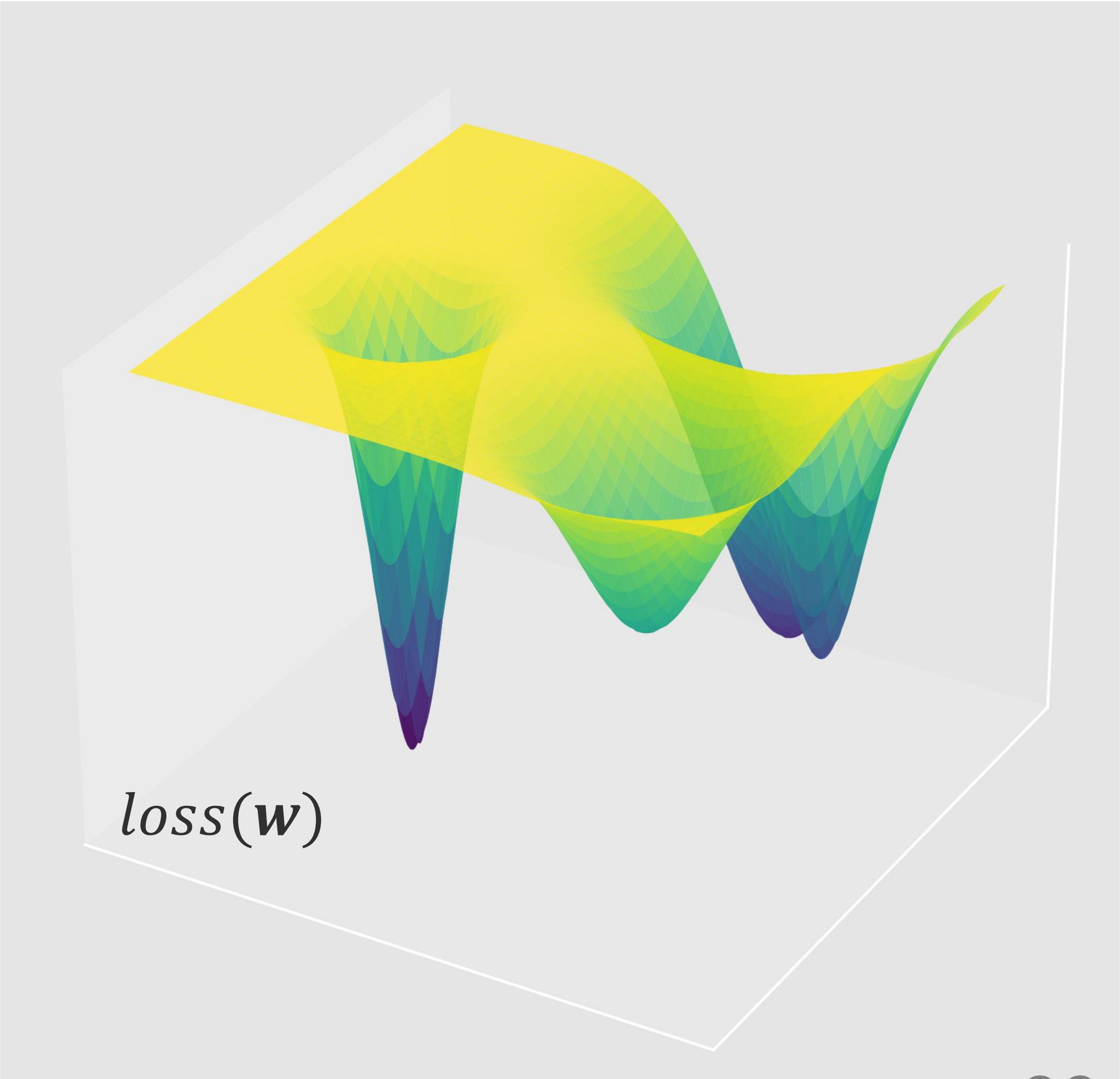
モデル

$$y = w_0 + w_1 x_1 + w_2 x_2 + \cdots$$

▲ 糖尿病リスク ▲ BMI ▲ 中性脂肪

損失関数

$$loss(\boldsymbol{w}) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i\boldsymbol{w})^2$$

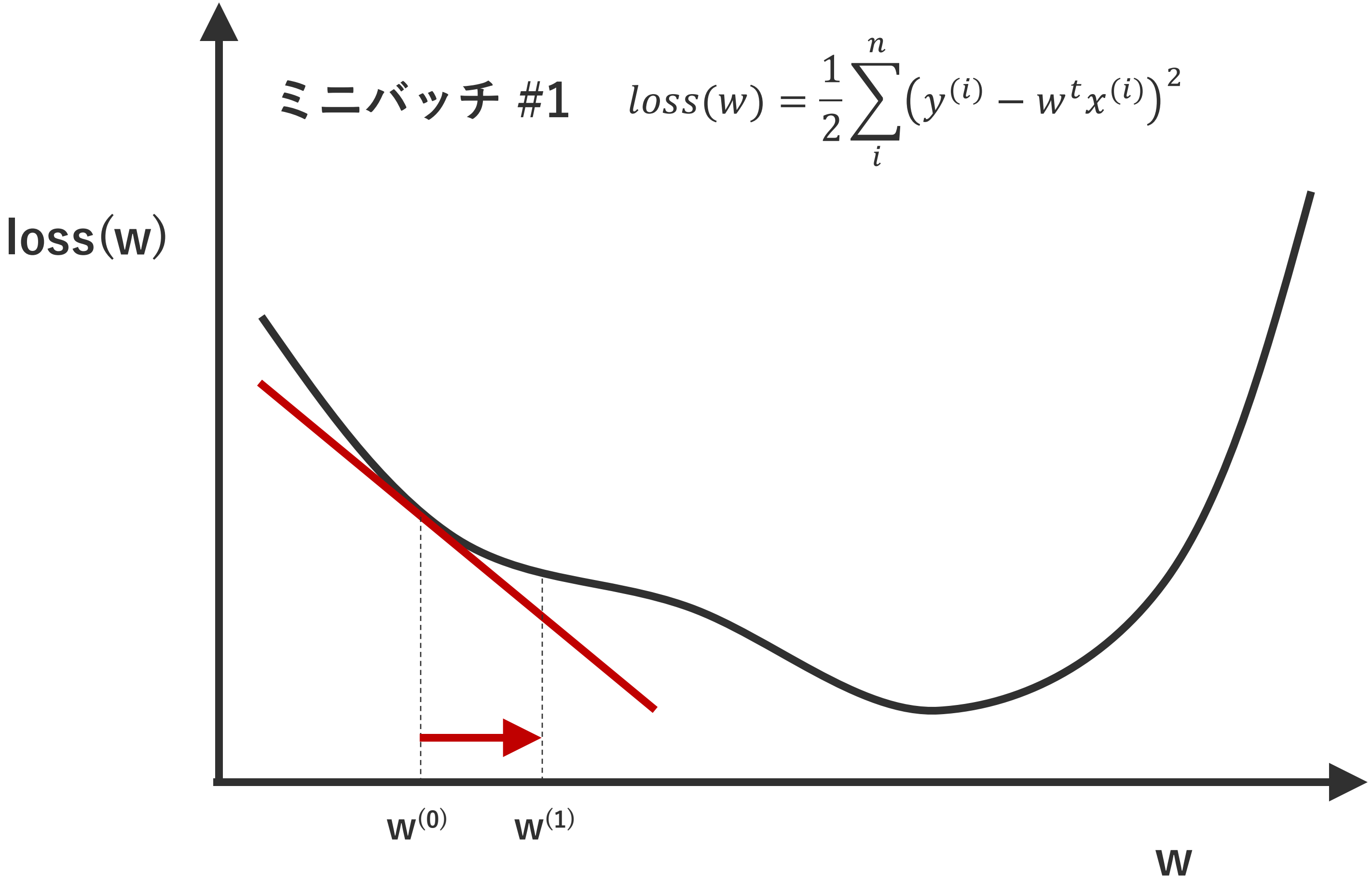


※画像はイメージです。

勾配降下法

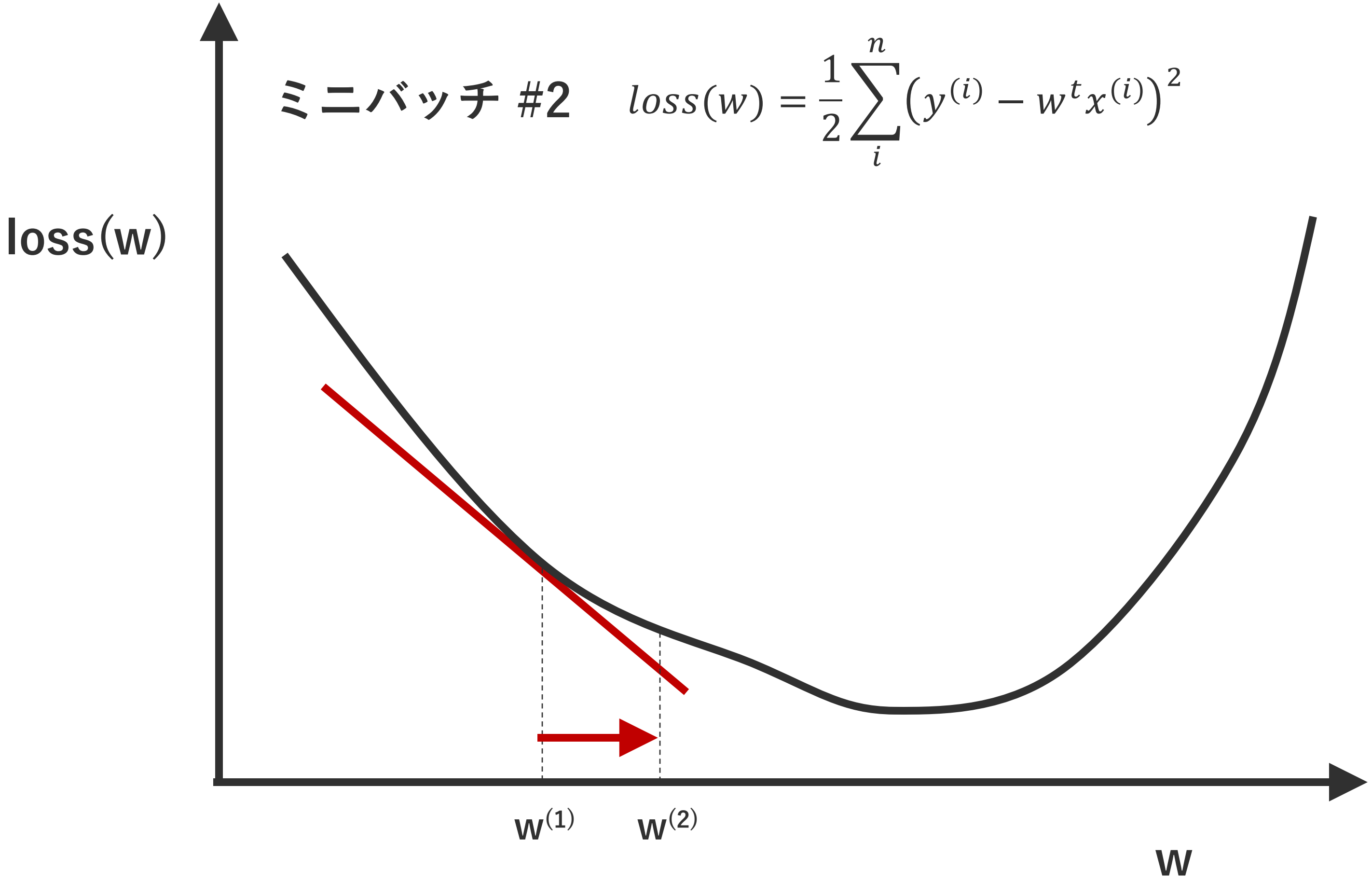
	バッチ勾配降下法	確率的勾配降下法（SGD）	ミニバッチ勾配降下法
計算対象	全データ	ランダムに抽出した1サンプル	ランダムに抽出した少数サンプル（ミニバッチ）
メリット	全データを利用するため、計算結果が比較的安定し、確実に収束する。	1 回の計算が速い。また、ノイズを含むので収束が不安定だが、最終的には良い解に収束することに繋がることが多い。	計算負担を軽減しつつ、SGD よりも安定した収束が得られる場合が多い。
デメリット	全データを利用するためメモリを要し、計算時間もかかる。	収束がノイズを含んで不安定になりやすい。	
利用場面	小規模なデータセットの場合や、精度重視で安定した最適化が求められる場面で使わる。	データセットが非常に大きい場合や、リアルタイムでモデルを更新したい場合に有効。	中規模以上のデータセットや、ニューラルネットワークのトレーニングでよく使用される。計算リソースとメモリの効率的な使用が求められる場合に有効。

