

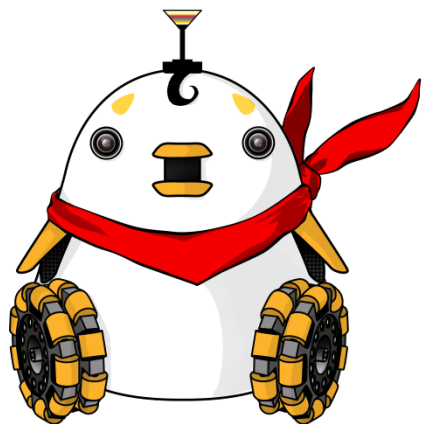
人工知能

第12章 学習と認識(2)

パターン認識と教師あり学習

立命館大学 情報理工学部

谷口彰



STORY パターン認識と教師あり学習

- ホイールダック2号はクラスタリングによって、目で見えた物体をいくつかのクラスタに分けた。これでこれで、**新しい物体を見たときにも、その物体がどのグループに属するかがわかる**だろう。そうすれば、ホイールダック2号は目の前にあるものが何かわかるに違いない。（例えば宝箱なのかゴールなのか）
- しかし、ホイールダック2号は宝箱を5つほど開けたところで気づいた。
- 「どうやら、宝箱には**財宝が入っているものと、罠が入っているものがある**らしい」
- その2種類は**宝箱の見た目が少し違う**ようなのだが、他のゴールや普通の道に比べると、よく似ていたために、クラスタリングの結果としては、同じクラスタになっていた。
- これではホイールダック2号にとっては区別がつかない。しかし、この「**財宝が入っていた**」宝箱の画像と「**罠が入っていた**」宝箱の画像を集めれば、その違いを学習することができるのではないだろうか。

仮定 パターン認識と教師あり学習

- ホイールダック2号は適切な画像特徴量を有限次元ベクトルで取得できるものとする.
- ホイールダック2号は分類のための教師信号を認識できるものとする.

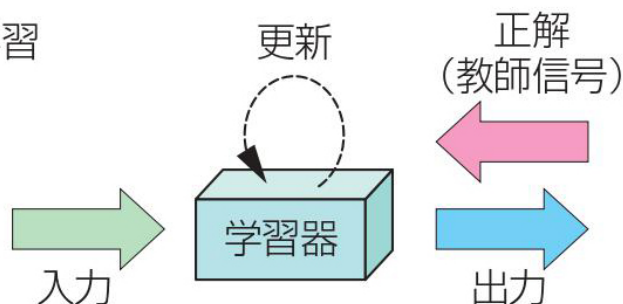


Contents

- 12.1 機械学習とは
- 12.2 機械学習の共通問題
- 12.2 パターン認識
- 12.3 教師あり学習の基礎

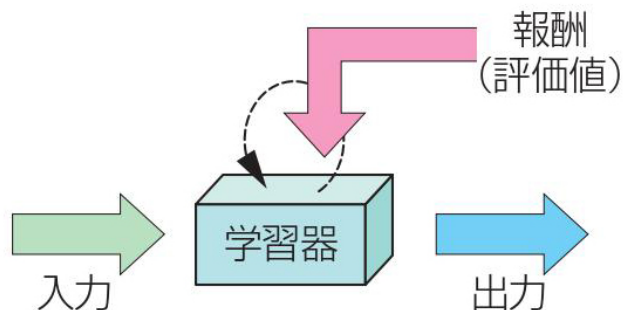
12.1.1 機械学習の分類

(1) 教師あり学習



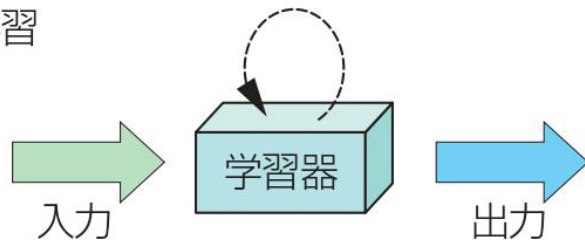
- ✚ パターン認識
- ✚ 時系列データ予測
- ✚ 回帰問題

(2) 強化学習



- ✚ 強化学習
- ✚ 最適化問題
- ✚ 最適制御問題

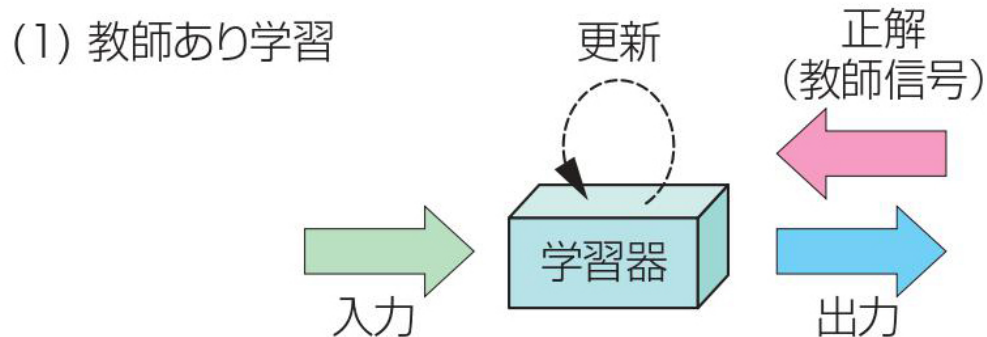
(3) 教師なし学習



- ✚ クラスタリング
- ✚ 低次元化
- ✚ 表現学習
- ✚ データマイニング

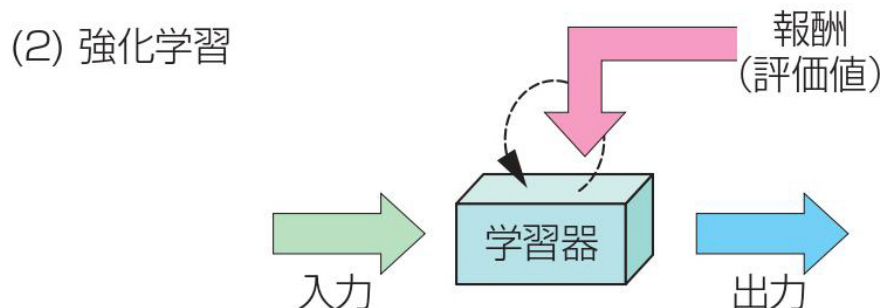
図 12.2 機械学習の分類

12.1.2 教師あり学習



- 学習器は入力に対して正しい出力を返すことを学習する.
- 学習器は訓練データとして正解の出力を与えられるものとする.
- この正解の出力は**目標値(target value)** や**教師信号(supervisory signal)** と呼ばれる. パターン認識の場合はクラスラベルに相当し**ラベル(label)** と呼ぶこともある.
- この情報を用いることにより, 徐々に入力と出力の関係を学習していく学習が**教師あり学習(supervised learning)** である.

