

# 深層学習基本語彙 (40分で！ 図付き！)

Graham Neubig  
奈良先端科学技術大学院大学 (NAIST)

# 予測問題

$x$  が与えられ  $y$  を予測

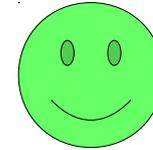
# 例：感情極性推定

x

I am happy

y

+1



I am sad

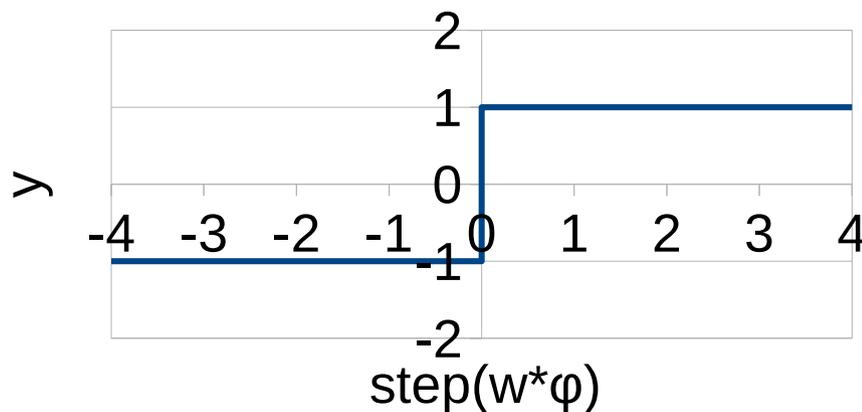
-1



# 線形分類器 (linear classifier)

$$y = \text{step}(w^T \phi(x))$$

- $x$ : 入力
- $\phi(x)$ : 素性列
- $w$ : 重み列
- $y$ : 予測結果、yes なら +1、no なら -1

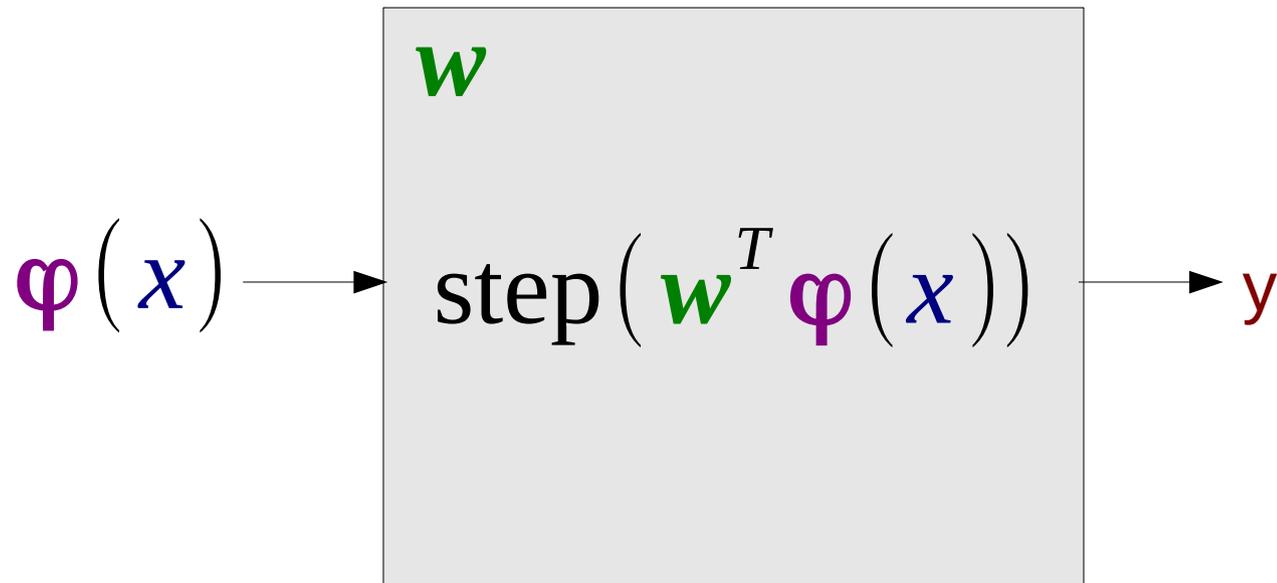


# 素性関数の例：ユニグラム素性

$$\begin{array}{l} \varphi(\text{I am happy}) \rightarrow \left( \begin{array}{c} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ \dots \end{array} \right) \begin{array}{l} \text{I} \\ \text{was} \\ \text{am} \\ \text{sad} \\ \text{the} \\ \text{happy} \\ \dots \end{array} \end{array} \quad \begin{array}{l} \varphi(\text{I am sad}) \rightarrow \left( \begin{array}{c} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ \dots \end{array} \right) \end{array}$$

# パーセプトロン

- 素性の重み付き和を計算し、活性化関数に代入する

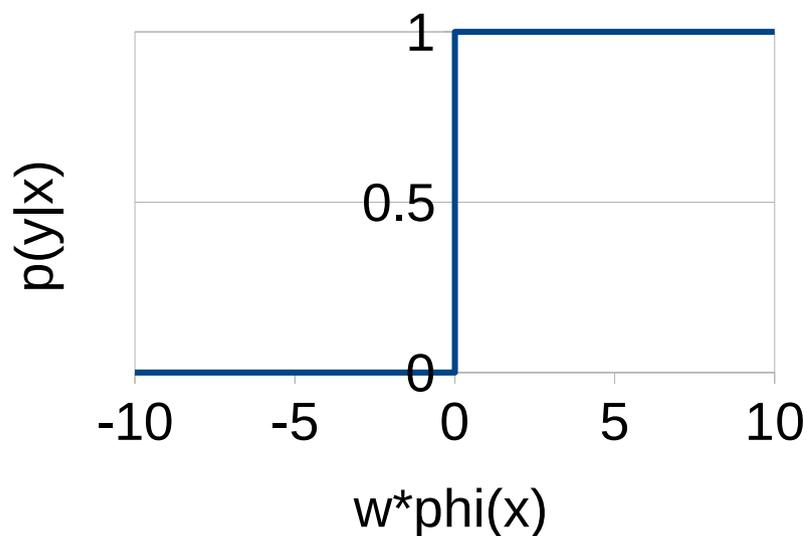


# シグモイド関数（ロジスティック関数）

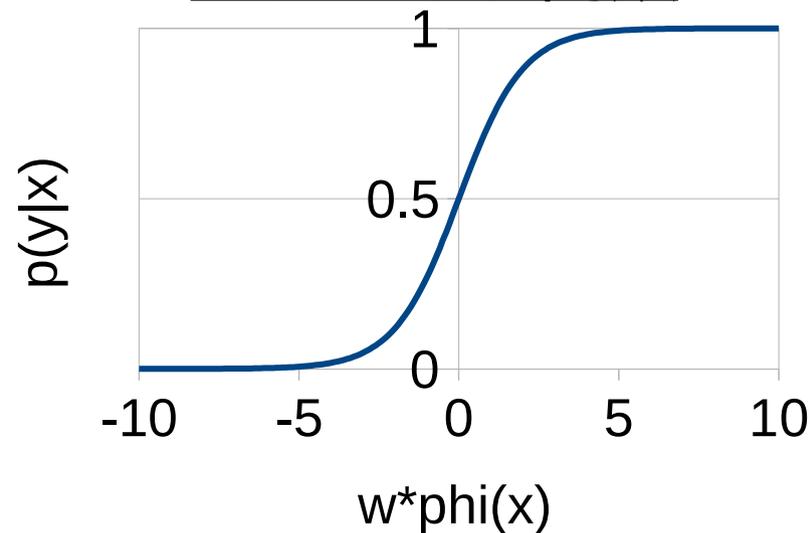
- シグモイド関数はステップ関数を柔らかくしたものの

$$P(y=1|x) = \frac{e^{w \cdot \varphi(x)}}{1 + e^{w \cdot \varphi(x)}}$$

ステップ関数



シグモイド関数



- 不確実性（確率）を考慮
- 微分可能

# ロジスティック回帰 (logistic regression; LR)

- 条件付き尤度最大化基準で学習
- $x$ が与えられたときの正解  $y$  の条件付き尤度を最大化するパラメータ  $w$  を獲得

$$\hat{w} = \operatorname{argmax}_w \prod_i P(y_i | x_i; w)$$

- 解き方は？

# 確率的勾配降下法 (stochastic gradient descent; SGD)

- ロジスティック回帰を含む確率モデルのための学習アルゴリズム

$w = 0$

**for** /iterations

**for each** labeled pair  $x, y$  in the data

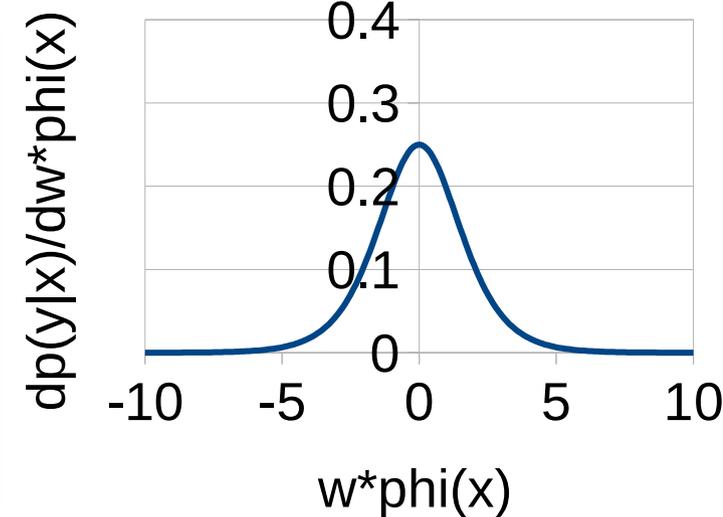
$w += \alpha * dP(y|x)/dw$

- つまり
  - 各学習例に対して勾配を計算  
( $y$ の確率が上がる方向)
  - その方向へ、学習率  $\alpha$  をかけた分だけ動かす

# シグモイド関数の勾配

- 確率の微分

$$\begin{aligned} \frac{d}{d w} P(y=1|x) &= \frac{d}{d w} \frac{e^{w \cdot \varphi(x)}}{1+e^{w \cdot \varphi(x)}} \\ &= \varphi(x) \frac{e^{w \cdot \varphi(x)}}{(1+e^{w \cdot \varphi(x)})^2} \end{aligned}$$