ミニ勾配降下法



$$w_j^{new} = w_j^{old} - \eta \left.\frac{\partial l}{\partial w}\right|_{w=w_j^{old}}$$

ミニ勾配降下法

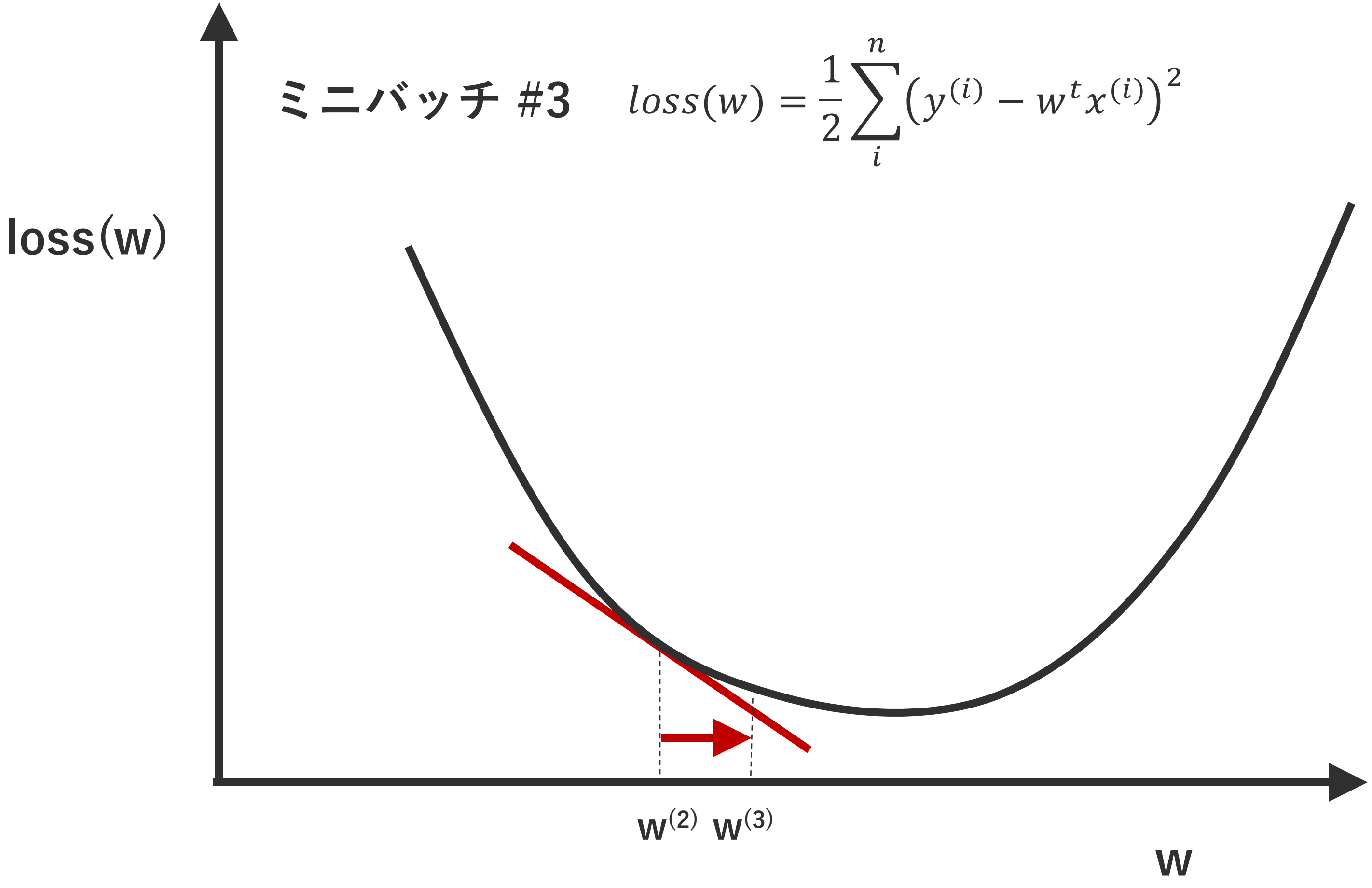


$$loss(w) = \frac{1}{2}\sum_{i}^{n}\left(y^{(i)} - w^{t}x^{(i)}\right)^{2}$$

$$w_j^{new} = w_j^{old} - \eta \left.\frac{\partial l}{\partial w}\right|_{w=w_j^{old}}$$

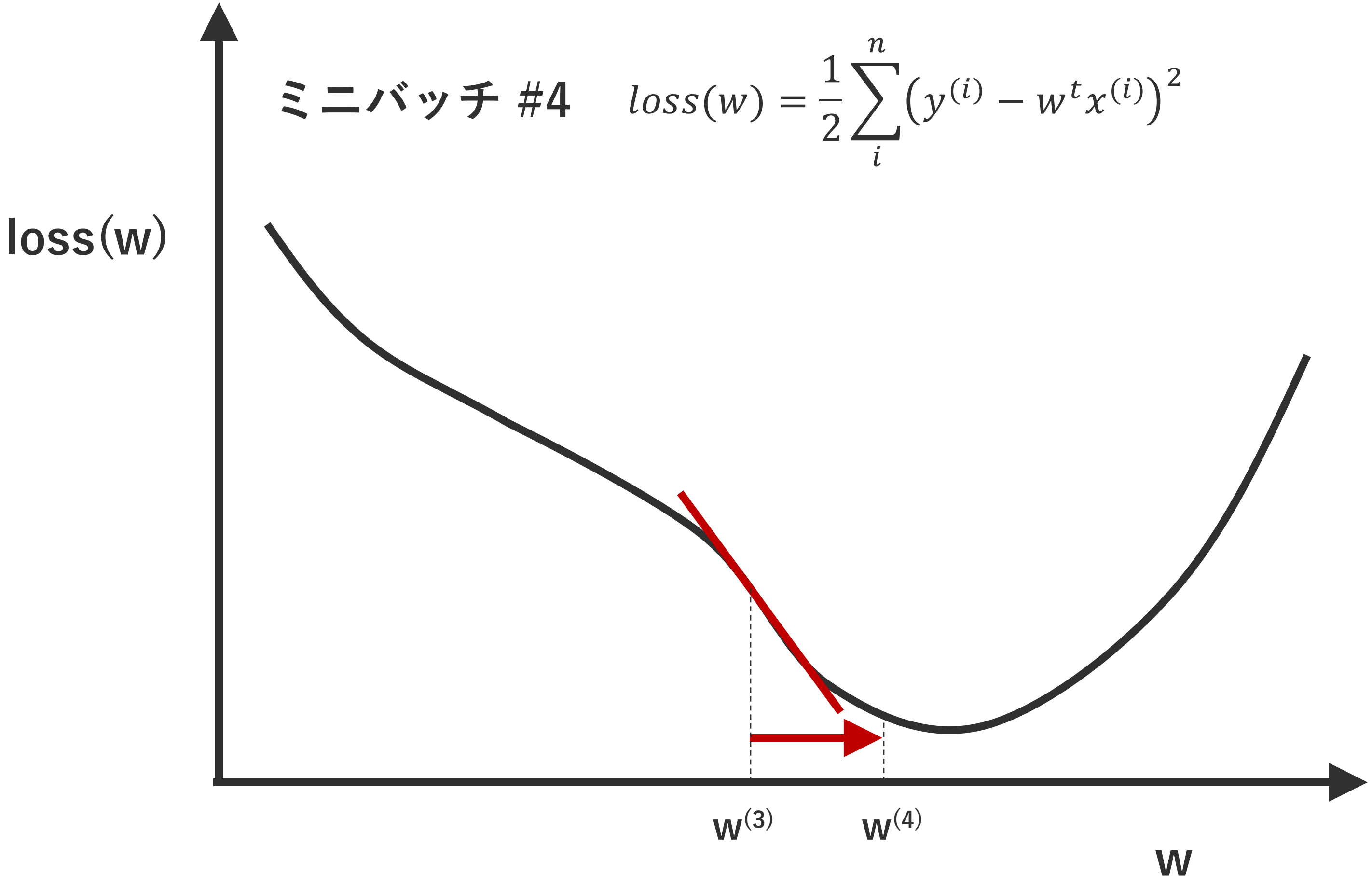
ミニ勾配降下法

ミニバッチ #3 $loss(w) = \frac{1}{2}\sum_{i}^{n}\left(y^{(i)} - w^{t}x^{(i)}\right)^{2}$



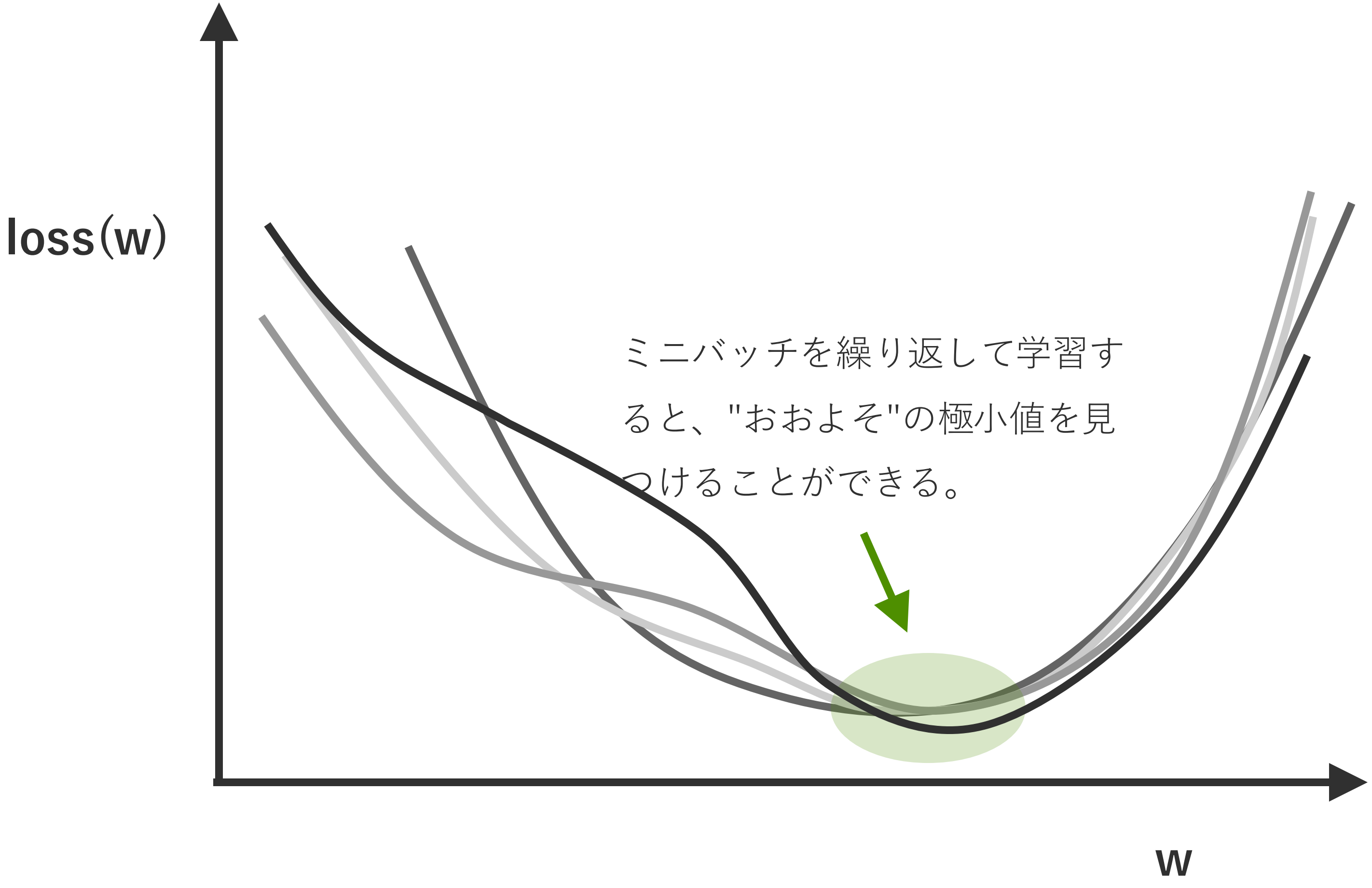
$$w_j^{new} = w_j^{old} - \eta \frac{\partial l}{\partial w}\bigg|_{w=w_j^{old}}$$