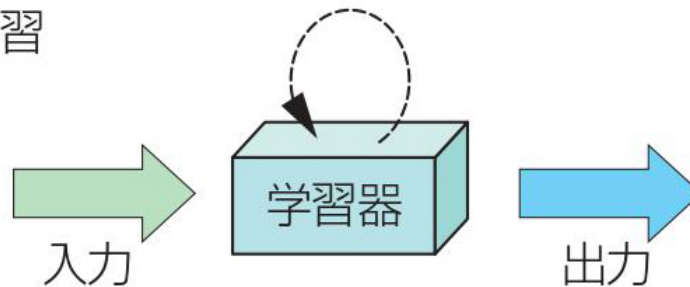
12.1.3 強化学習



- 人間が試行錯誤で学習を行うような場合，必ずしも正しい答え（出力）を教えてもらえるわけではない。
- たとえば，サッカーのフリーキックの練習をするとき，ゴールにうまく入ったかどうかは観測できるが，どう蹴るべきだったか，つまり正解は教えてもらえない。
- このように，行動に対する時間遅れを含んだ事後的な評価に基づいて行う学習を**強化学習(reinforcement learning)**と呼ぶ。この評価値のことを強化学習では**報酬**と呼ぶ。

12.1.4 教師なし学習

(3) 教師なし学習



- 外部からの正解としての教師信号は与えられずに、学習器が入力の情報のみに従って学習を進める学習を**教師なし学習 (unsupervised learning)** と呼ぶ.
- 学習は事前に学習器内部に埋め込まれた規範やアルゴリズムに従って進められる. 多くの場合, 何らかの評価関数を最大化 (最小化) するようにアルゴリズムが構成されていることが多い.
- 教師なし学習の主なものとして**クラスタリング (clustering)** が挙げられる.

演習12-1 機械学習の分類

- 以下の機械学習はそれぞれ「教師あり学習」「教師なし学習」「強化学習」のいずれにあたるか？
 1. 問題を解くと得点だけがしめされて、「後のことは自分で考えなさい！」と言われる試験
 2. 問題を解くとそれぞれの解答が示されて「後のことは自分で考えなさい！」と言われる試験
 3. 100人のマンガのキャラの絵を見せられて「キャラの類似性にもとづいて10グループにわけよ」と言われる課題
 4. カピバラの写真10枚を「これがカピバラだ」と見せられた後に、デグーの写真10枚を「これがデグーだ」と見せられる。その後どちらかの写真を見せられて、それが何かを当てる課題.
 5. 自分一人でペットボトルに入れるビー玉の数を変えては、風呂に投げ入れ、沈むかどうかを判定し、何個入れれば風呂の水に沈むかというルールを学習すること.
 6. 100件のワンルーム不動産の物件に対して、駅からの距離、床面積、風呂トイレの有無、賃料を収集し、駅からの距離、床面積、風呂トイレの有無から賃料を予測出来るようにするタスク.

Contents

- 12.1 機械学習とは
- 12.2 機械学習の共通問題
- 12.2 パターン認識
- 12.3 教師あり学習の基礎

12.2 学習器とパラメータ／汎化

• 結局は関数 f の最適化？

- 学習器は結局は入力から出力への変換を学習する数学的存在としてモデル化される.
- より具体的に言うと, 学習器は何らかの関数 f を持ち, これを関数 f の内部パラメータ θ を変化させることで学習する.
- この θ はニューラルネットワークの結合重みであったり, 強化学習器の Q 値であったりする.

• 訓練データ(training data)と汎化

- 機械学習においては**学習用データ**と**テスト用データ(test data)**を区別することが重要である. 特に教師あり学習では学習用データに対しては教師信号として「答え」が与えられるため, 正しい「答え」を出力できるようになるのは当たり前である.
- 訓練データから訓練データ以外にも適用可能な普遍的知識を見出すことを**汎化(generalization)**という.
- 逆に訓練データに特化した学習をしすぎて, 新しいデータに対応できなくなってしまうことがある. このような現象を**過学習(overtraining)** (もしくは**過剰適合, overfitting**) と呼ぶ. 過学習を避けて高い**汎化性能(generalization performance)**を持つ学習器をつくることが重要となる.

12.2.3 テストデータと交差検証

- データセットを k 分割し $(k-1)$ つのデータを学習に, 1つのデータをテストに用いる. これらを全組み合わせに対して行い評価値を集計する.
- $K=N$ (データ数) の特殊な場合は**LOO(Leave-one-out)法**
- ハイパーパラメータの最適化にはテストデータの他に**検証データ(validation data)**を準備する。

1 回目	Test	Training	Training	Training	Training	Score1	平均 Score
2回目						Score2	
3回目						Score3	
4回目						Score4	
5回目						Score5	

Contents

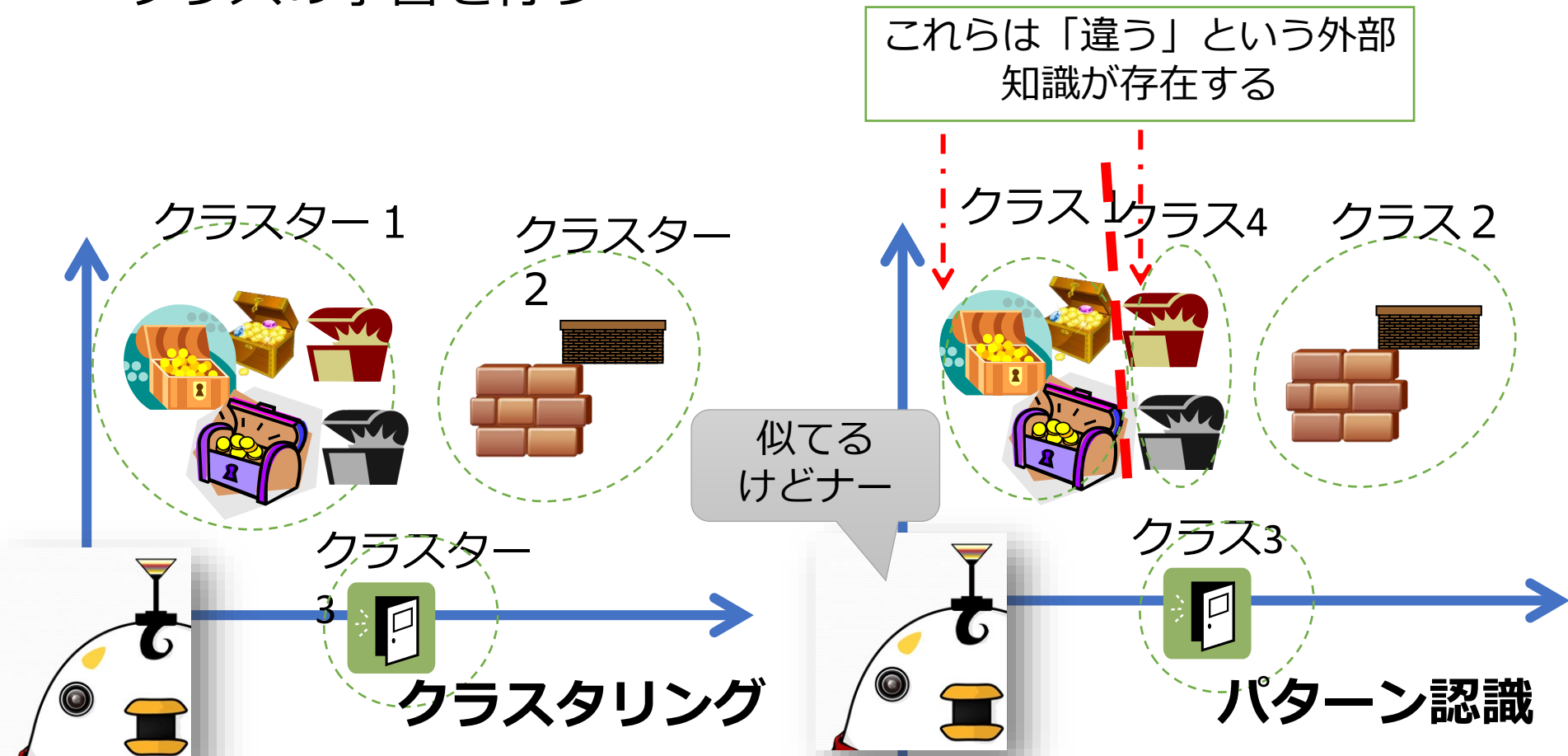
- 12.1 機械学習とは
- 12.2 機械学習の共通問題
- 12.2 パターン認識
- 12.3 教師あり学習の基礎