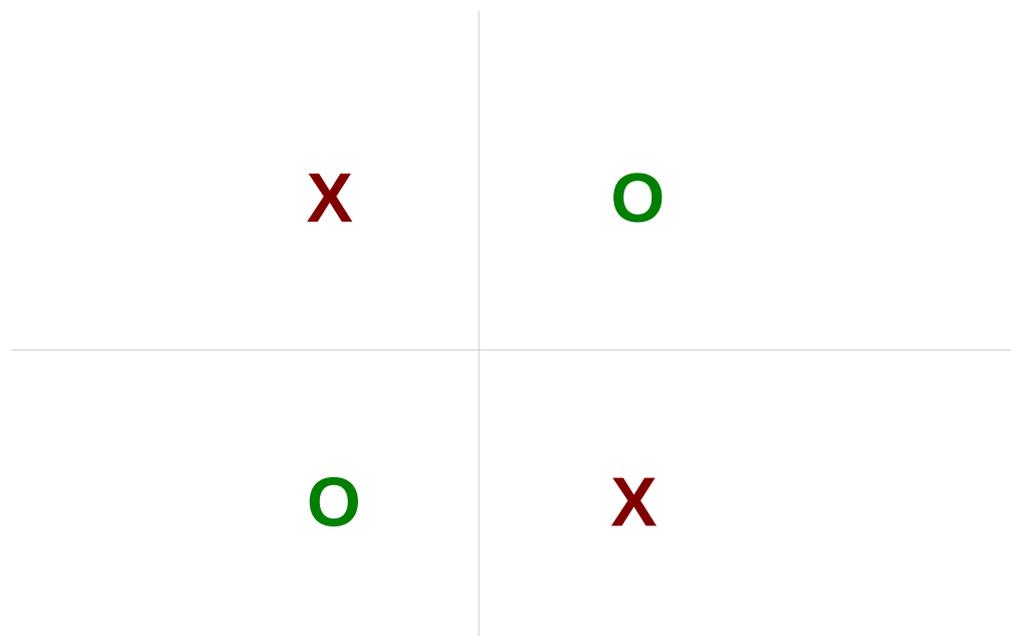


$$\begin{aligned} \frac{d}{d w} P(y=-1|x) &= \frac{d}{d w} \left( 1 - \frac{e^{w \cdot \varphi(x)}}{1+e^{w \cdot \varphi(x)}} \right) \\ &= -\varphi(x) \frac{e^{w \cdot \varphi(x)}}{(1+e^{w \cdot \varphi(x)})^2} \end{aligned}$$

# ニューラルネット

# 問題：線形分類不可能な関数の学習

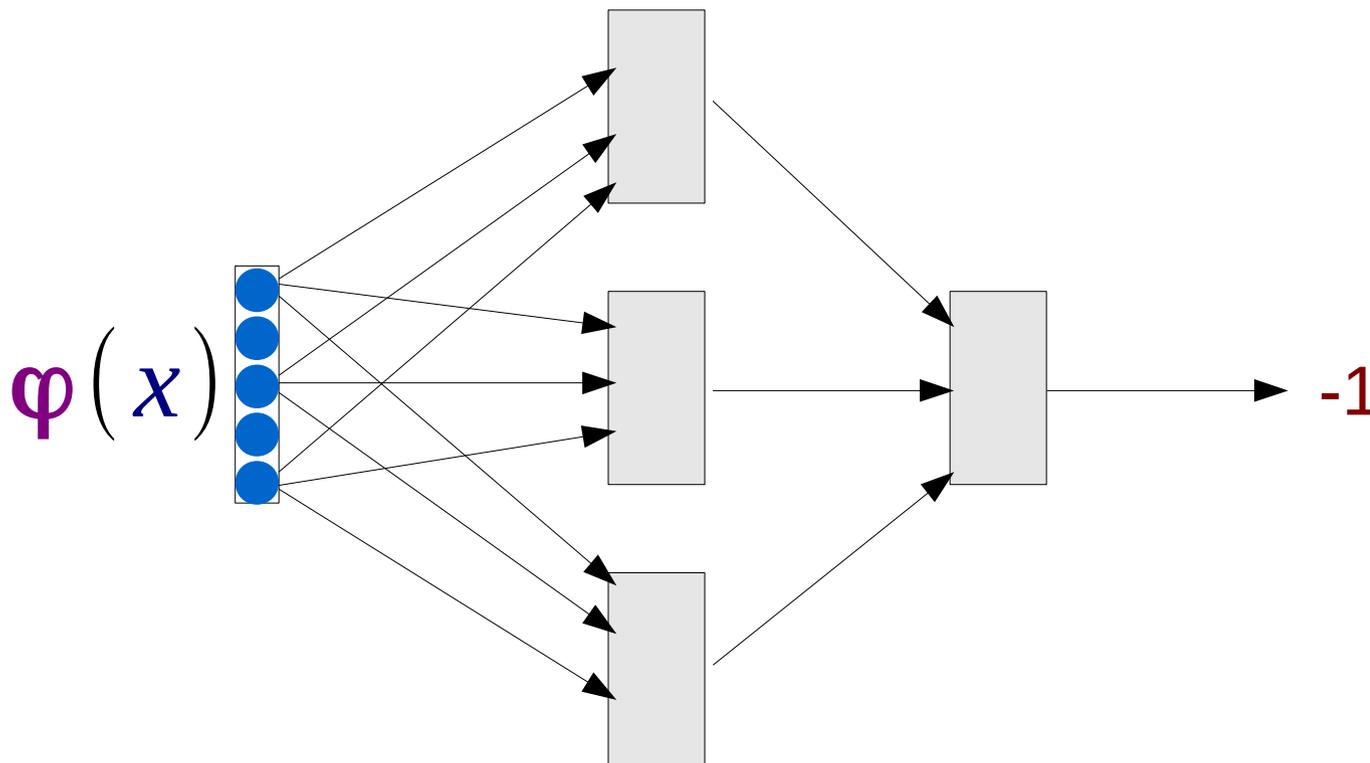
- パーセプトロンは線形分類しかできない



- 例：「I am **not** happy」

# ニューラルネット (多層パーセプトロン)

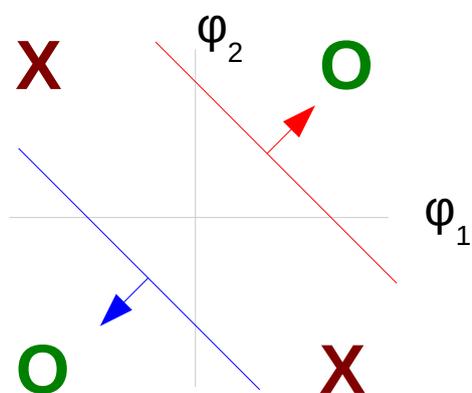
- ニューラルネットで、複数のパーセプトロンを連結



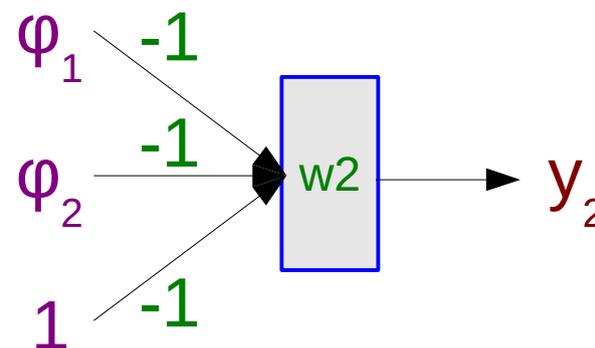
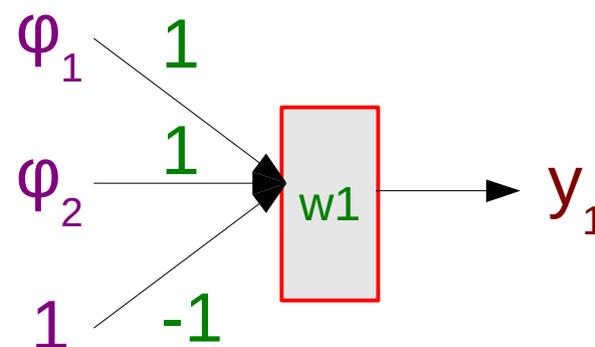
# 例：非線形分類