バッチ勾配降下法



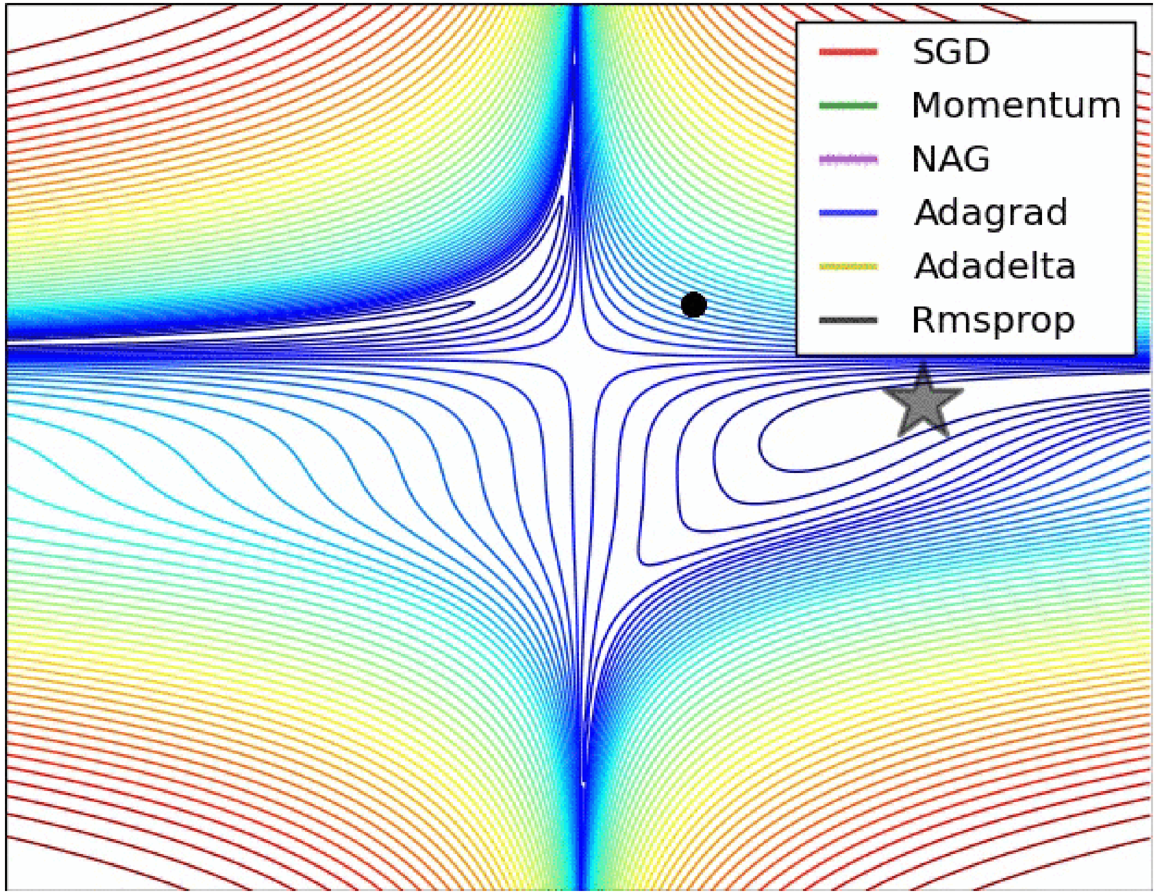
$$w_j^{new} = w_j^{old} - \eta \left.\frac{\partial l}{\partial w}\right|_{w=w_j^{old}}$$

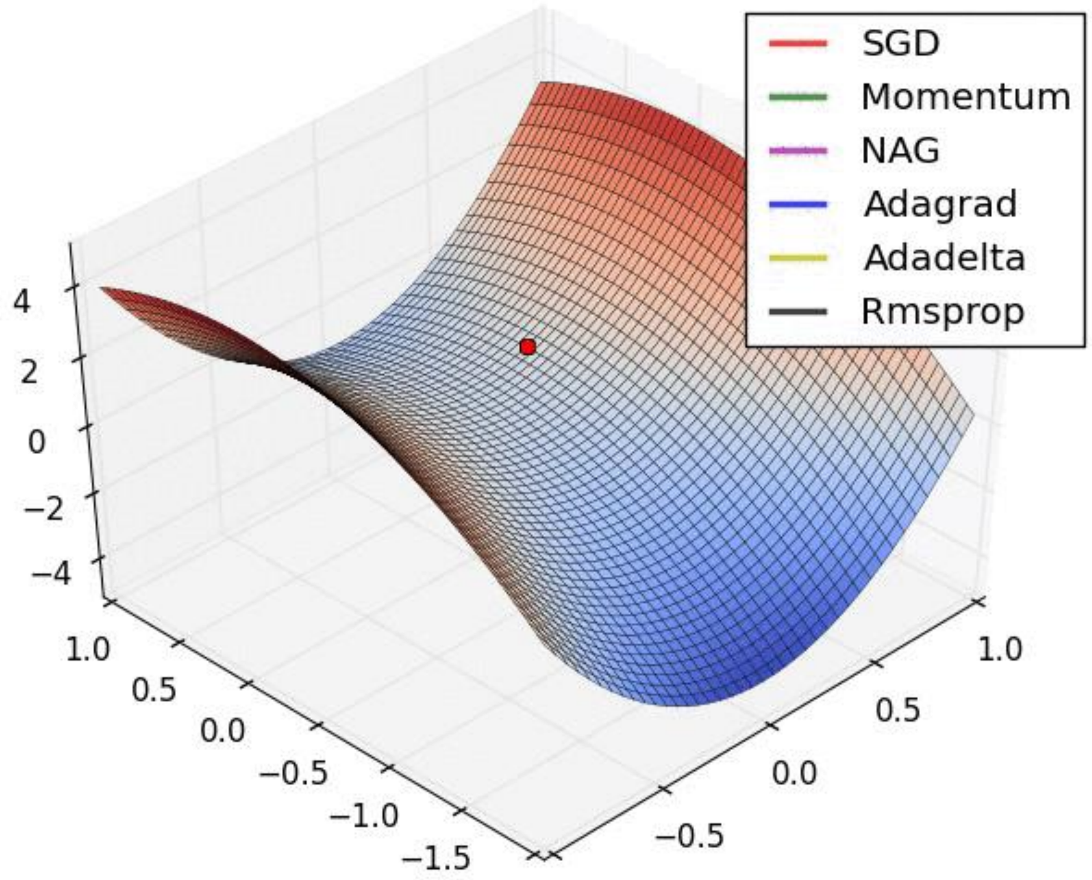
ミニ勾配降下法



$$w_j^{new} = w_j^{old} - \eta \left.\frac{\partial l}{\partial w}\right|_{w=w_j^{old}}$$

最適化アルゴリズム


SGD
Momentum
NAG
Adagrad
Adadelta
Rmsprop


SGD
Momentum
NAG
Adagrad
Adadelta
Rmsprop
4
2
0
−2
−4
1.0
0.5
0.0
−0.5
−1.0
−1.5
1.0
0.5
0.0
−0.5

機械学習

機械学習

回帰分析

ロジスティック回帰

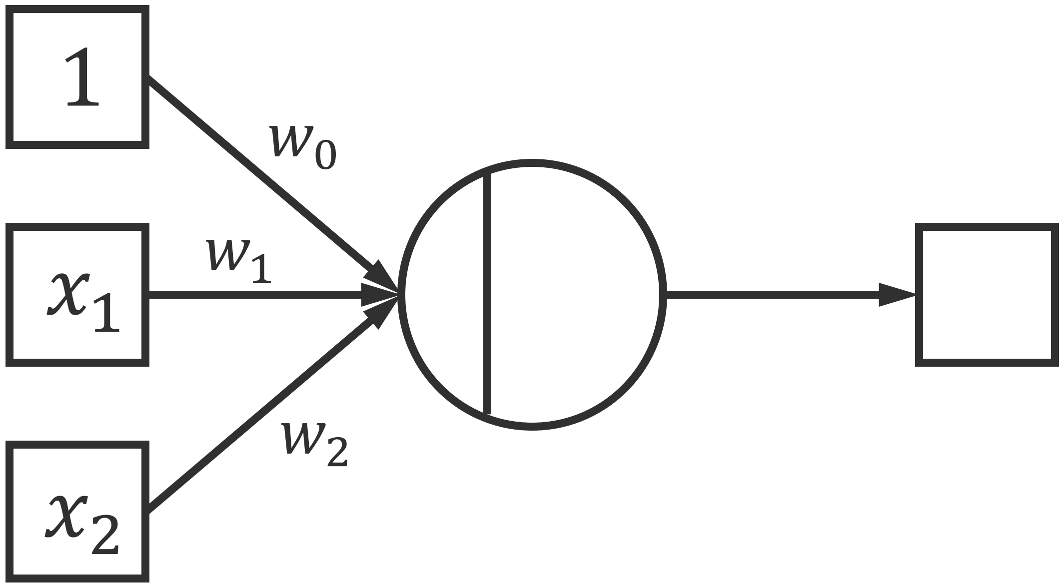
パーセプトロン・ADALINE

勾配降下法

ニューラルネットワーク

ニューラルネットワーク

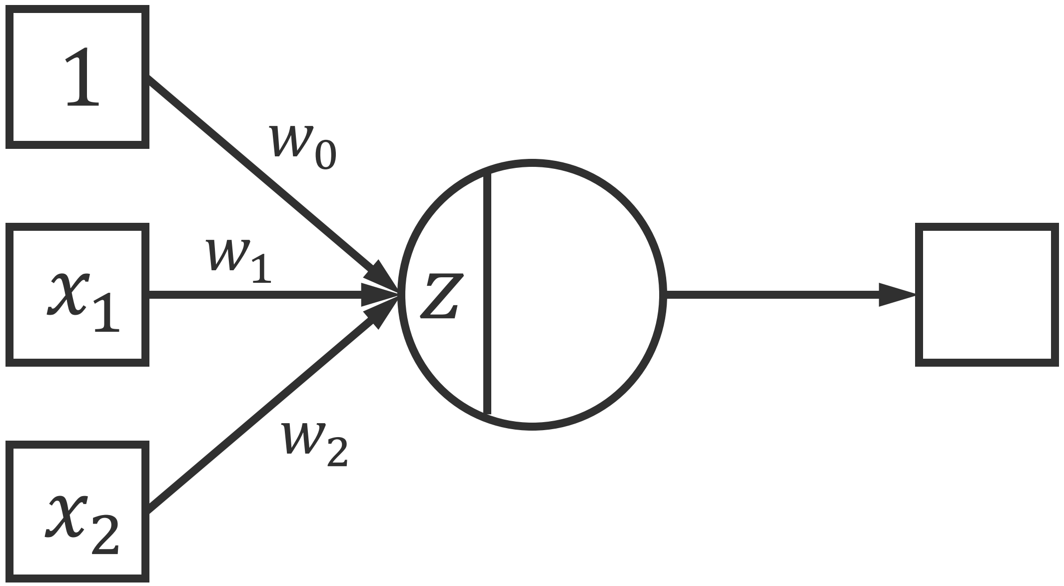
入力　x_1　　x_2



ニューラルネットワーク

入力 x_1 x_2

処理 $z = w_0 + w_1 x_1 + w_2 x_2$

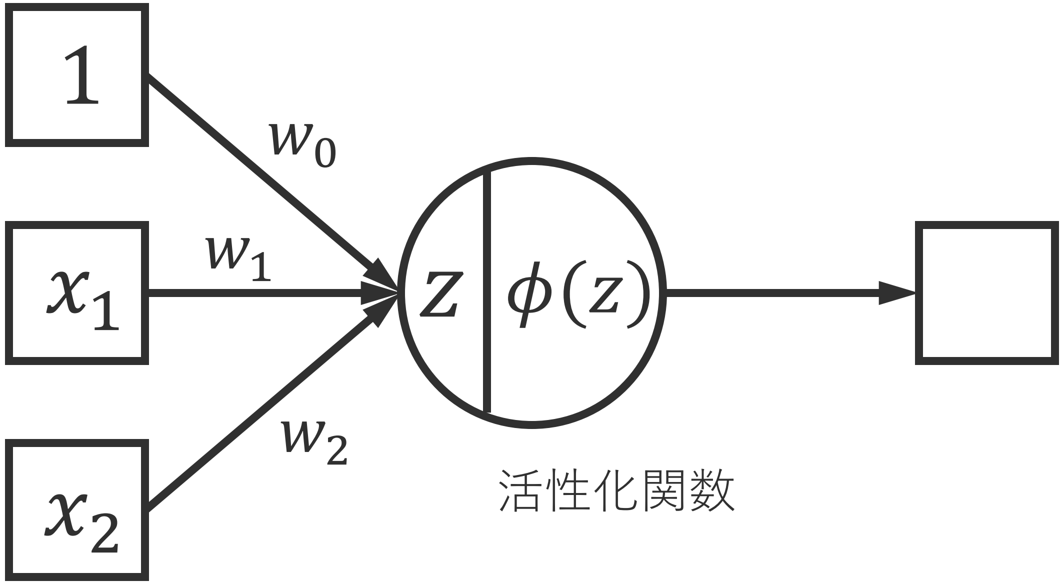


ニューラルネットワーク

入力 x_1 x_2

処理 $z = w_0 + w_1 x_1 + w_2 x_2$

処理 $y = \phi(z)$



アルゴリズム	活性化関数
回帰分析	恒等関数
ロジスティック回帰	シグモイド関数
パーセプトロン	ステップ関数
ニューラルネットワーク	ReLU 関数、シグモイド関数など