12.2.1 パターン認識と教師あり学習

- パターン認識とは画像や音声などデータに対して行う情報処理で、観測されたデータを予め定められた複数の**クラス**のうちの一つに対応させる処理である。文字認識(character recognition)
 - 画像データを認識して文字の種類を認識する
 - タッチペン入力の書き文字認識など
- 音声認識(speech recognition)
 - 人間の声を認識して文字列として解釈する。
 - モバイルデバイスでの音声情報検索など
- 画像認識
 - カメラ画像に写った物体が何の物体であるか認識する一般物体認識，表情認識などがある。

教師なし学習によるクラスタリングと 教師あり学習によるパターン認識の違い

□画像の異なり具合を基準にしてクラスの境界を引くよりも、**外部的な知識**に基いて、その違いを見分けるようにクラスの学習を行う



12.3.2 回帰問題と分類問題

□目的

- 入力ベクトル x に対して正しい出力ベクトル y を出力出来るようになること.

□問題の分類

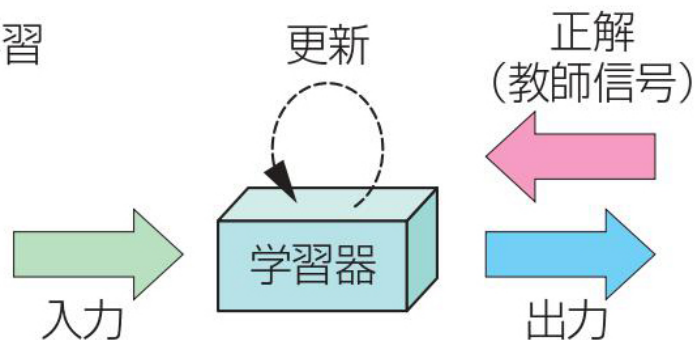
□分類問題 (classification)

- 入力ベクトルに対して正事例であるか負事例であるかの二値 $\{1,0\}$ の値を返すことで分類を行う. (多値のものもあり)
- 学習データとしては正負のラベルの付けられたデータセットを用いる.

□回帰問題 (regression)

- 入力ベクトルに対して通常実数値の値を返し, 未知入力に対する出力の予測を行う.
- 学習データとしては (x,y) の値の組が渡される.

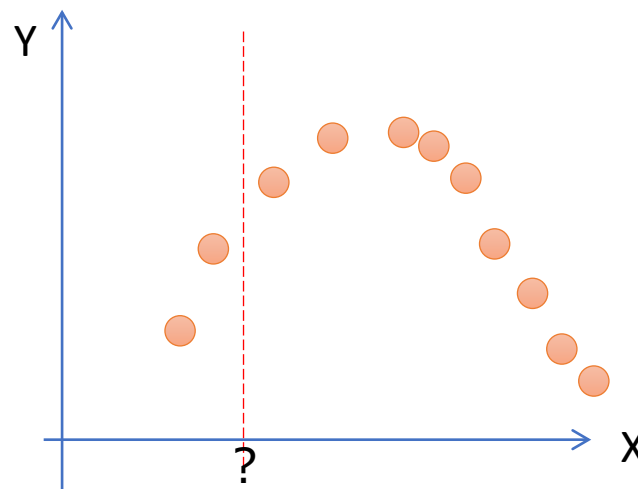
(1) 教師あり学習



12.3.2 回帰問題(regression)

- 入力ベクトルに対して実数値の値を返す連続的な関数関係を学習する問題である。
- 学習後は未知入力に対する出力値の予測を行う。学習データとしては**入力ベクトル x と出力ベクトル（もしくは出力値） y の組み合わせ (x,y) の集合**が学習器に渡される。
- 様々な (x, y) 上の点を与えられた時に未知の入力、たとえば、**？マーク**の位置の入力に対する出力 y を答えるのが回帰問題である。