- 2つの分類器を作成

$$\varphi(x_1) = \{-1, 1\} \quad \varphi(x_2) = \{1, 1\}$$



$$\varphi(x_3) = \{-1, -1\} \quad \varphi(x_4) = \{1, -1\}$$

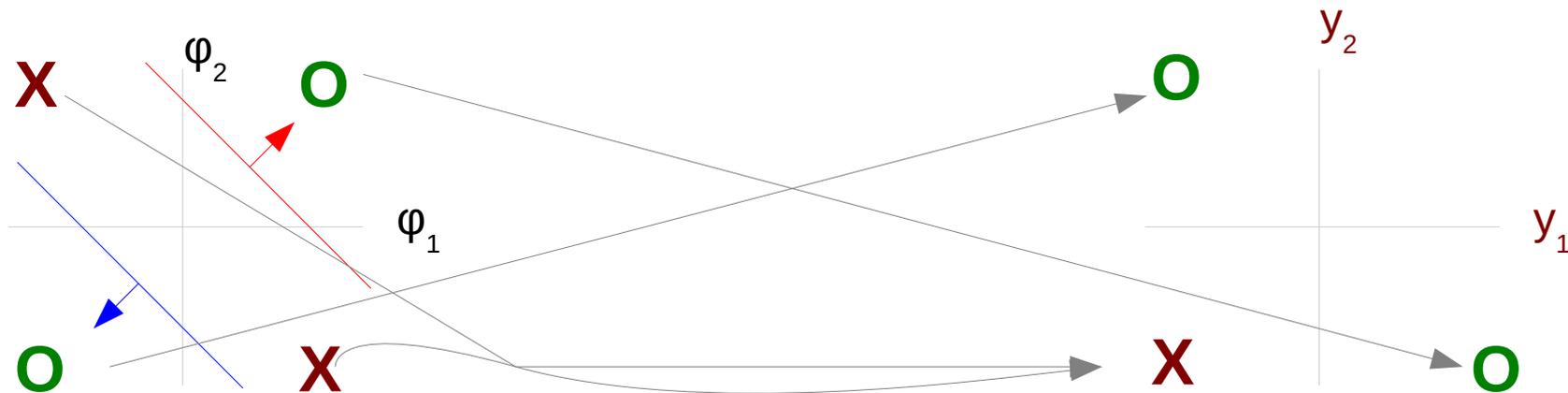


# 例：非線形分類

- 入力を新しい空間へ写像

$\varphi(x_1) = \{-1, 1\}$     $\varphi(x_2) = \{1, 1\}$

$y(x_3) = \{-1, 1\}$

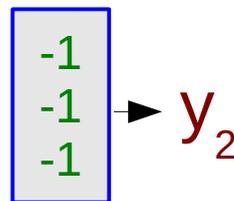
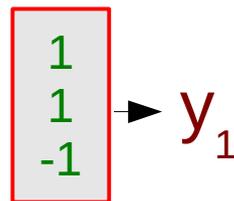


$\varphi(x_3) = \{-1, -1\}$     $\varphi(x_4) = \{1, -1\}$

$y(x_1) = \{-1, -1\}$

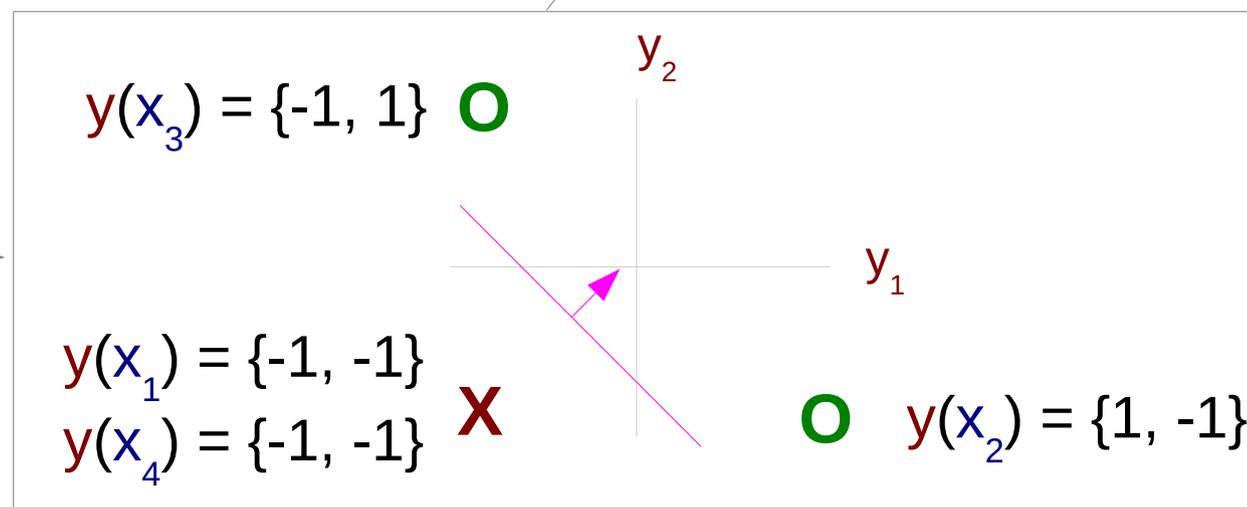
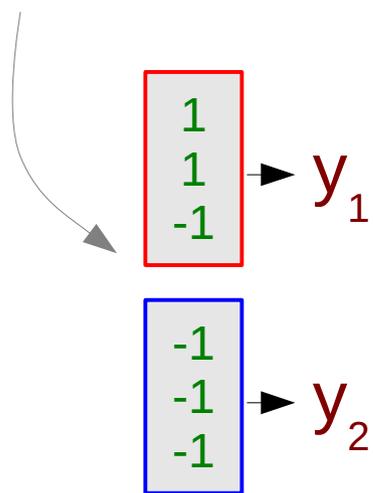
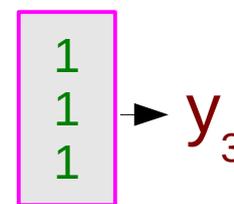
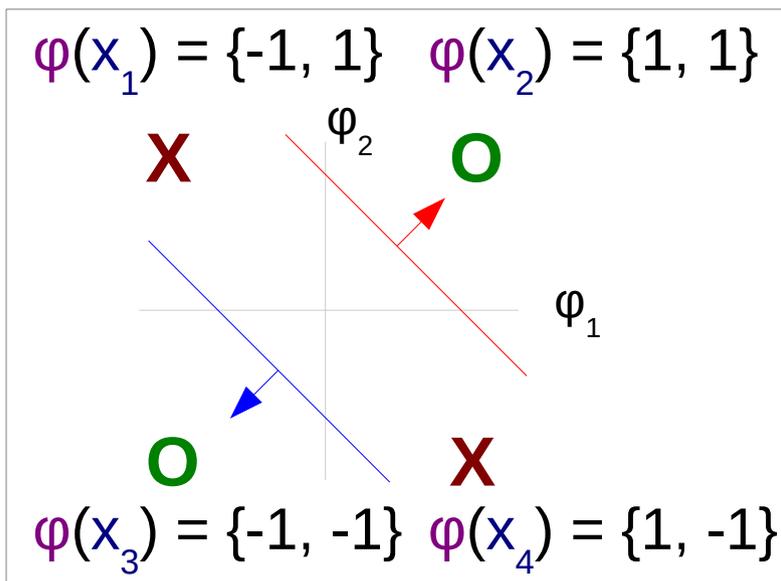
$y(x_2) = \{1, -1\}$