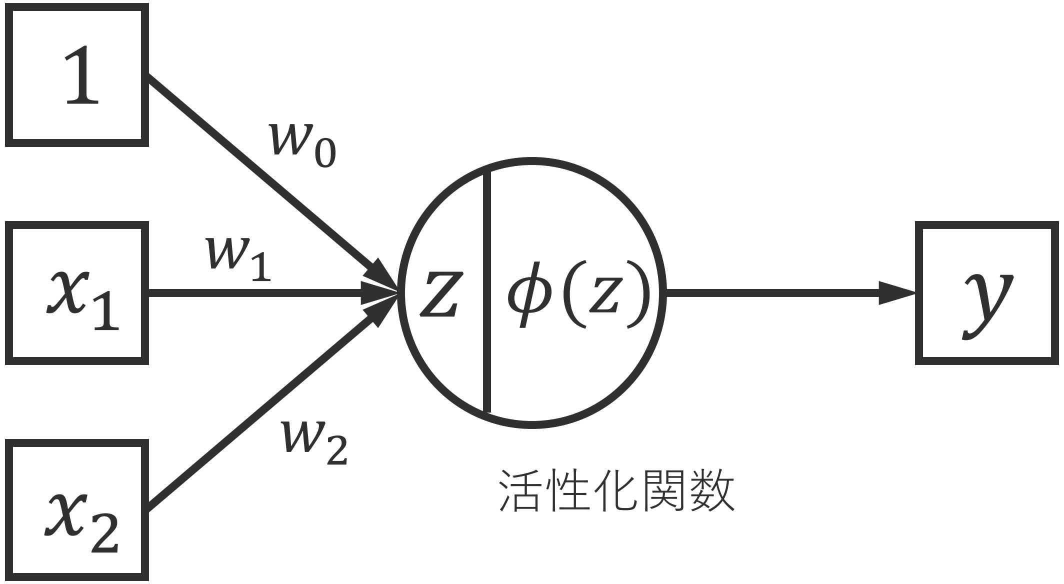
ニューラルネットワーク

入力 x_1 x_2

処理 $z = w_0 + w_1 x_1 + w_2 x_2$

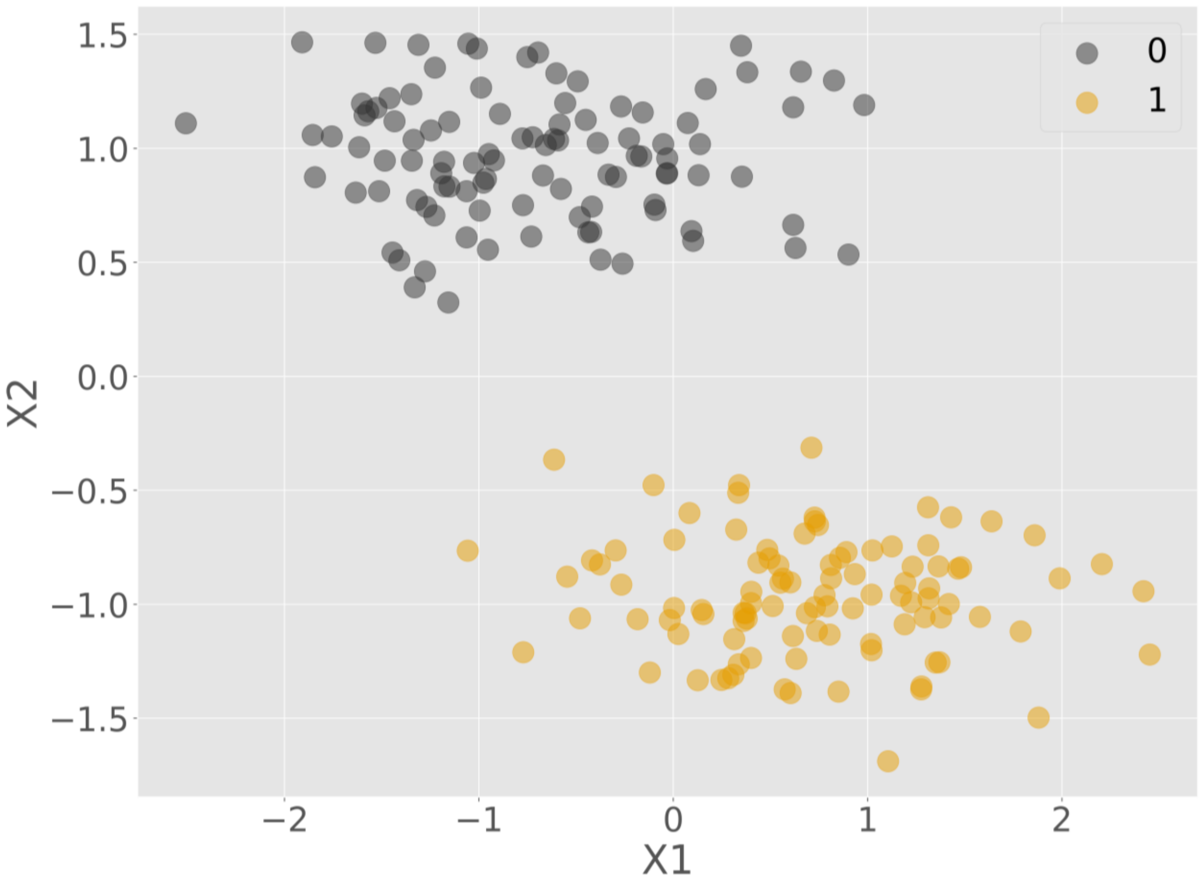
処理 $y = \phi(z)$

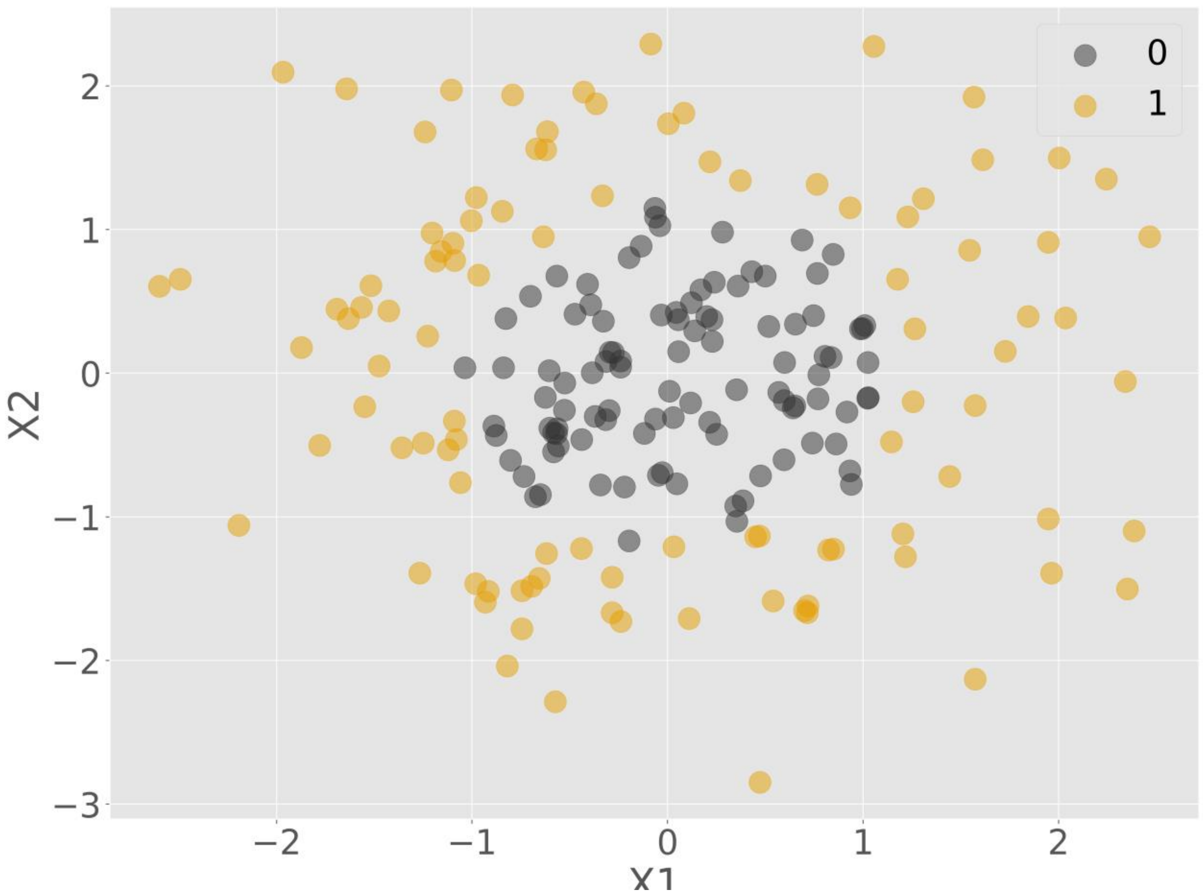
出力 y



アルゴリズム	活性化関数
回帰分析	恒等関数
ロジスティック回帰	シグモイド関数
パーセプトロン	ステップ関数
ニューラルネットワーク	ReLU 関数、シグモイド関数など