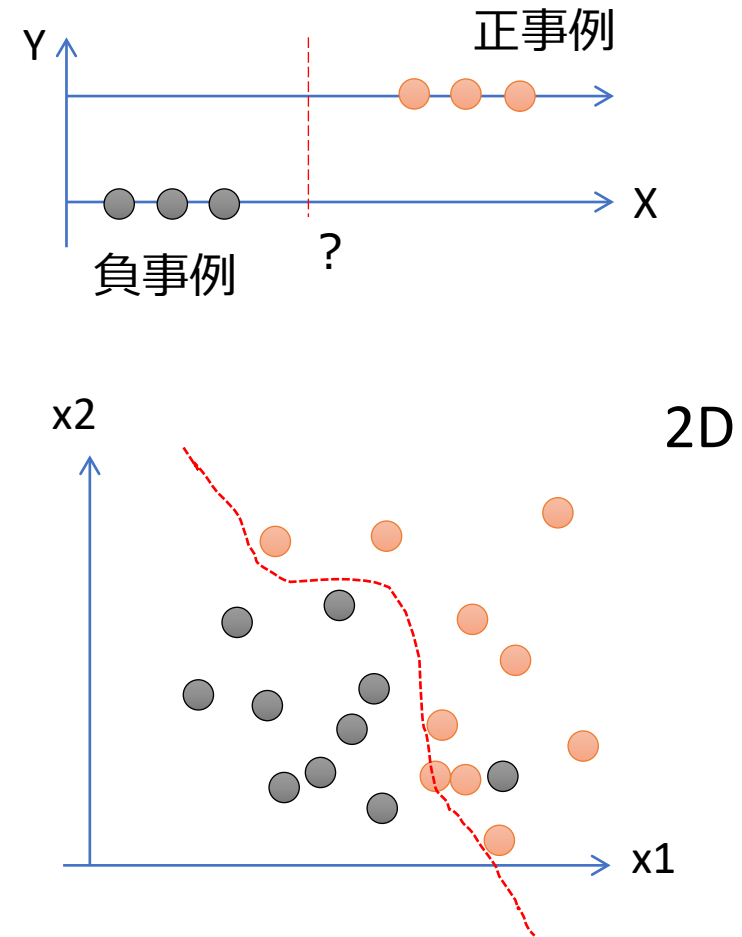
回帰問題



結局は $y=f(x)$ もしくは $P(y|x)$ の推定問題となる場合が多い

12.3.3 分類問題(classification)

- 入力ベクトルに対して正事例(true)か負事例(false)かを返す法則を学習する問題である。
- もしくは、有限個のクラスのどれに属するかを学習する問題。
- 様々な (x, y) 上の点を与えられた時に未知の入力、たとえば、？マークの位置の入力に対する出力 y を答えるのが分類問題である。



結局は $y=f(x)$ もしくは $P(y|x)$ の推定問題となる場合が多い

12.3.4 教師あり学習のさまざまな手法

回帰問題

- 線形回帰
- 一般線形モデル
- ニューラルネットワーク
- カーネル回帰
- ガウス過程回帰(GP)
- その他

分類問題

- パーセプトロン
- ニューラルネットワーク
- SVM(サポートベクターマシン)
- ランダムフォレスト
- 混合ガウス分布
- ナイーブベイズフィルタ
- その他

ニューラルネットワークが近年は主に使われる。
一方に用いられる手法がもう一方に用いられるように拡張されることもよくある。
区別を理解すると同時に共通性を理解するのが大事。

12.3.5 識別モデルと生成モデル

□識別モデル

□正事例と負事例を区別するための境界線を訓練データから直接的に求めようとする。

□生成モデル

□分類対象となるデータがどのような確率モデルから生成されたかをモデル化し、そのモデルに基づいて分類を行う。

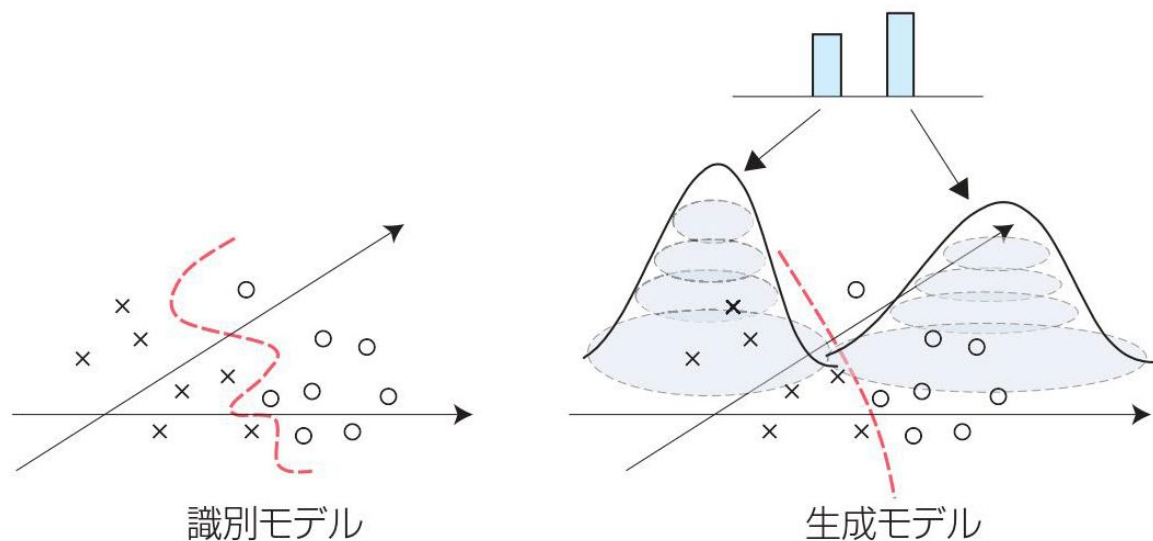
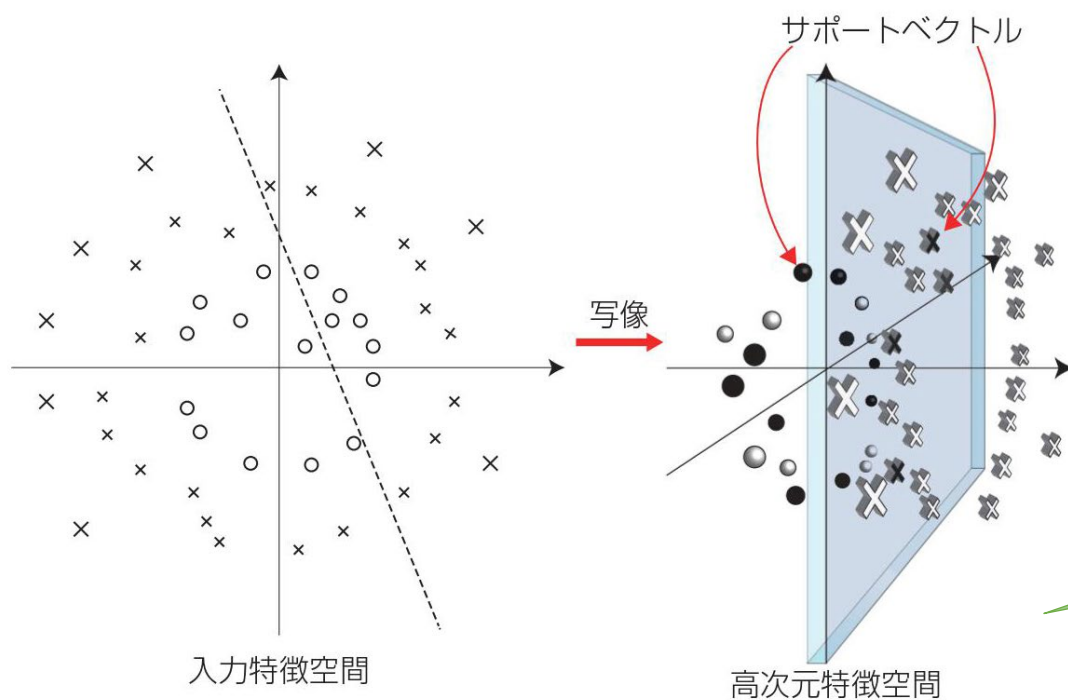


図 12.4

識別モデルと生成モデル

サポートベクトルマシン

□SVM は線形分類器であるパーセプトロンに**カーネル法** (kernel method)を組み合わせることによって実現される



複雑な分離面も表現可能. 汎化性能が大変高い! 便利!