$y(x_4) = \{-1, -1\}$



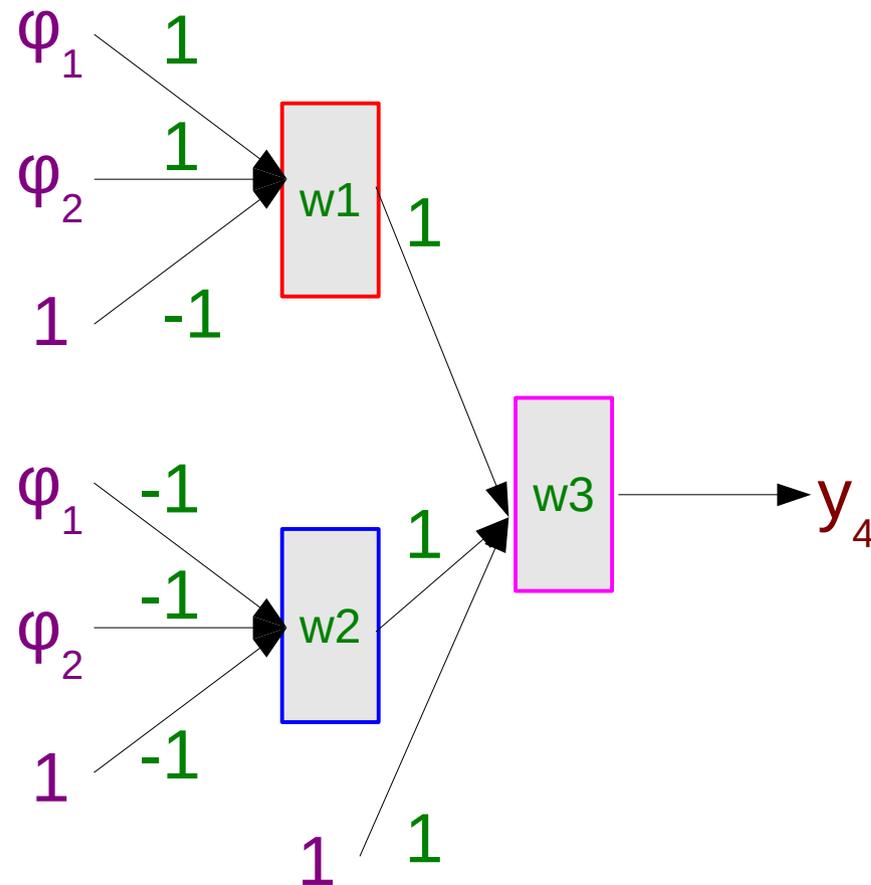
# 例：非線形分類

- 新しい空間で線形分類が可能！



# 例：非線形分類

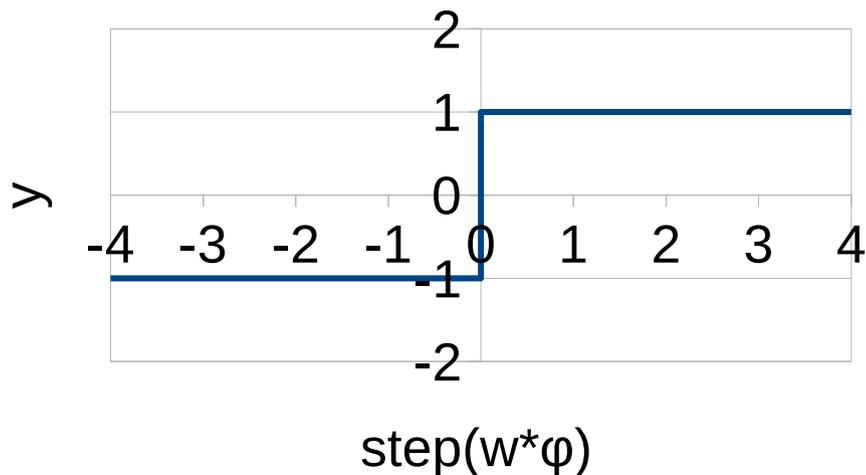
- 最終的なニューラルネット



# 隠れ層活性化関数

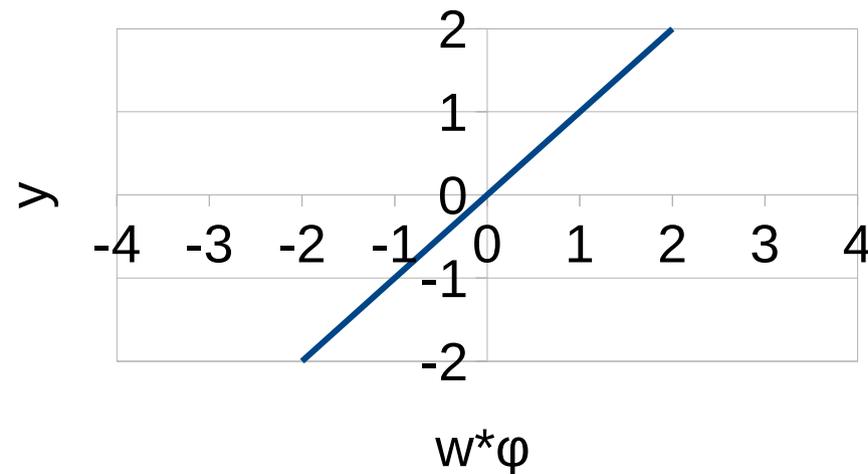
ステップ関数：

- 勾配が計算不可



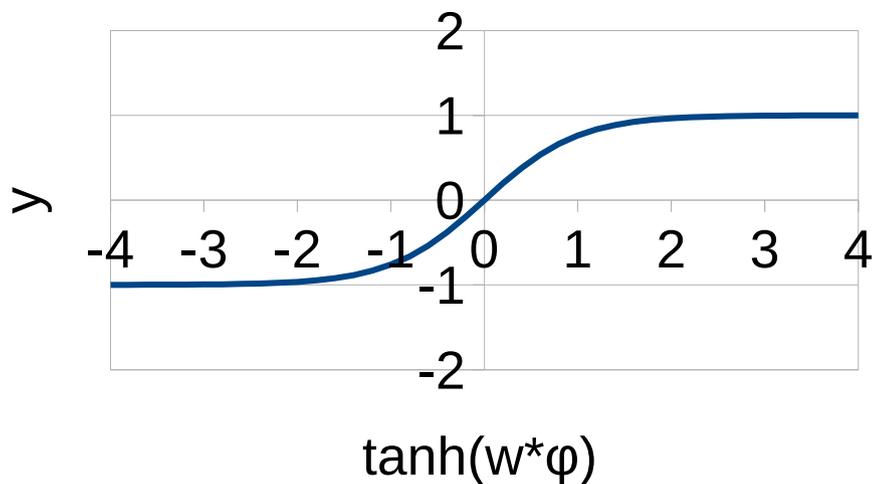
線形関数：

- ネット全体が線形関数



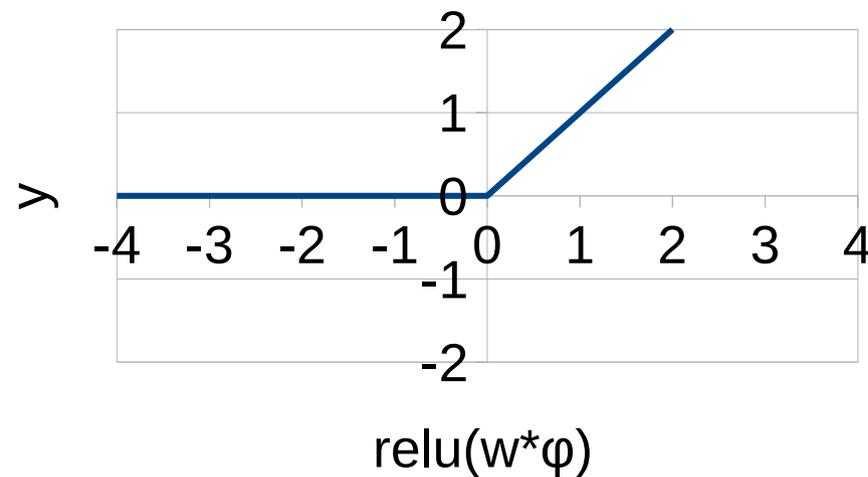
双曲線関数：

標準 (0-1 のシグモイドも)