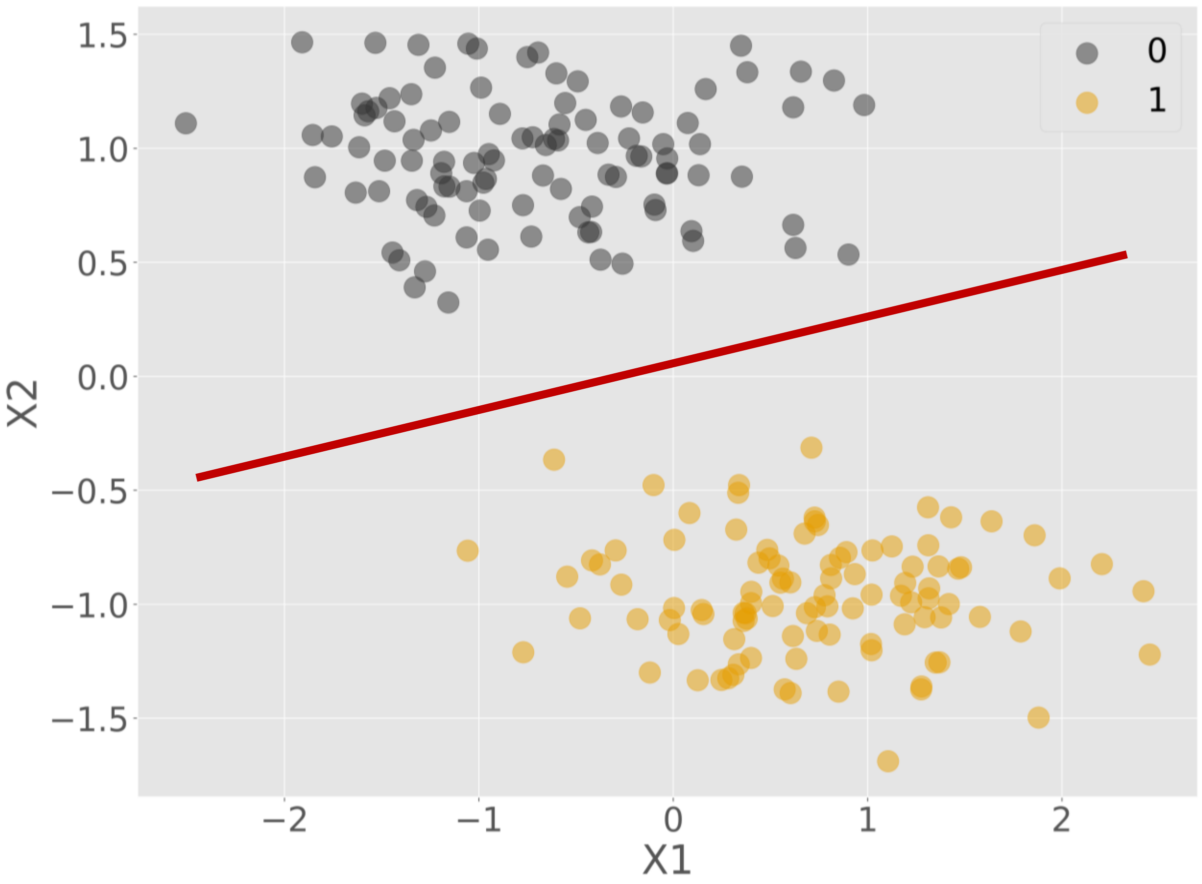
線形分離可能



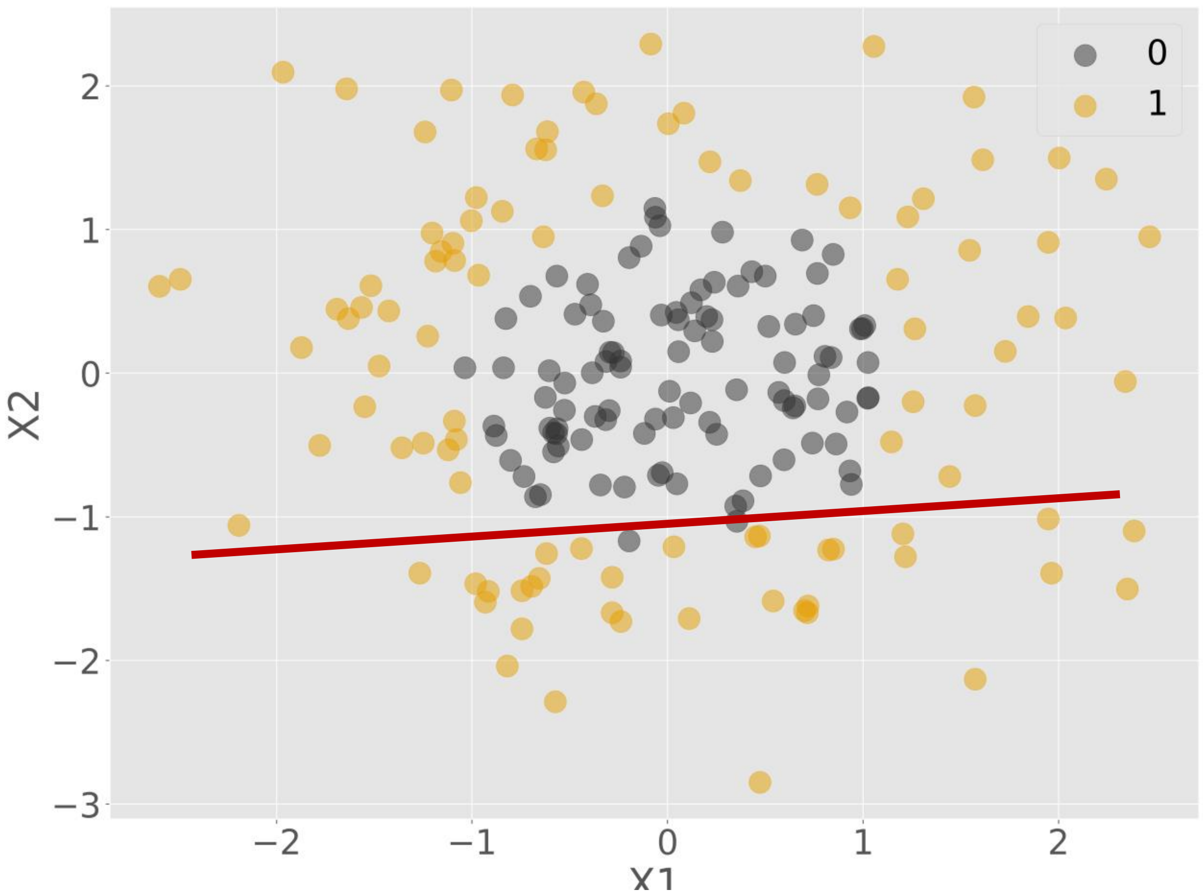


線形分離可能

線形分離可能

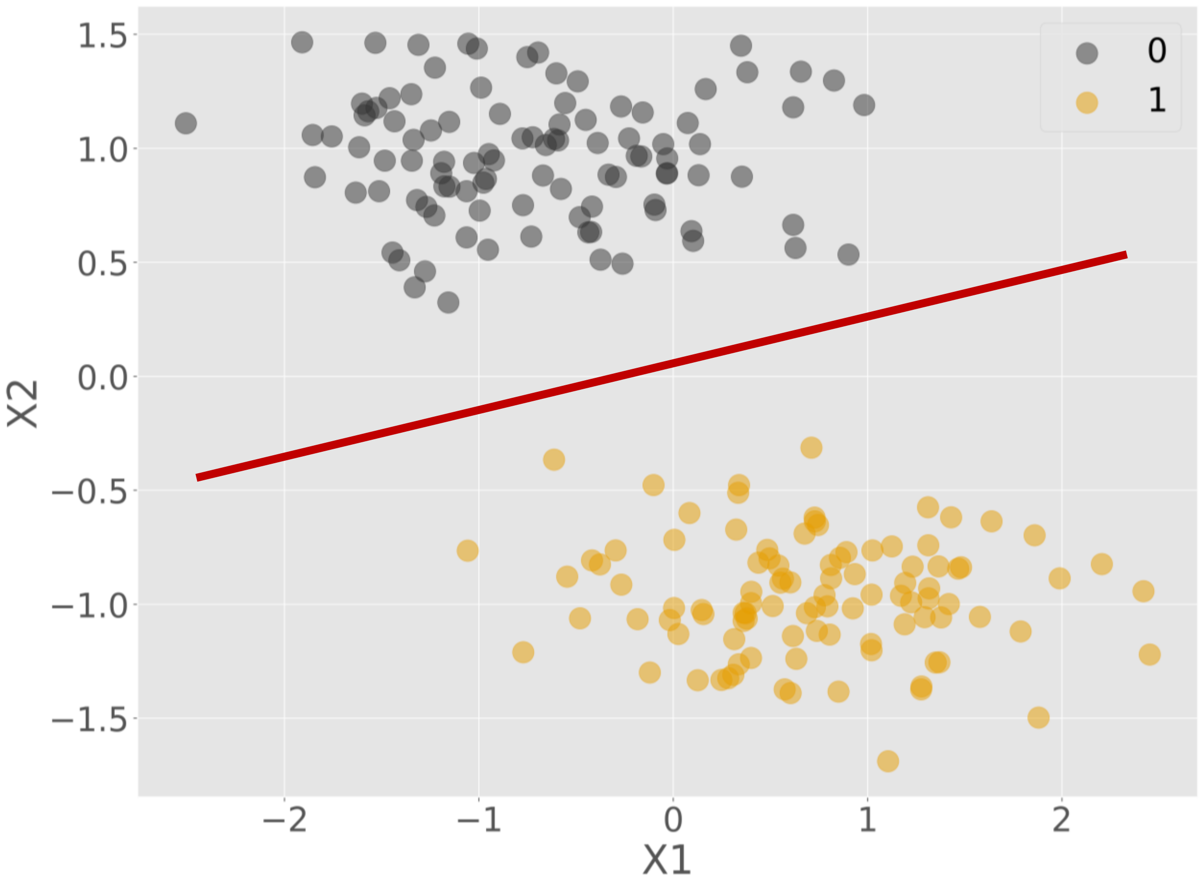


線形分離不可能

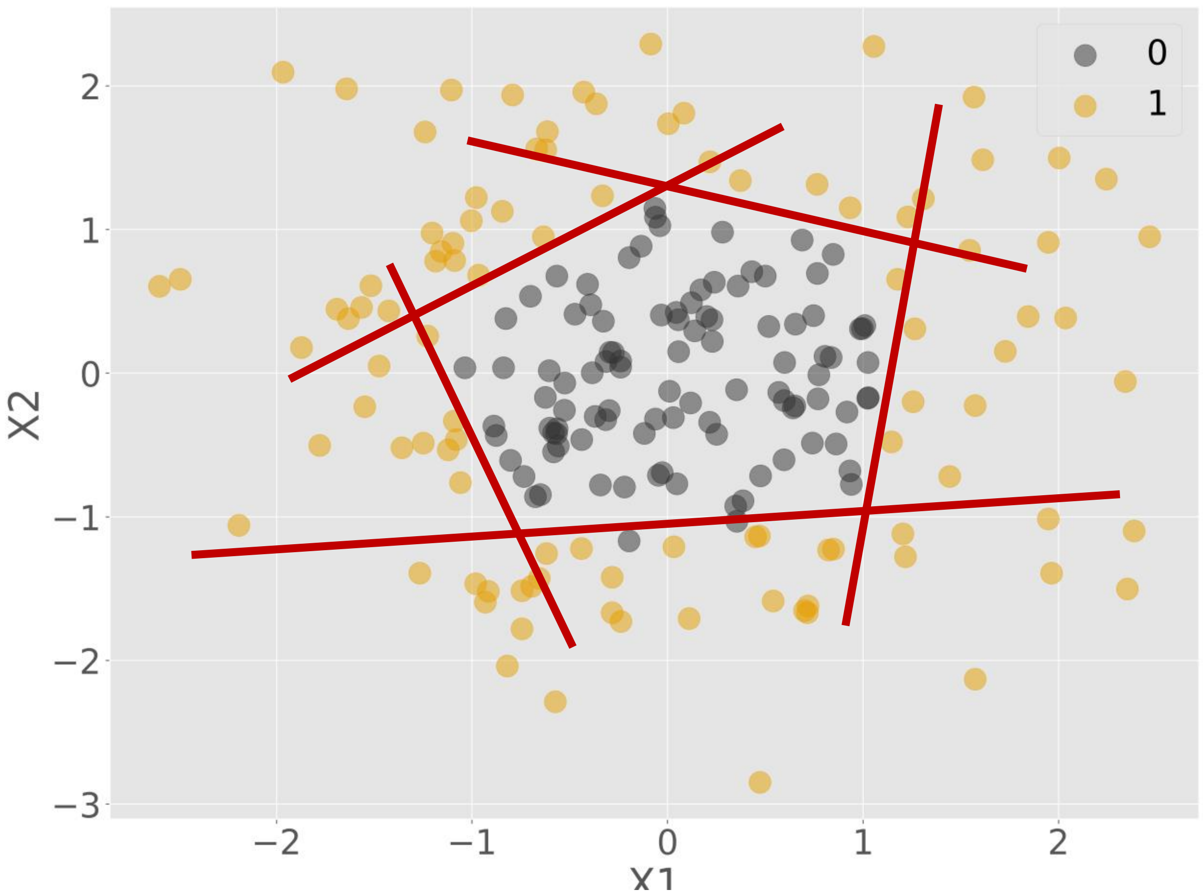


線形分離可能

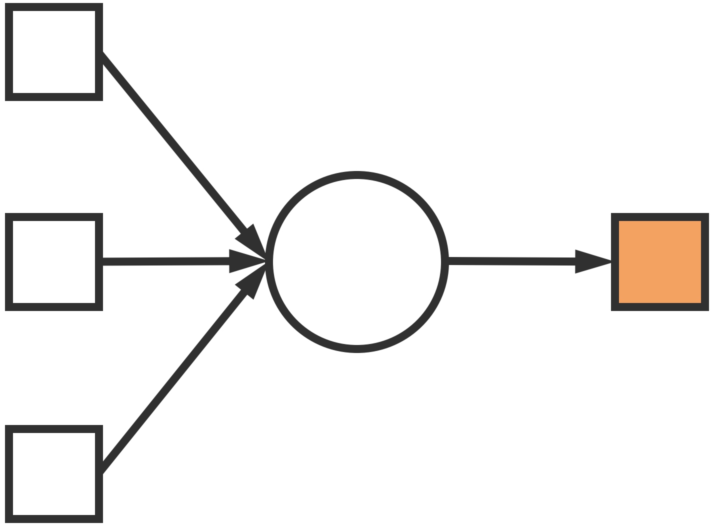
線形分離可能



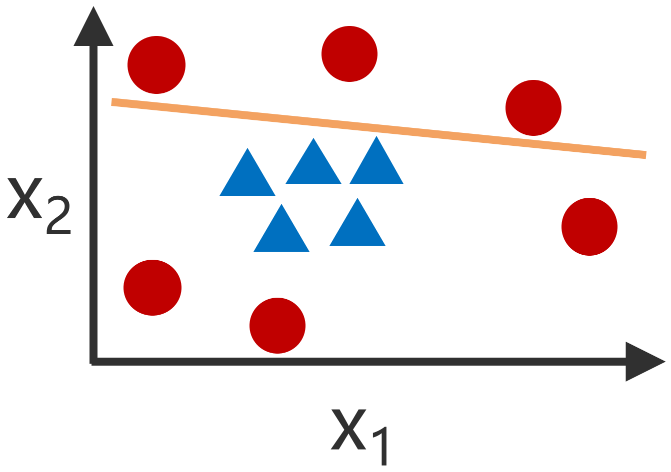
線形分離不可能



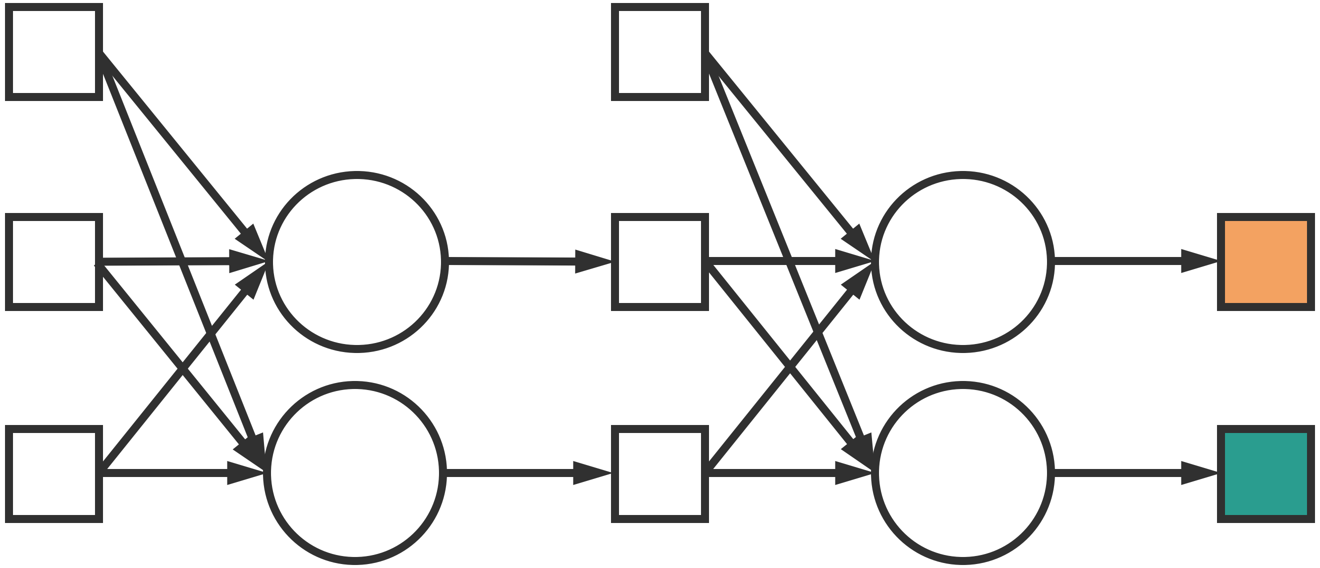
ニューラルネットワーク