識別モデル

図 12.8

サポートベクトルマシンのイメージ図

ナイーブベイズモデル

ロナイーブベイズモデル(naive Bayes model) は生成モデルに基づき分類を行うために用いられる最も単純なモデルの一つである。

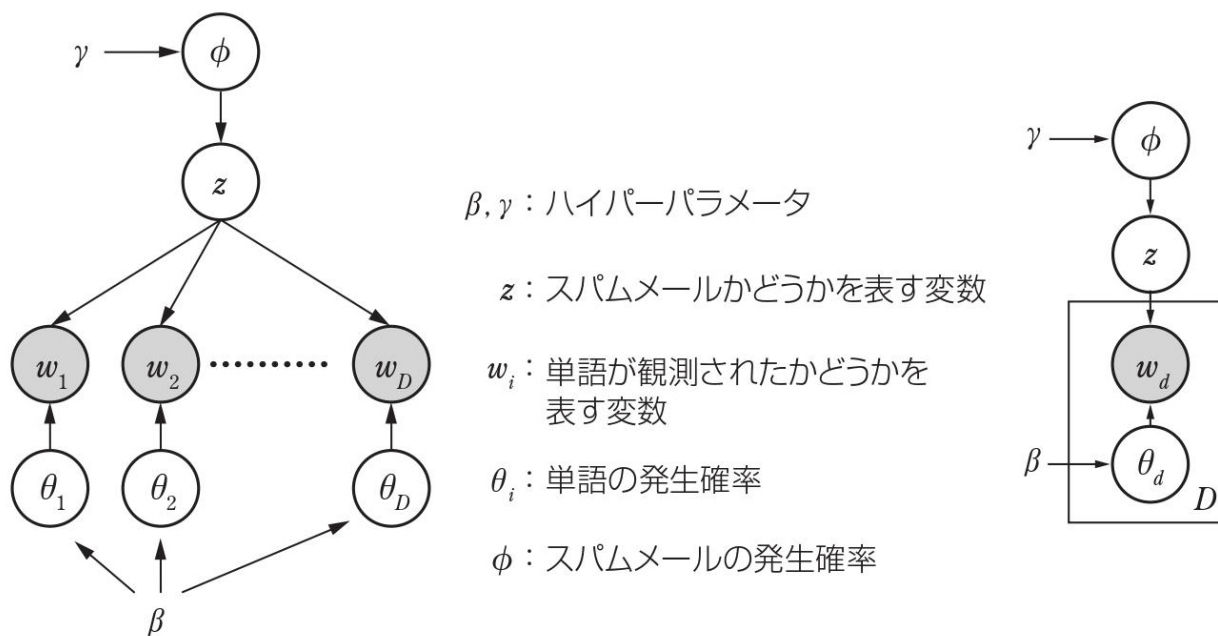


図 7.9 ナイーブベイズモデルのグラフィカルモデル

生成モデル

演習12-2 教師あり学習の分類

- 以下の学習はそれぞれ「分類問題」「回帰問題」のいずれにあたるか？
 1. カピバラの写真10枚を「これがカピバラだ」と見せられた後に、デグーの写真10枚を「これがデグーだ」と見せられる。その後どちらかの写真を見せられて、それが何かを当てる課題。
 2. 自分一人でペットボトルに入れるビー玉の数を変えては、風呂に投げ入れ、沈むかどうかを判定し、何個入れれば風呂の水に沈むかというルールを学習すること。
 3. 100件のワンルーム不動産の物件に対して、駅からの距離、床面積、風呂トイレの有無、賃料を収集し、駅からの距離、床面積、風呂トイレの有無から賃料を予測出来るようにするタスク。
 4. 初速度を V [m/s]として弾丸を射出し、その落下点 x [m]を多数計測することで V - x の関係を学習し一般法則を導き出そうとすること。

Contents

- 12.1 機械学習とは
- 12.2 機械学習の共通問題
- 12.2 パターン認識
- 12.3 教師あり学習の基礎

12.4.1 関数近似と勾配法

- 回帰問題を解くための最も基本的な方法は、入力 x と出力 y の関係が $y = f(x; \theta)$ という関係にあると考え、予測誤差を最小化するように学習器の最適なパラメータ θ^* を求める方法である。