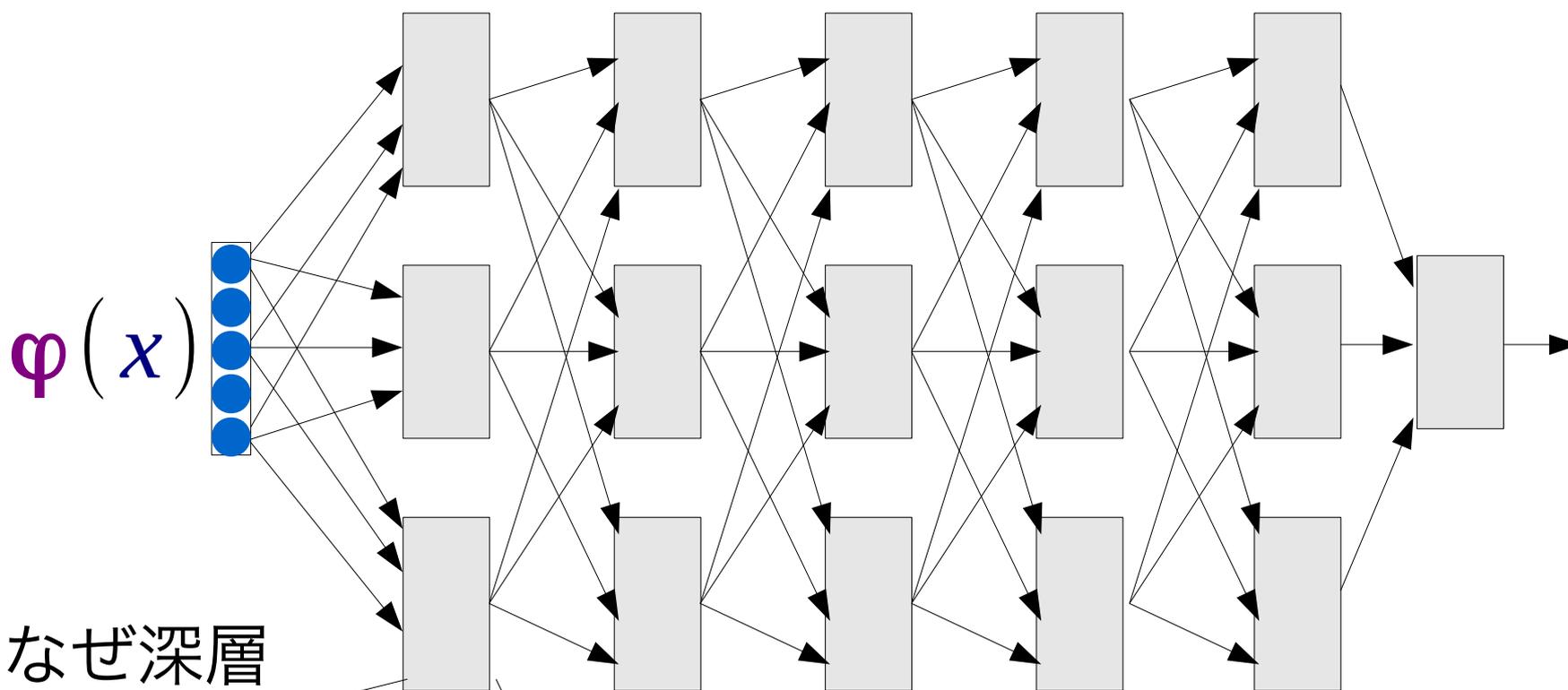


Rectified 線形関数：

+ 大きな値でも勾配が大きい



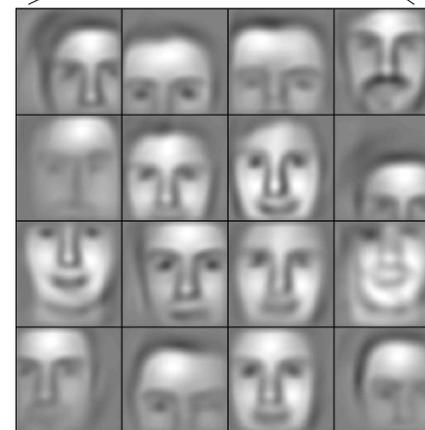
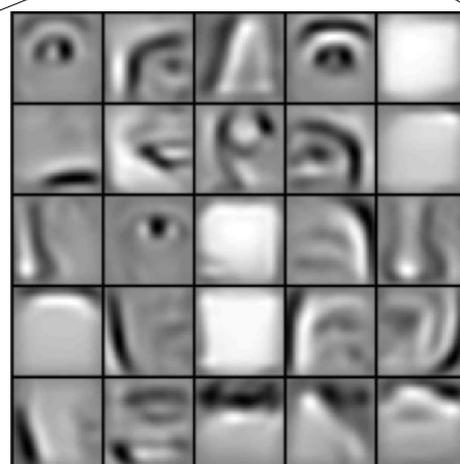
# 深層ネット (deep networks)



- なぜ深層



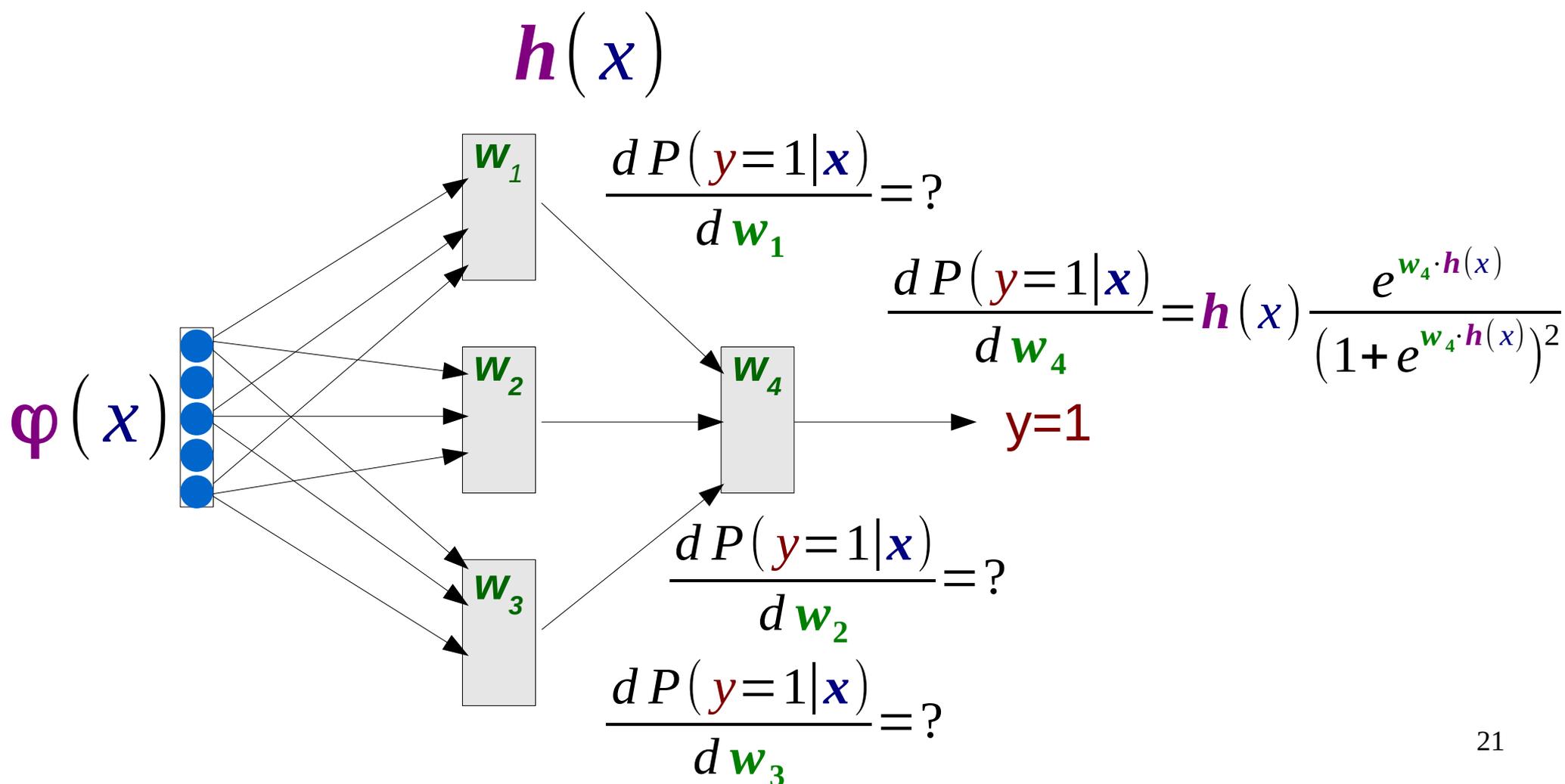
- 徐々に複雑な関数を表現  
(例 [Lee+ 09])



# ニューラルネットの学習

# 学習：隠れ層の勾配が分からない

- 出力層のタグしか分からない



# 逆伝搬 (back propagation)

- 連鎖律を使って計算

$$\frac{dP(y=1|x)}{dw_1} = \frac{dP(y=1|x)}{dw_4 h(x)} \frac{dw_4 h(x)}{dh_1(x)} \frac{dh_1(x)}{dw_1}$$

$$\frac{e^{w_4 \cdot h(x)}}{(1+e^{w_4 \cdot h(x)})^2}$$

↓

次の層の  
エラー ( $\delta_4$ )