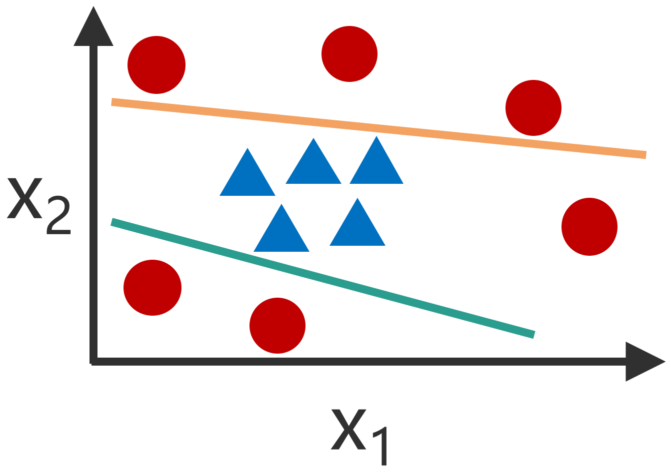
※ 図はイメージである。



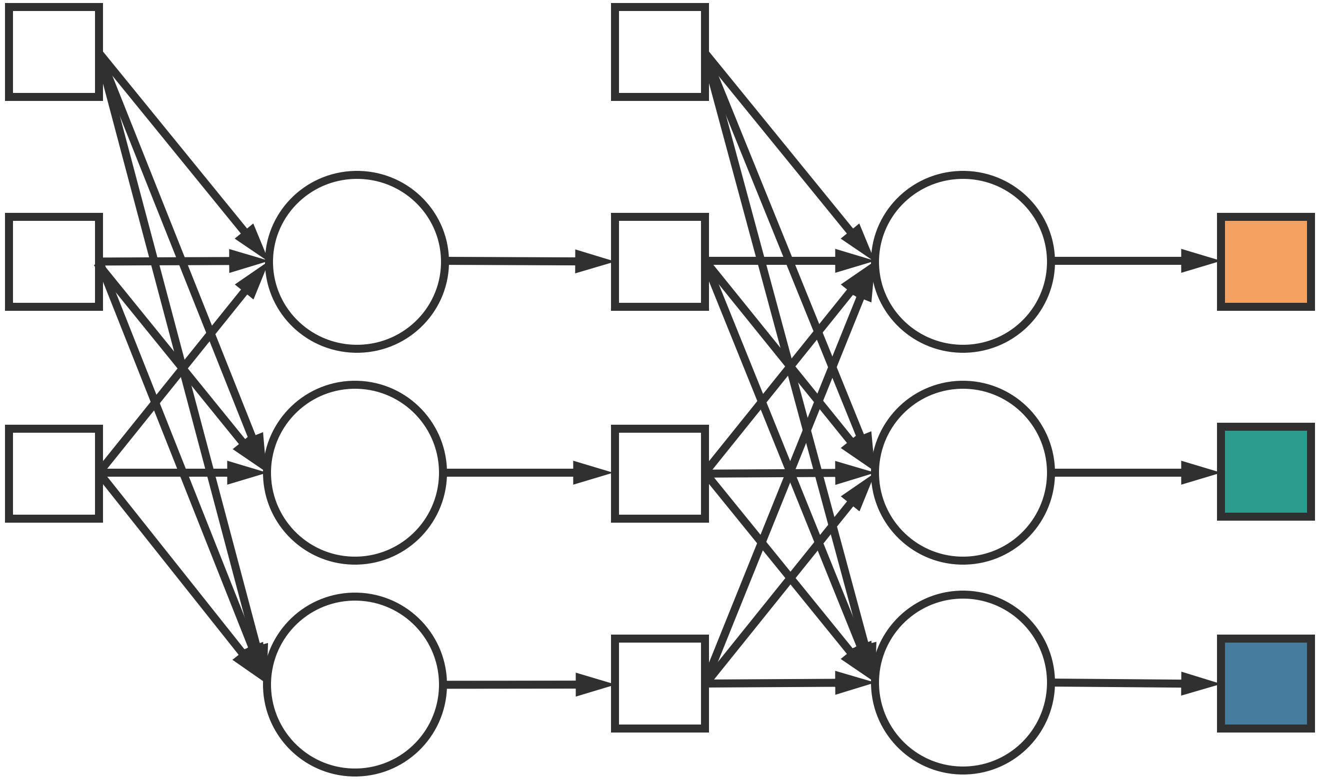
ニューラルネットワーク



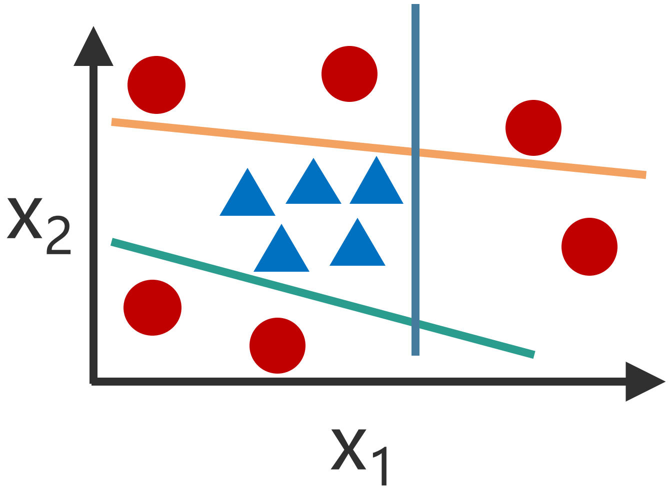
※ 図はイメージである。



ニューラルネットワーク

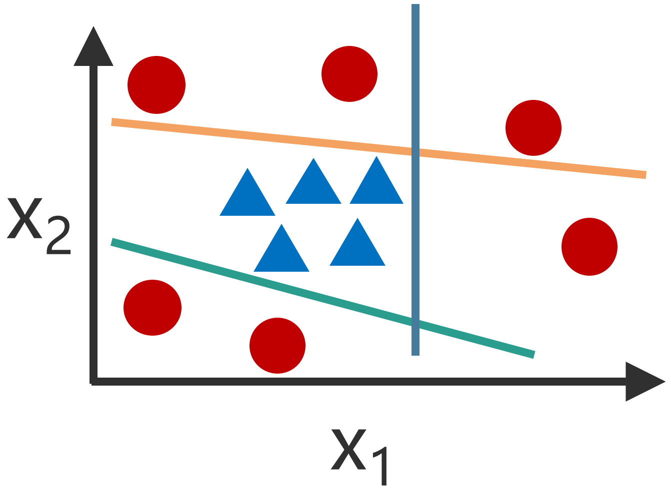


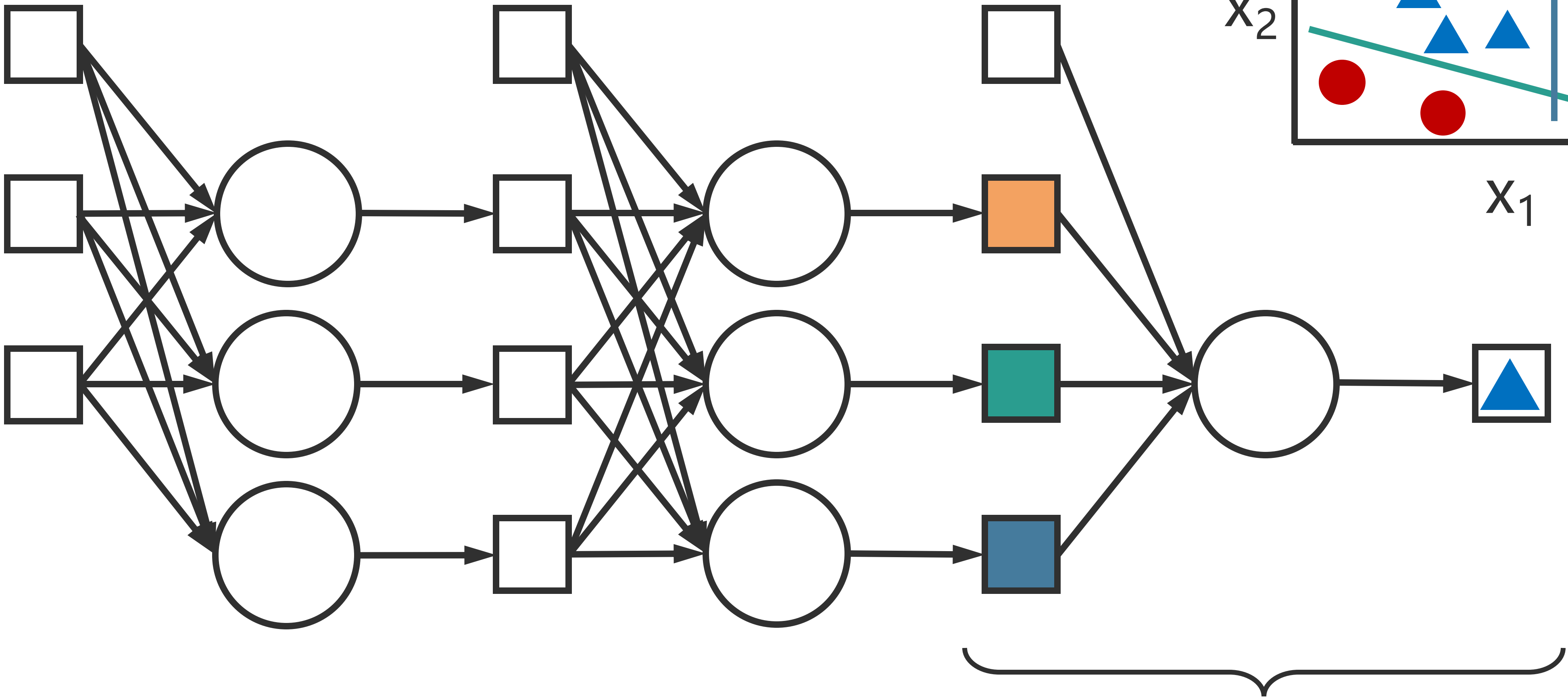
※ 図はイメージである。



ニューラルネットワーク

※ 図はイメージである。

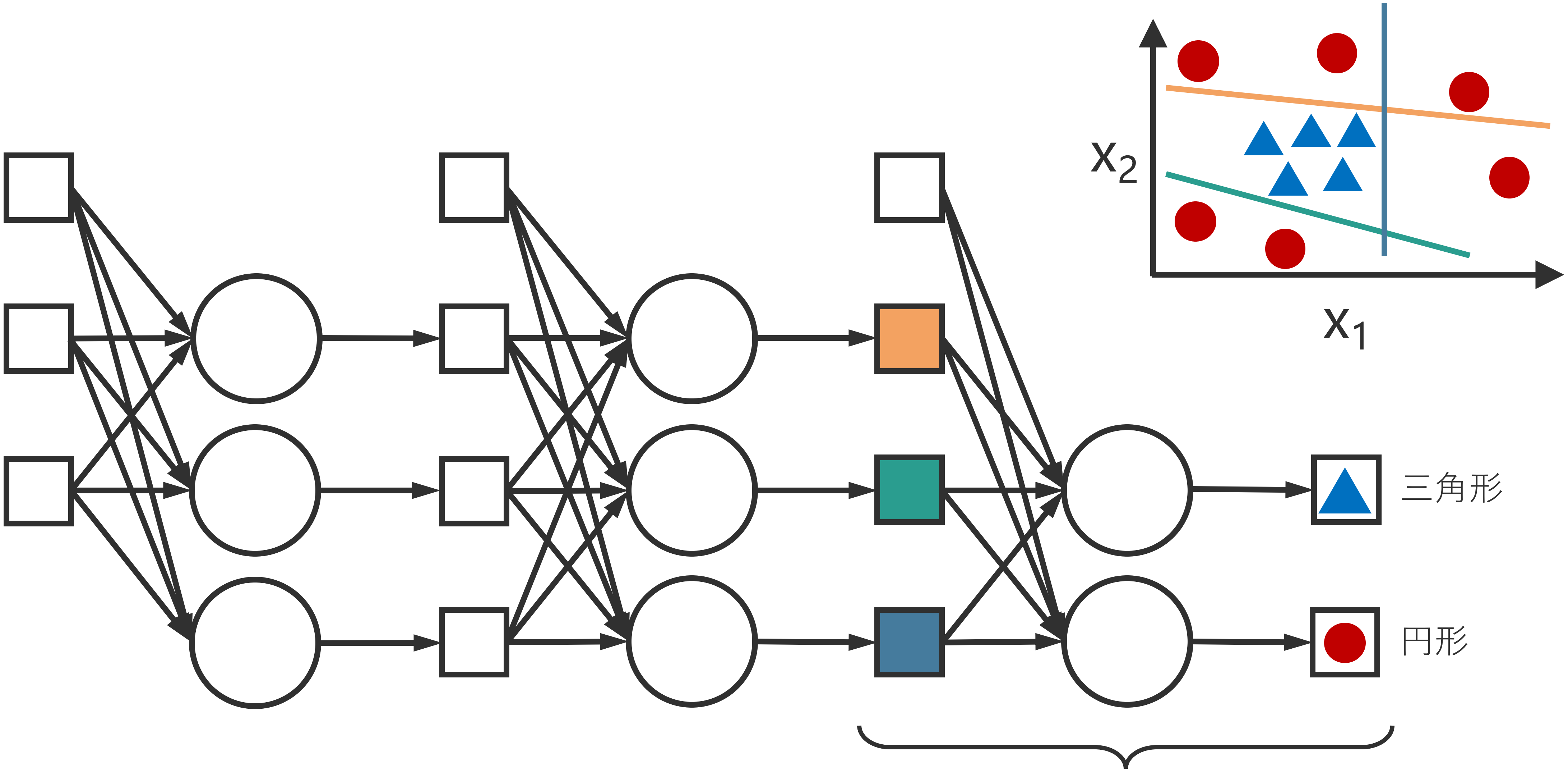




分類対象のスコアを自動的に計算できるように出力数を調整

ニューラルネットワーク