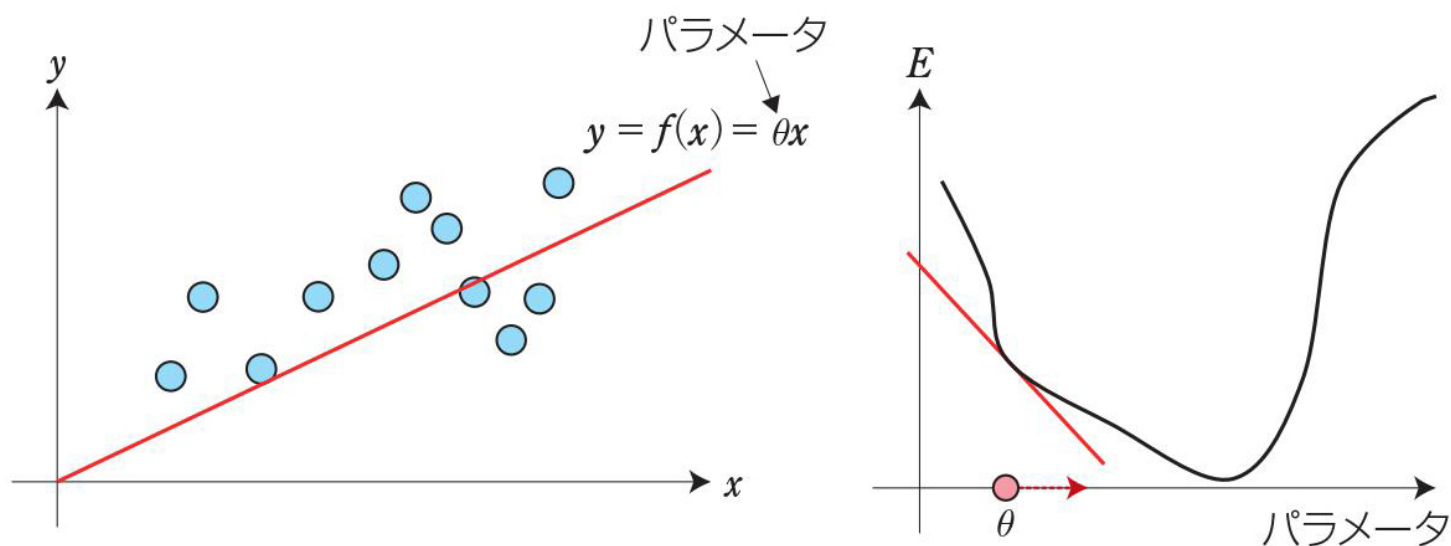
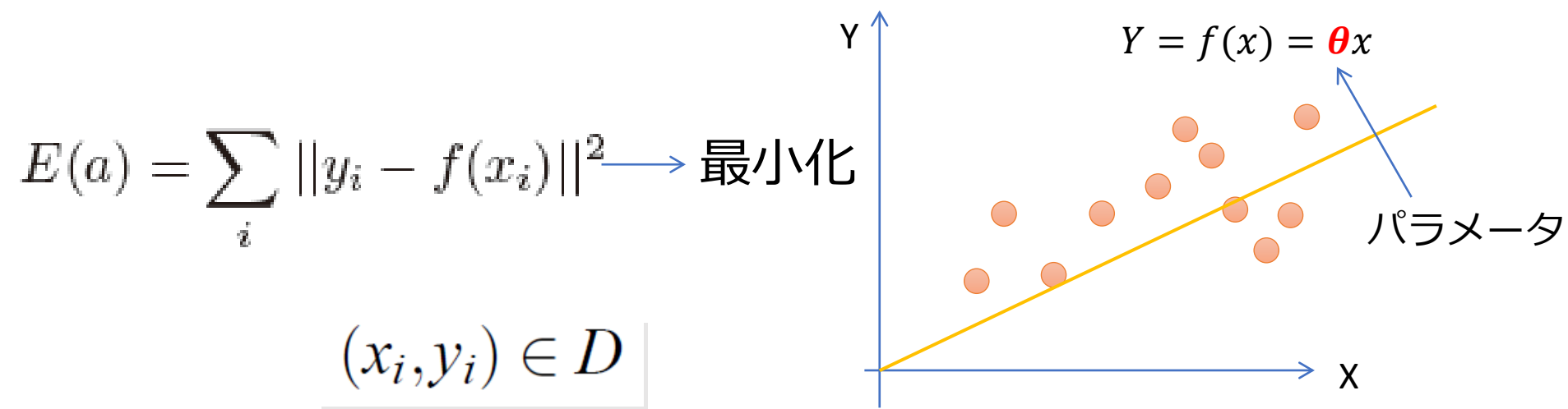


図 12.5 関数近似と勾配法によるパラメータの最適化

最小二乗法

- 与えられたデータに対して近似誤差が最小になるように関数 $f(x)$ のパラメータを調整する.
- **最小二乗法**
 - $f(x)$ が線形関数の場合は解析的に（閉形式で）解ける
- **再急降下法（勾配法）**
 - 誤差が徐々に小さくなるように，誤差の偏微分を計算して逆方向にパラメータを修正



最急降下法（勾配法）

- 誤差 E が徐々に小さくなるように、誤差の偏微分を計算して逆方向にパラメータを修正
- 具体的には**勾配 (gradient)** $\partial E / \partial \theta$ を計算し、その逆方向に更新する.

f のパラメータを θ とすると、 k ステップにおいて

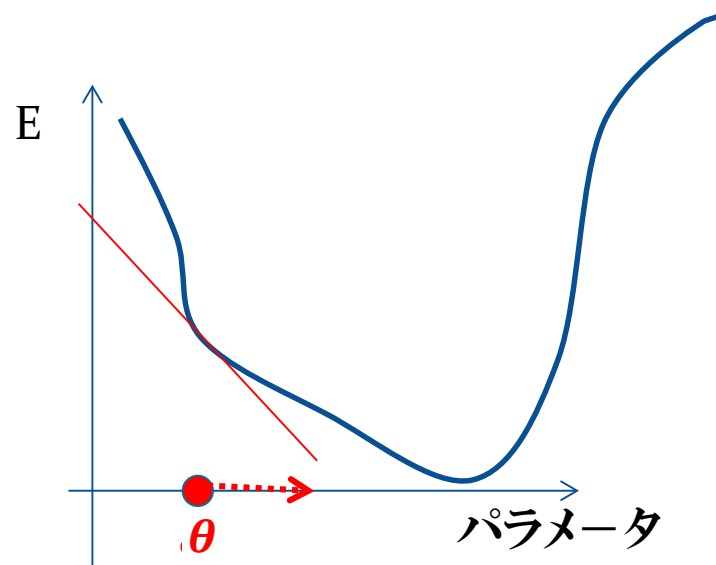
$$\theta^{(k+1)} \leftarrow \theta^{(k)} - \eta \frac{\partial E}{\partial \theta} \Big|_{\theta=\theta^{(k)}} \quad (12.4)$$

として更新していく. η は学習率を表す正の実数であり、 $\theta^{(k)}$ は k ステップ目における θ の推定値である.

パラメータが多変数 $\theta = (\theta_1, \theta_2, \dots, \theta_D)$ の場合、より一般的に勾配を表す記号ナブラ ∇ を用いて最急降下法は以下のように書ける.

$$\theta^{(k+1)} \leftarrow \theta^{(k)} - \eta \nabla_{\theta} E \quad (12.5)$$

ここで、 $\nabla_{\theta} = \left(\frac{\partial}{\partial \theta_1}, \dots, \frac{\partial}{\partial \theta_D} \right)^{\top}$ である.



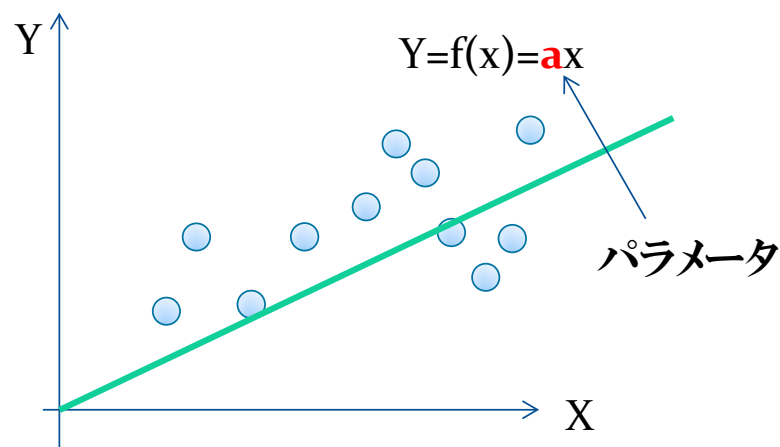
線形回帰：最小二乗法（狭義）

- 最小二乗法

- $f(x)$ が線形関数の場合は解析的に（閉形式で）解ける

$$E = \sum_i ||y_i - f(x_i)||^2 \rightarrow \text{最小化}$$

$(x_i, y_i) \in D$



一変数線形関数については一般形を解いてみよう！

12.4.3 一般線形モデル

- 線形回帰では，線形な関数，つまりグラフにプロットしたときに直線や平面になる関数関係しかモデル化できない。
- 線形回帰の枠組みをそのまま拡張し，非線形関数に対応する簡便な方法として**一般線形モデル**が存在する。

$$f(x) = \sum_m w_m b_m(x) = w^\top b(x) \quad (12.19)$$

ここで $b(x) = (b_1(x), b_2(x), \dots, b_M(x))^\top \in B = \mathbb{R}^M$ は基底関数の出力を並べた列ベクトルであり， w_i は重み， $w = (w_1, w_2, \dots, w_M)^\top$ は重みベクトルである。