$$w_{1,4}$$

↓

重み

$$1$$

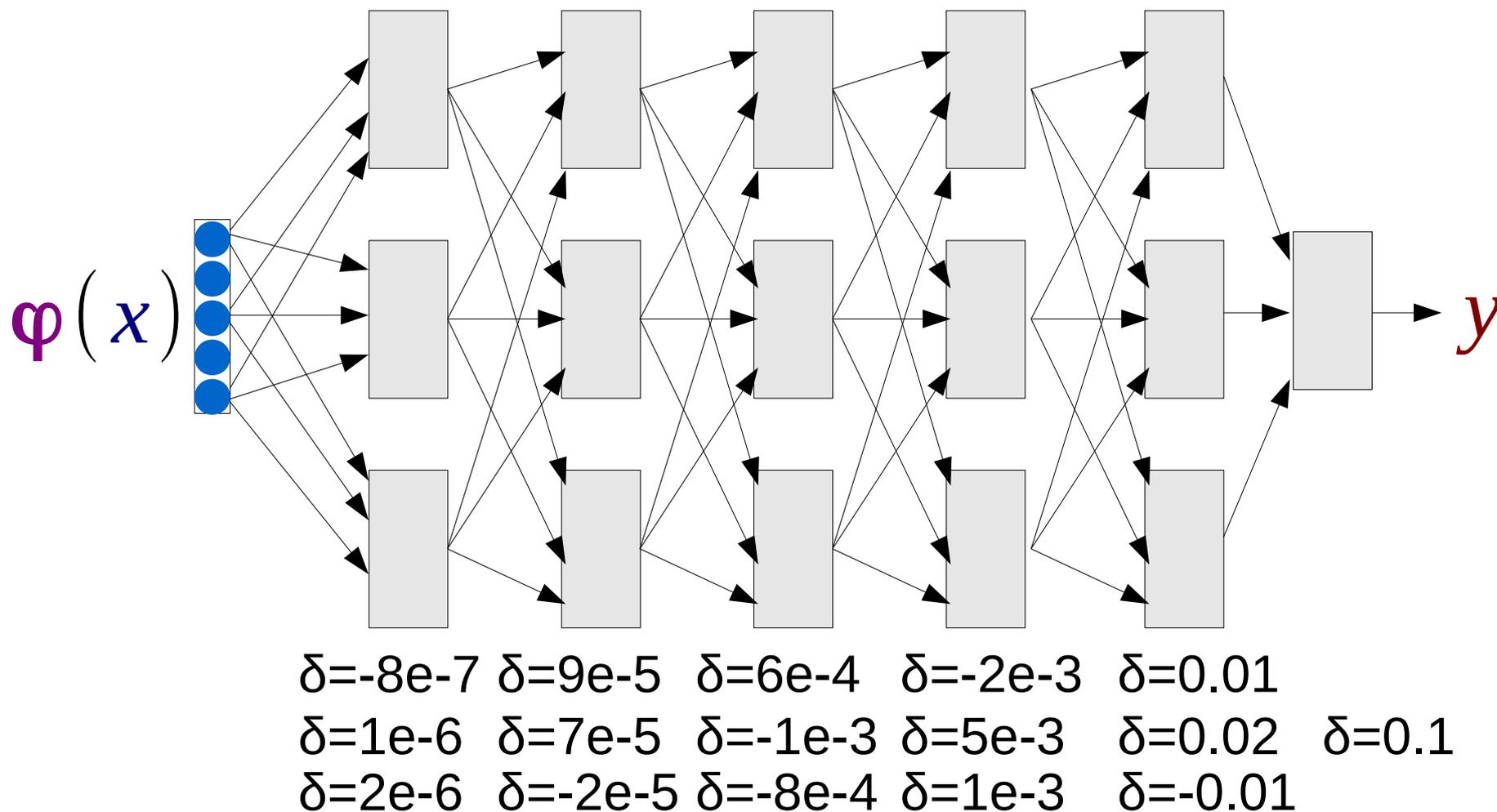
↓

勾配

一般には  
 $i$ を次の層の  
 $j$ のエラーで計算

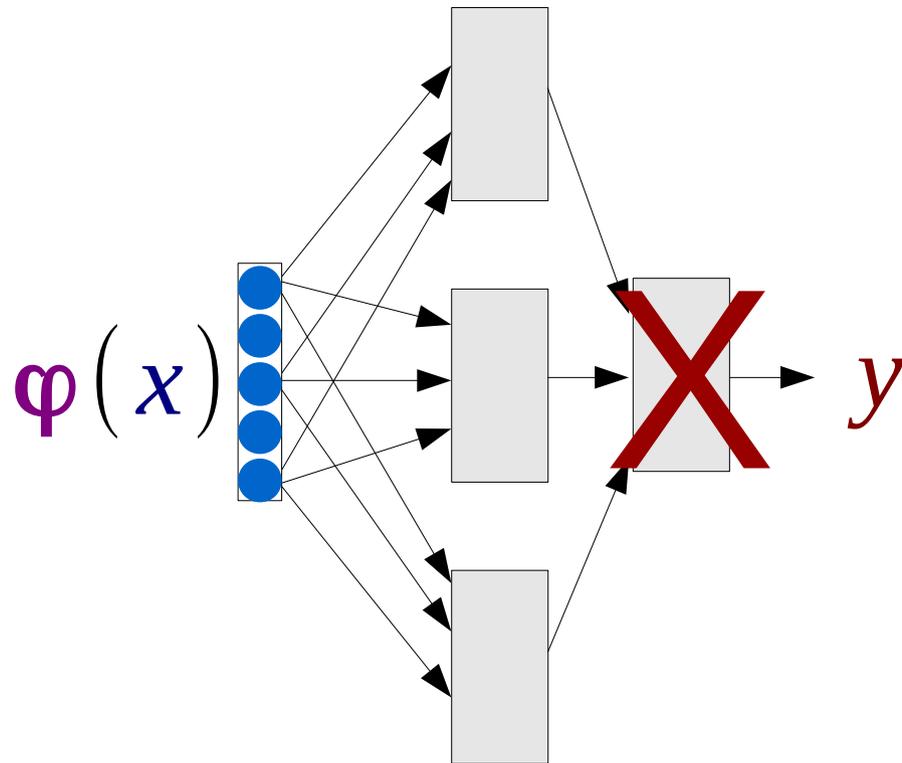
$$\frac{dP(y=1|x)}{dw_i} = \frac{dh_i(x)}{dw_i} \sum_j \delta_j w_{i,j}$$