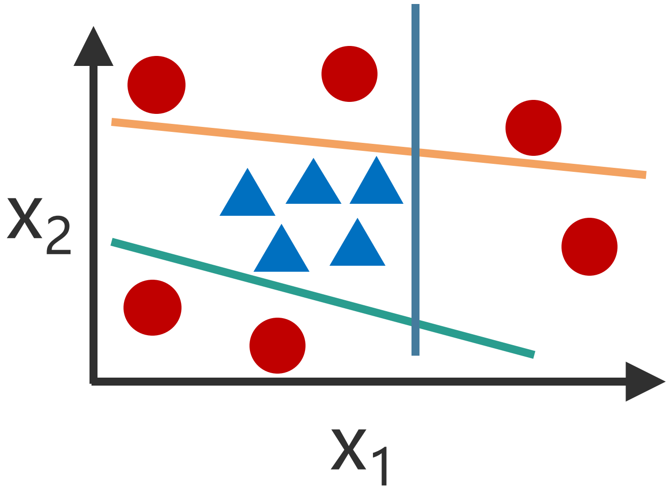
※ 図はイメージである。

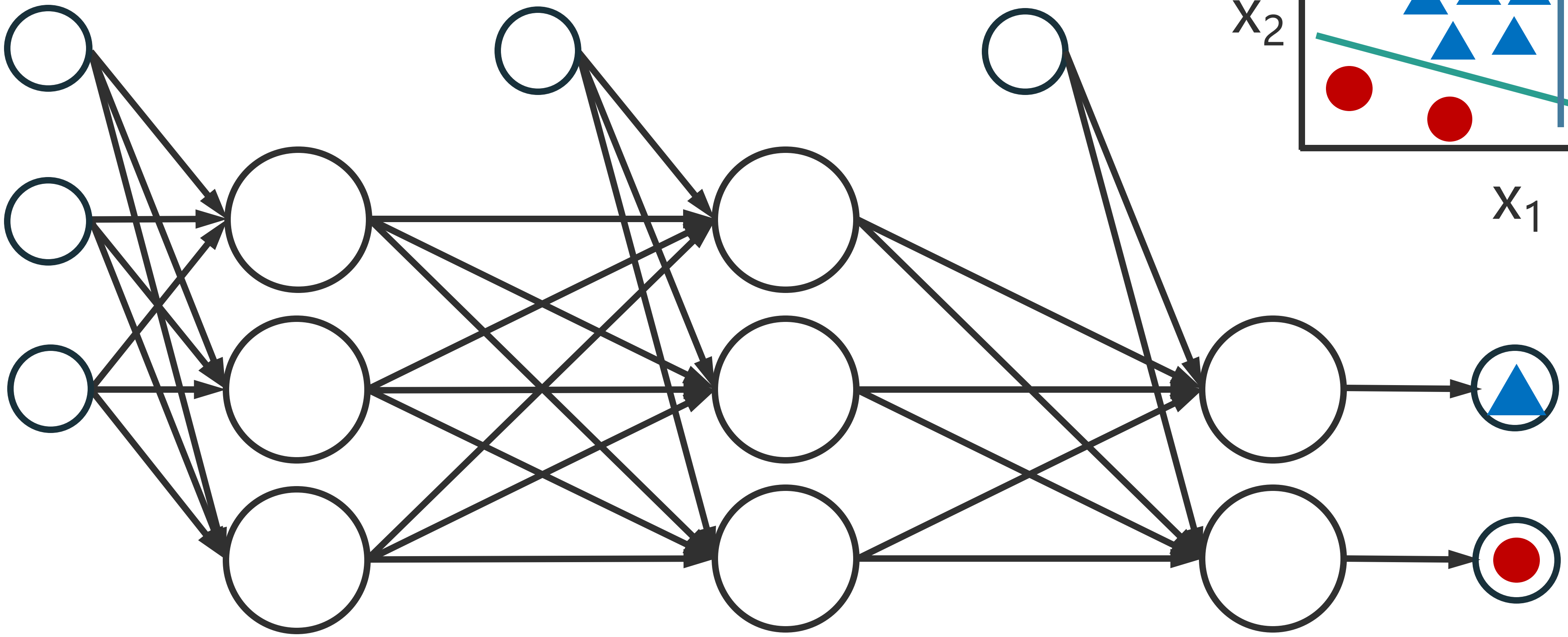


分類対象のスコアを自動的に計算できるように出力数を調整

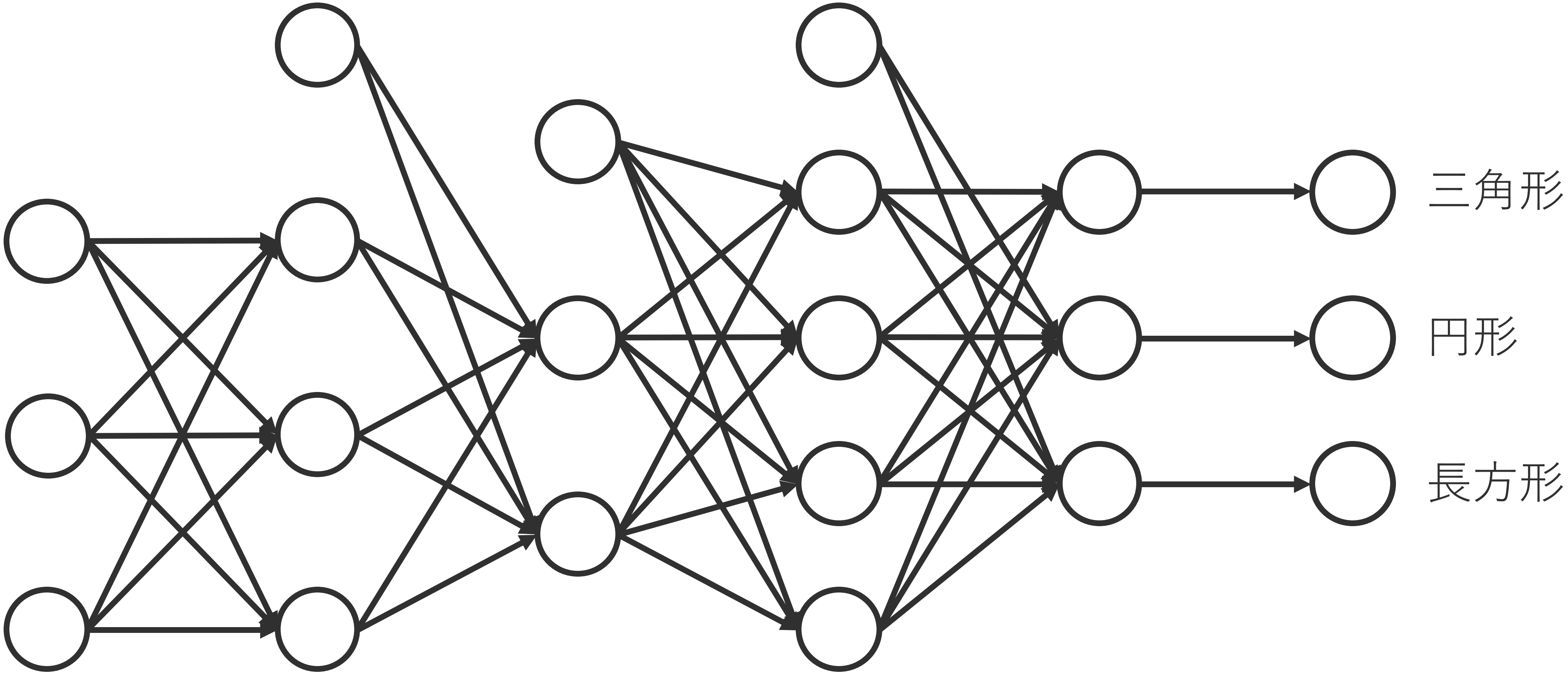
ニューラルネットワーク

※ 図はイメージである。

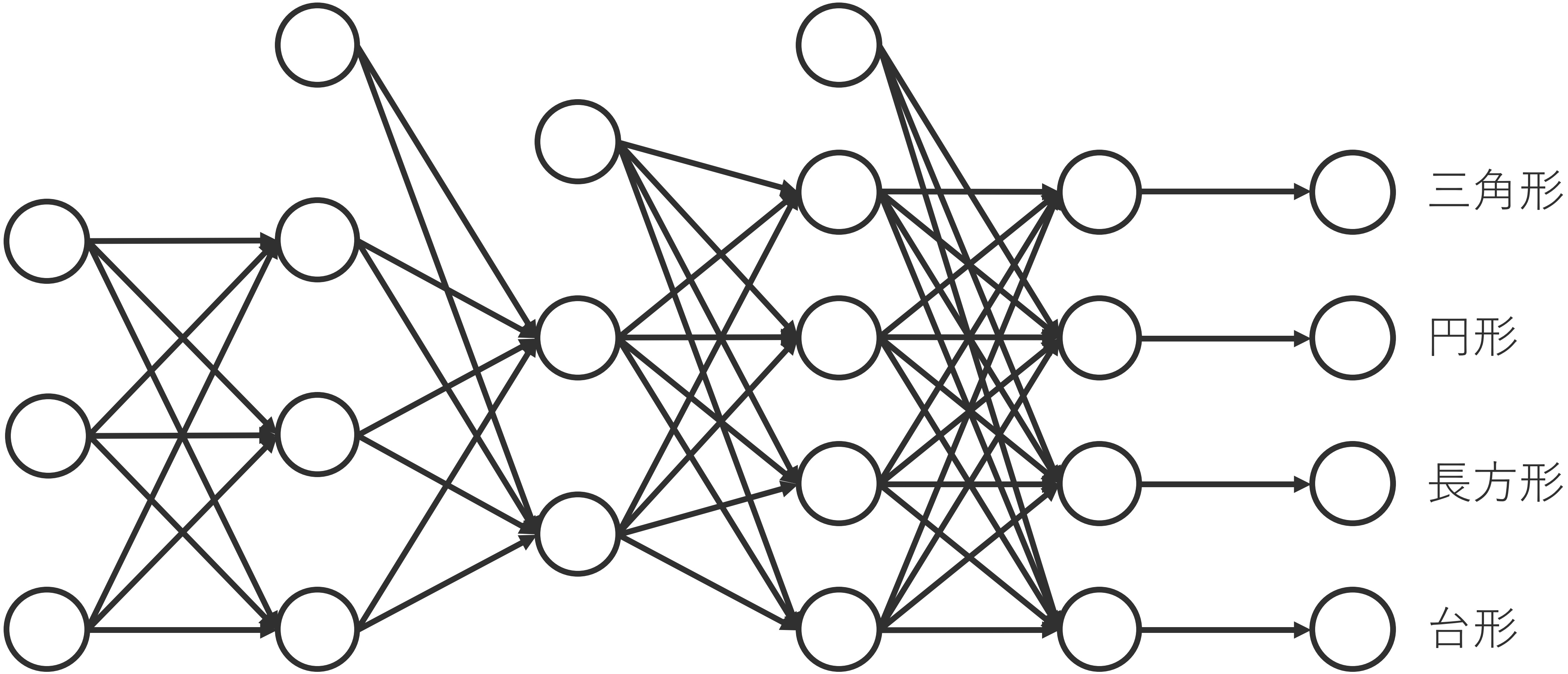




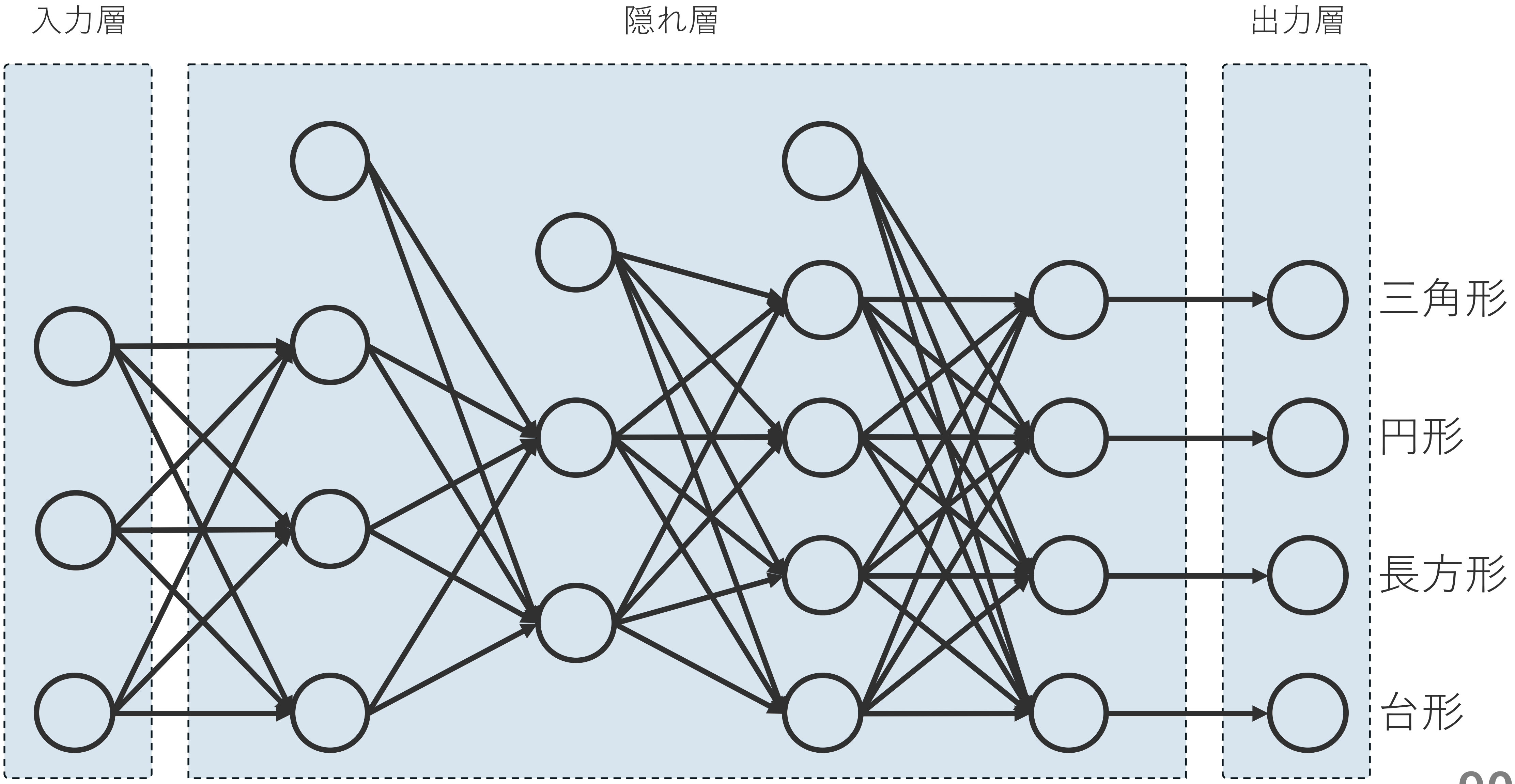
ニューラルネットワーク



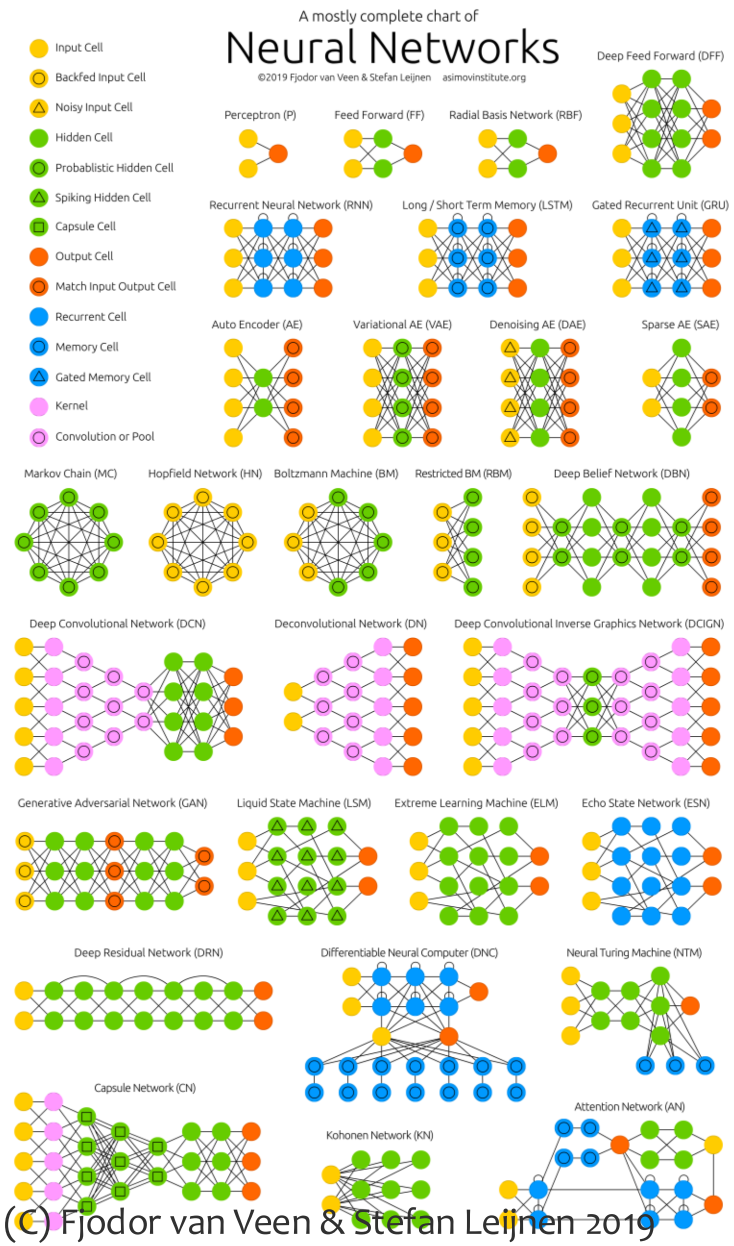
ニューラルネットワーク



ニューラルネットワーク



ニューラルネットワークの種類



(C) Fjodor van Veen & Stefan Leijnen 2019

- 多層化パーセプトロン
  入力から出力まですべての層で完全に結合されているニューラルネットワーク。分類や回帰に用いられる。

- 畳み込みニューラルネットワーク（CNN）
  畳み込み演算（移動平均）により画像の特徴を抽出する層を加えたニューラルネットワーク。画像分類や物体検出に広く使われている。

- 再帰型ニューラルネットワーク（RNN）
  過去の情報を記録するパラメータを加えたニューラルネットワーク。時系列データを利用した予測などに使われる。LSTM や GRU などの改良型がある。

- 敵対的生成ネットワーク（GAN）
  敵対的生成ネットワークは、2 つのニューラルネットワーク（生成器と識別器）を対立的に学習させ、その後、生成器だけを利用してデータ生成させる。