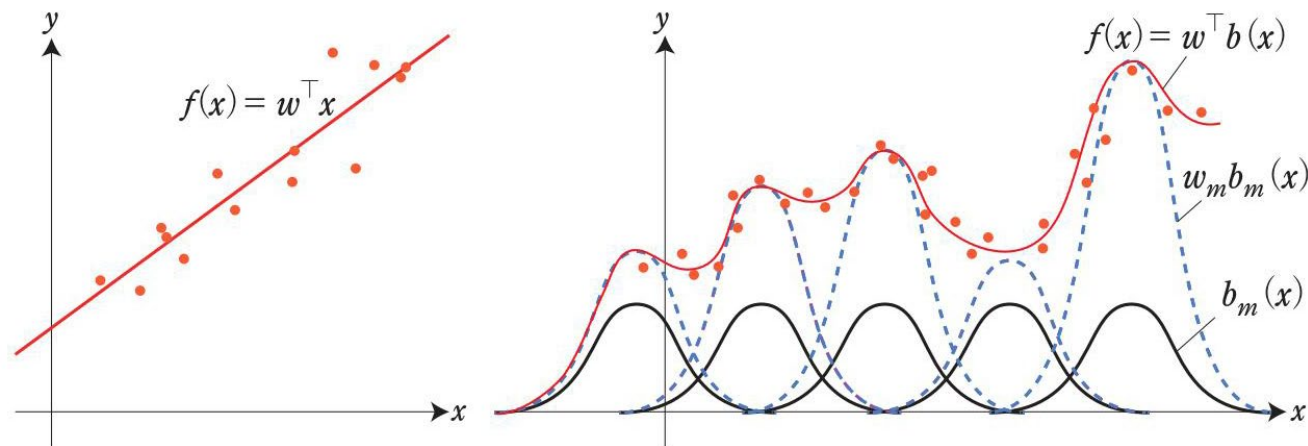


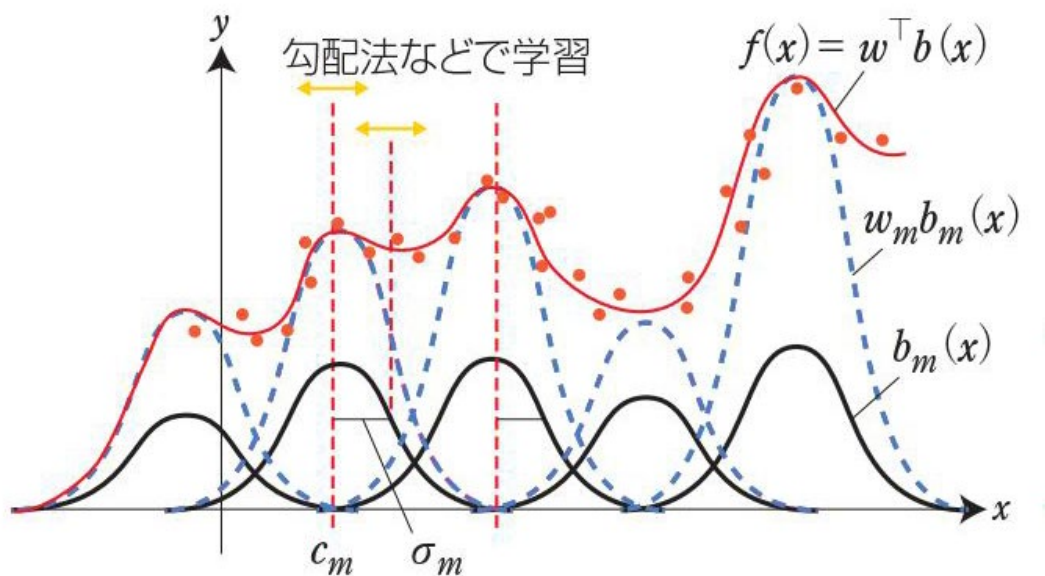
図 12.6

線形モデルと一般線形モデルによる回帰

12.4.4 動径基底関数ネットワーク

- **動径基底関数 (RBF: Radial basis function)** は $b_m(x)$ の値がその中心とそこからの距離のみによって決まる関数であり, その典型がガウス型関数である。
- この基底関数のパラメータ自体も学習する。

その典型がガウス型関数 $b_m(x) = \exp(-(\frac{x-c_m}{\sigma_m})^2)$



ニューラルネットワーク

- 人の脳で行われている情報処理を模倣した情報処理モデルである.
- 回帰問題にも分類問題にも用いることができる.
- BP法（勾配法の効率的計算方法）によって学習する.