# 深層ネットの逆伝搬：消える勾配



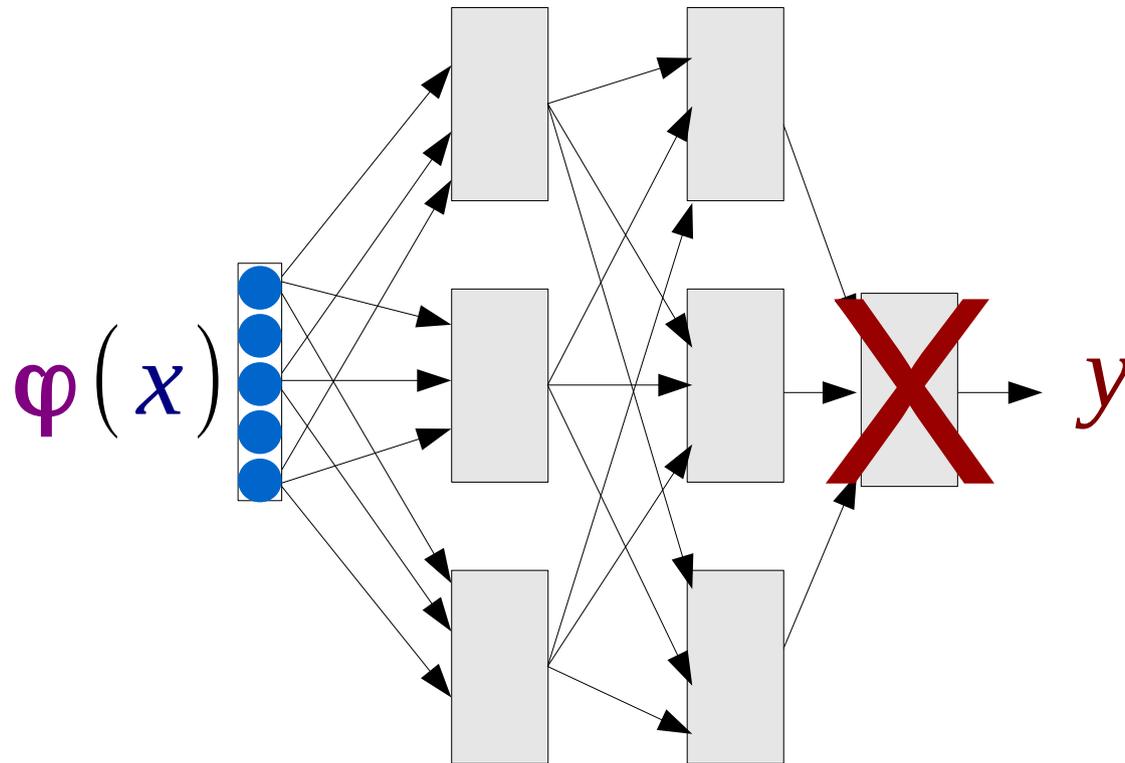
- 爆発的に増える勾配も

# 層ごとの学習 (layerwise training)



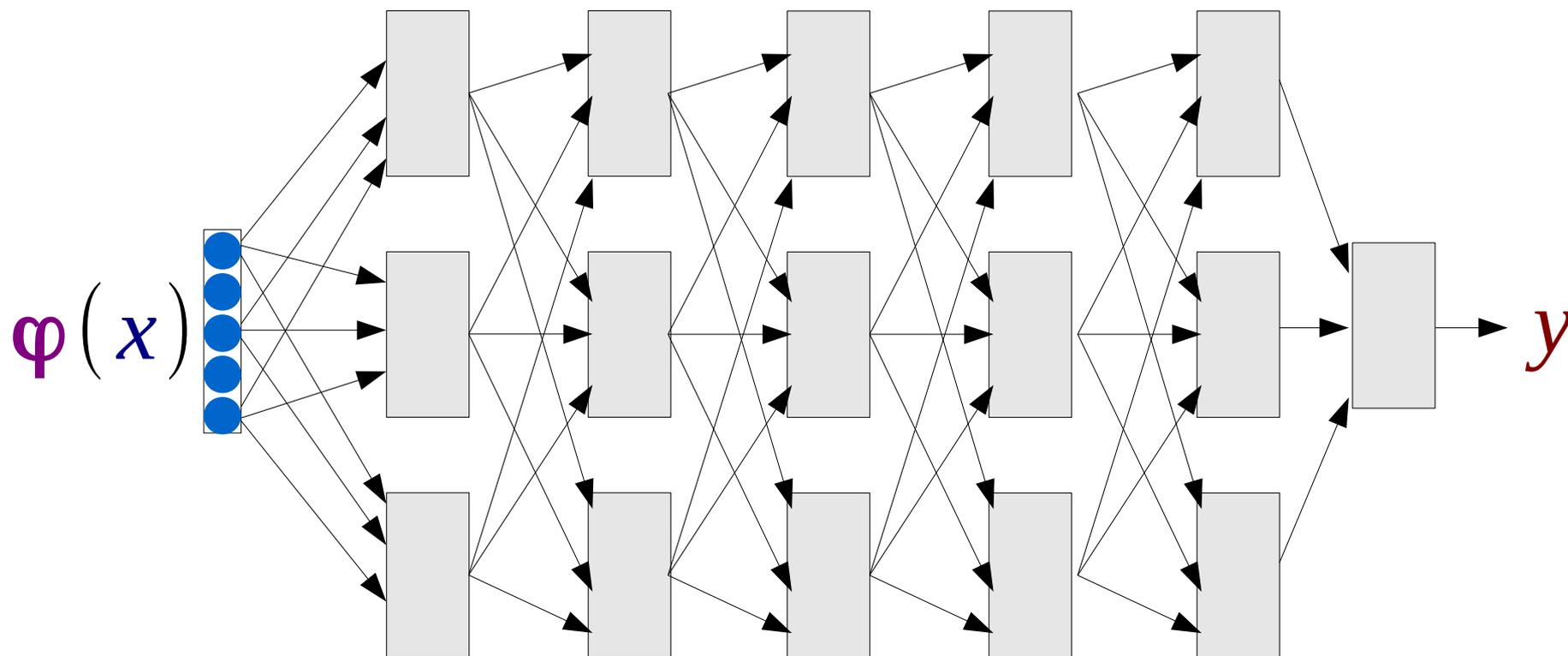
- 層を、入力に近い側から1つずつ学習

# 層ごとの学習 (layerwise training)



- 層を、入りに近い側から1つずつ学習

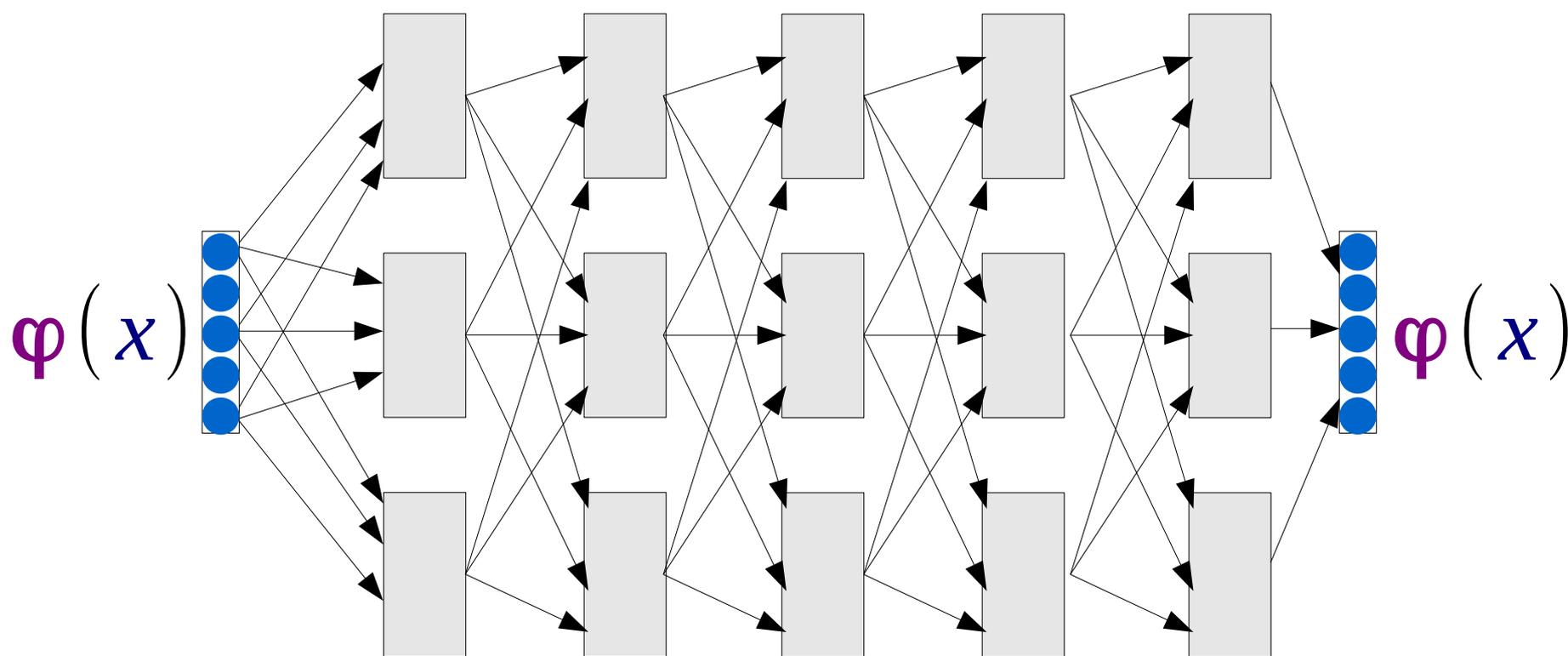
# 層ごとの学習 (layerwise training)