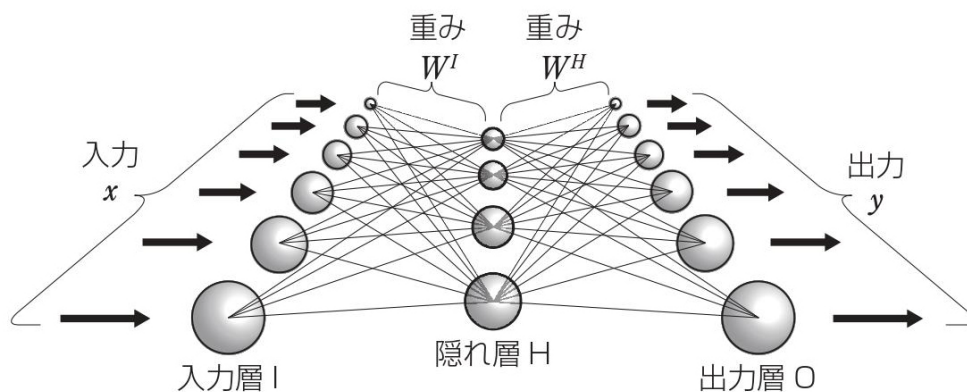


図 13.5 3 層のニューラルネットワーク

12.4.5 分類問題と交差エントロピー誤差関数



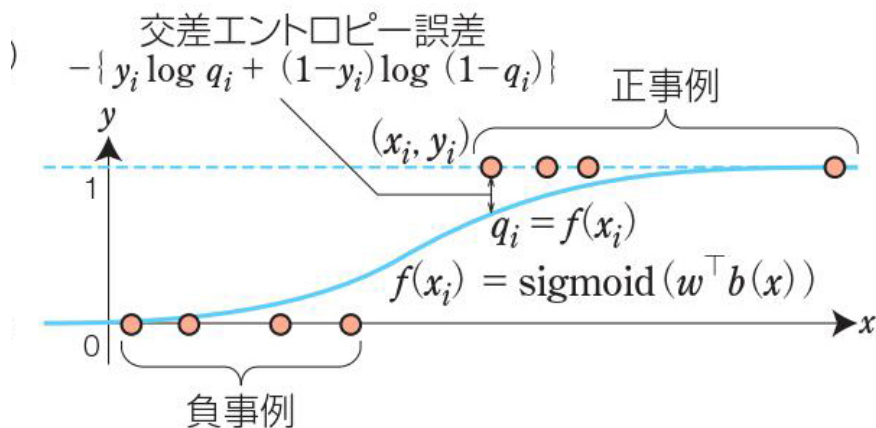
- 尤度関数 $L(\theta) = P(Y|X) = \prod_i q_i^{y_i} (1 - q_i)^{(1-y_i)}$
- 交差エントロピー誤差関数

$$E(\theta) = -\log L = -\log P(Y|X) = -\sum_i y_i \log q_i + (1 - y_i) \log (1 - q_i)$$

ロジスティック回帰

$$f(x) = \text{sigmoid}(w^\top b(x))$$

$$\text{sigmoid}(x) = \frac{1}{1 + \exp(-x)}$$



分類問題も「誤差最小化」
で解くことができます！



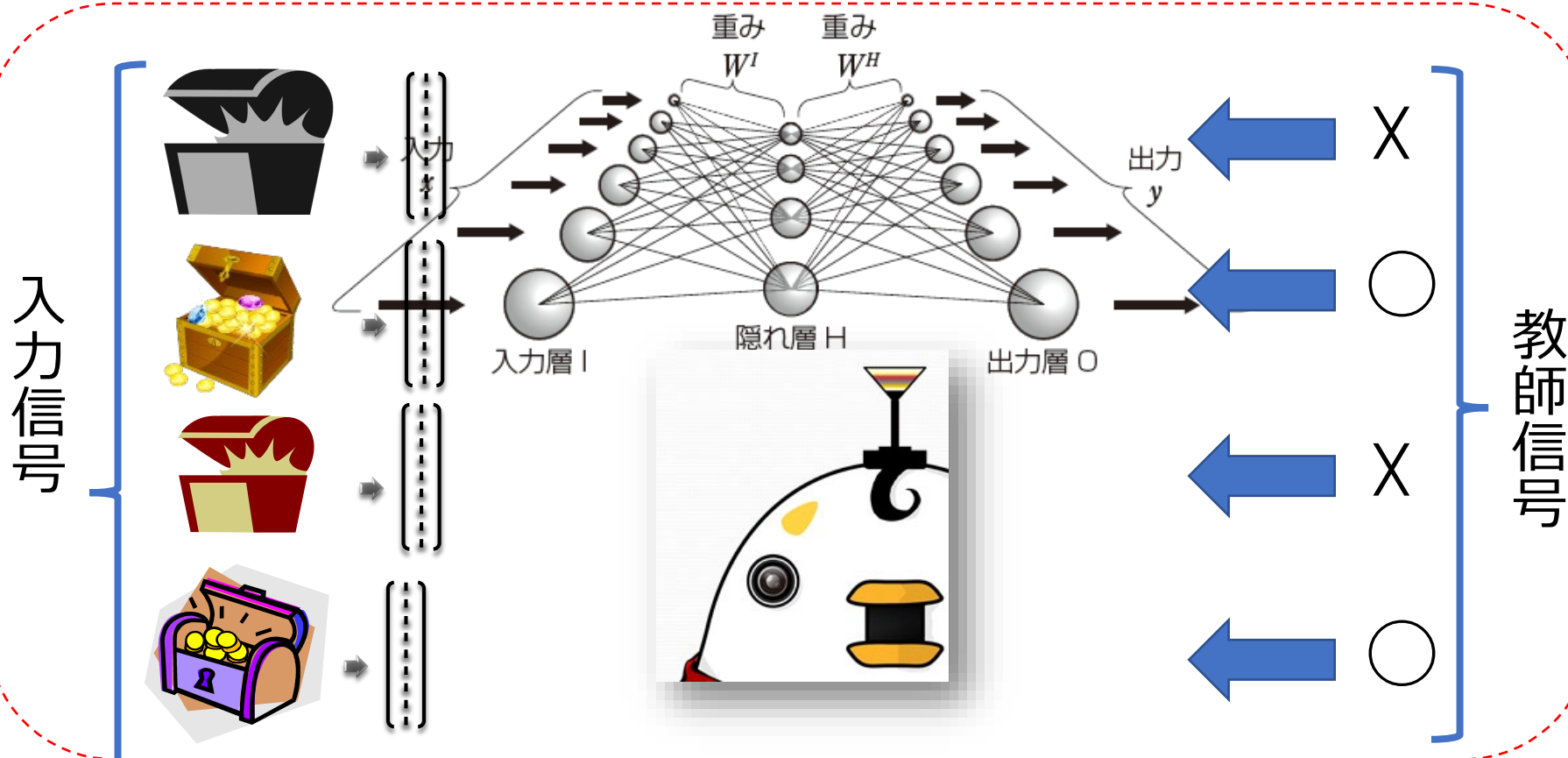
演習12-4 教師あり学習

• 以下の中で最も適切でないものを選べ.

1. 分類問題は交差エントロピー誤差関数の最小化問題として定式化することができる.
2. 一般線形モデルの最小二乗法は線形モデルの最小二乗法と同様に, 解析的に解くことができる.
3. 微分可能な関数でなくてもあらゆる関数の最適化は勾配法で解くことができる.
4. 勾配とはパラメータの各次元に関する偏微分の結果を並べてベクトルにしたものである.

ホイールダック2号の学習

訓練データセット



テストデータセット

○ ?

まとめ

- 機械学習の分類法について学んだ.
- パターン認識とその応用事例について概要を学んだ.
- 回帰問題と分類問題の区別について学んだ.
- 線形回帰および一般線形モデルにおける最小二乗法について学んだ.
- 識別モデルと生成モデルの区別について学んだ.
- 分類問題の学習に用いる交差エントロピー誤差関数について学んだ.

□次回

- 言語と論理（１）

- 自然言語処理