- 層を、入力に近い側から1つずつ学習

# オートエンコーダ（自己符号化器）

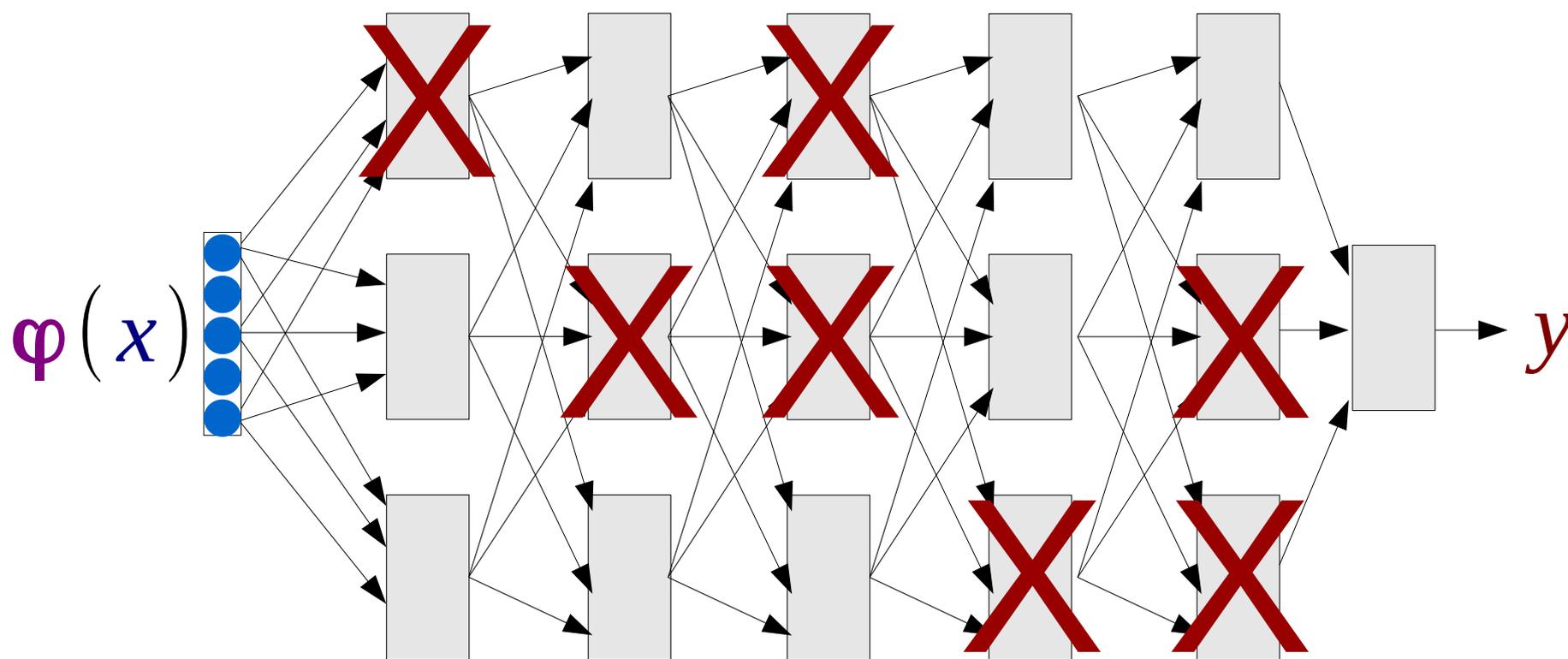
- ネットの初期化として、入力を復元するように学習



- 利点：正解ラベル  $y$  が不要！

# ドロップアウト

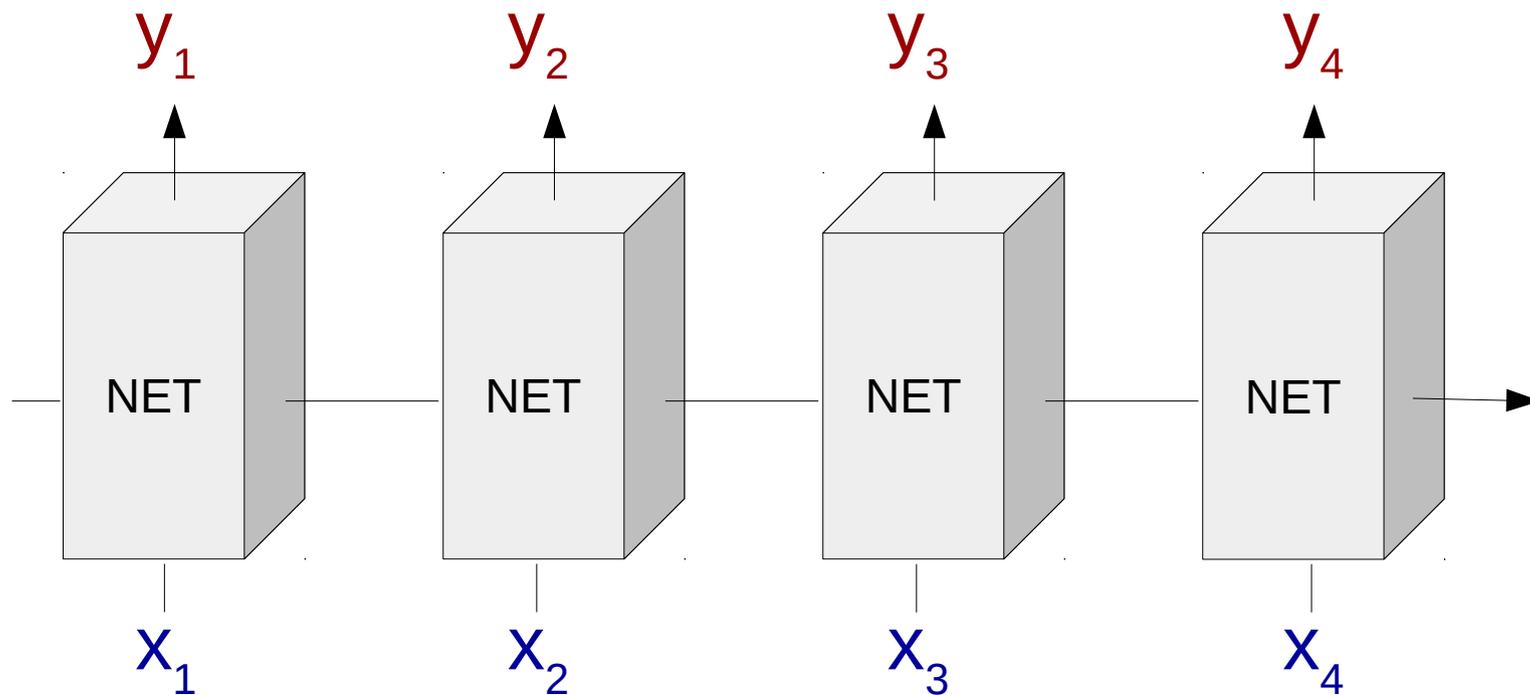
- 問題：過学習



- 各更新ごとに、ネットの一部を無効化
- アンサンブル学習の一種

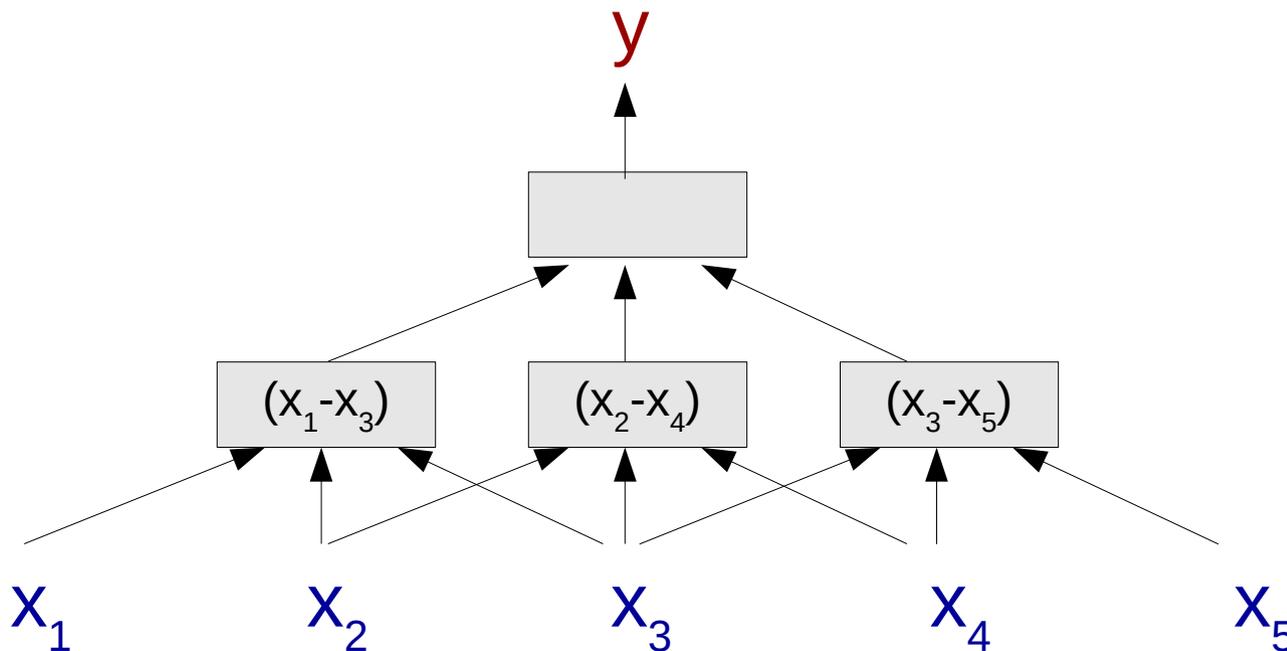
# 特殊な構造のネット

# リカレントニューラルネット (recurrent neural net; RNN)



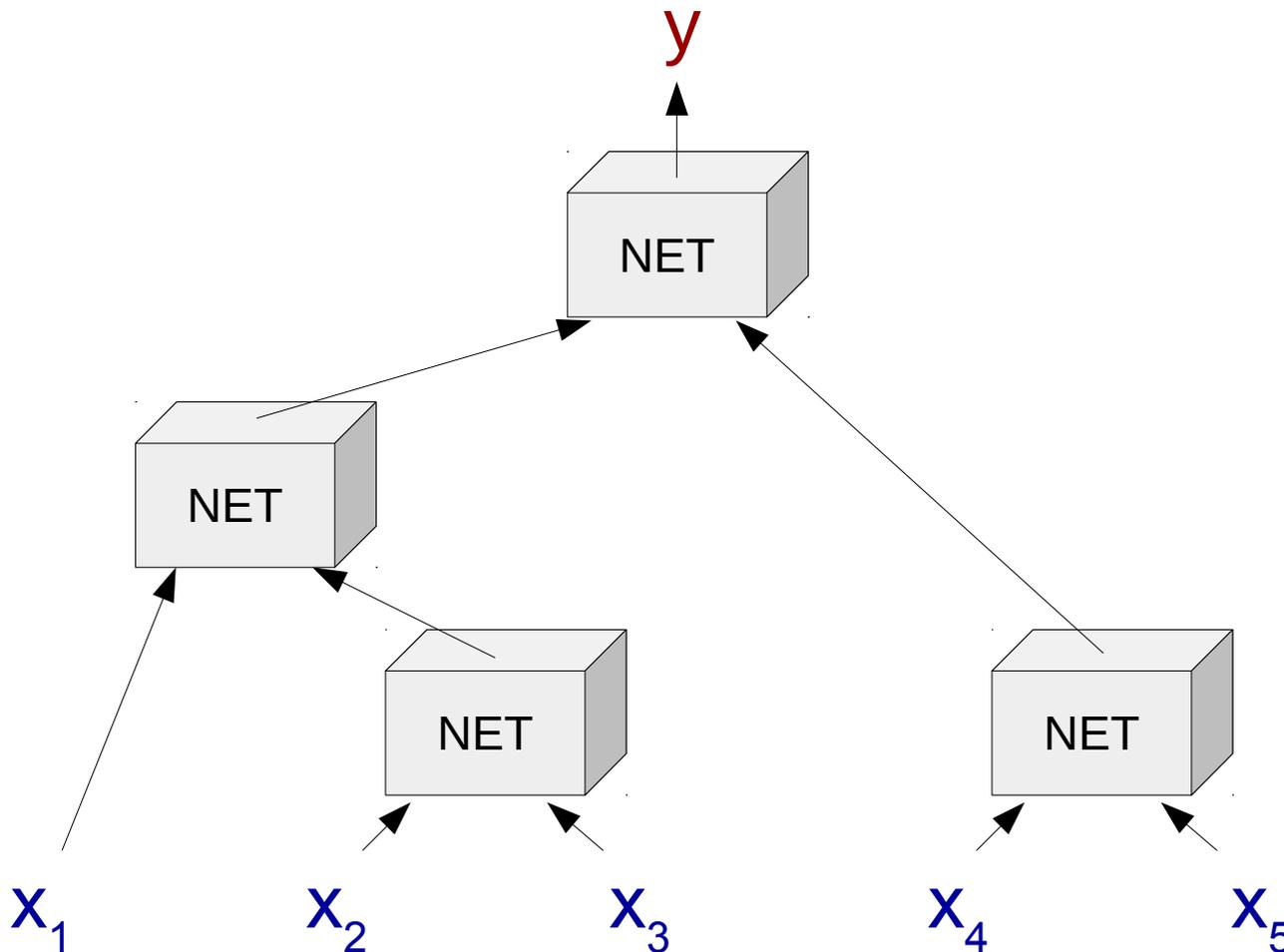
- 系列データのモデル化に有用
- 例：音声の音響モデル

# 畳み込みネット (convolutional neural net; CNN)



- 空間内の「距離」に意味のあるデータをモデル化
- 例：画像

# 再帰的ニューラルネット (recursive neural nets)



- 木構造のデータのモデル化に有用
- 例：言語データ

# その他のトピック

## その他のトピック

- DBN (deep belief net)
  - 制限付きボルツマンマシン
  - コントラストティブ・ダイバージェンス
- ニューラルネットによる生成
- ネットの状態の検証・可視化
- バッチ・ミニバッチ更新
- LSTM (long short-term memory)
- GPU を用いた学習
- 深層学習ライブラリ