機械学習への導入

二反田篤史

2021年4月12日

認識

物体を見たときにそれが何かわかる機能.

認識

リンゴ

リンゴを見ればリンゴとわかるような機能のことで，周辺環境が変化したりはじめて見るリンゴでも正しく動作することが期待される.

コンピュータによる認識

認識機能をコンピュータ上で実現したい．

ルールベースのアプローチでは人が事前に記述したルールに従い認識を試みる．

リンゴ認識のルール

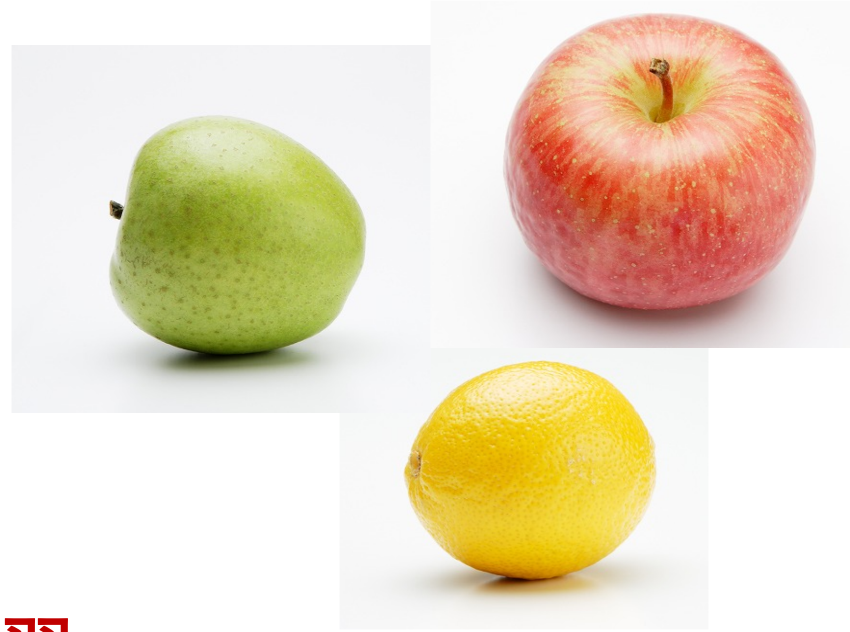
- 赤あるいは緑
- 丸い
- くぼみがある
- 果軸がある

しかし光源，周辺環境，形，色などの変化は非常に多彩．

複雑なルールを事前にプログラムすることは実質的に不可能．

学習

人が持つ認識機能は過去の経験からの学習により形成される.

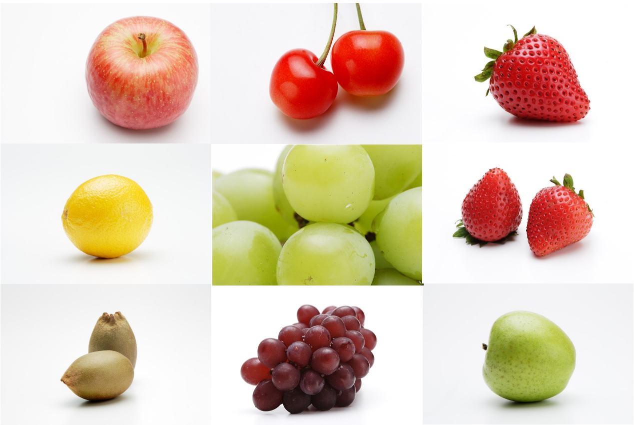


経験
リンゴと呼ばれる物体,
それ以外の物体の複数回の観測.

学習
限られた経験からリンゴ間で不偏的な視覚パターンを認知,
物体をリンゴを認識するルールを獲得.

機械学習

ルールを事前にプログラムするのではなく，有限の観測データからデータに潜むルールを発見する学習方法をプログラムする．



観測データ

機械学習

g:  $\mapsto$ リンゴ

仮説関数

機械学習とは観測データからルールの仮説を出力する過程を実現する技術．

機械学習を搭載したシステム

- **画像認識**
  顔認証，自動タグ付け，自動運転
- **音声認識**
  動画の字幕生成，スマートスピーカー
- **アイテム推薦**
  視聴履歴から動画推薦，購入履歴から商品推薦
- **自然言語処理**
  機械翻訳，チャットボット
- **強化学習**
  ゲームAI，ロボット制御

人工知能・機械学習・深層学習

人工知能（Artificial Intelligence, AI）

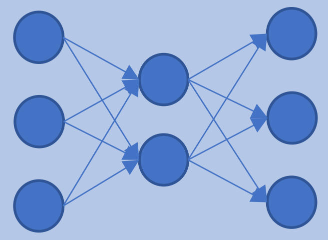
人間のような知能の獲得.
人間の作業の代行.

機械学習（Machine Learning, ML）

有限の観測データから背後に潜む規則を獲得.
教師あり学習・教師なし学習・強化学習という枠組みがある.

深層学習（Deep Learning, DL）

ニューラルネットワークを用いた機械学習.
AIにパラダイムシフトを起こした.



ニューラルネットワーク：
人の脳を模した学習機械.

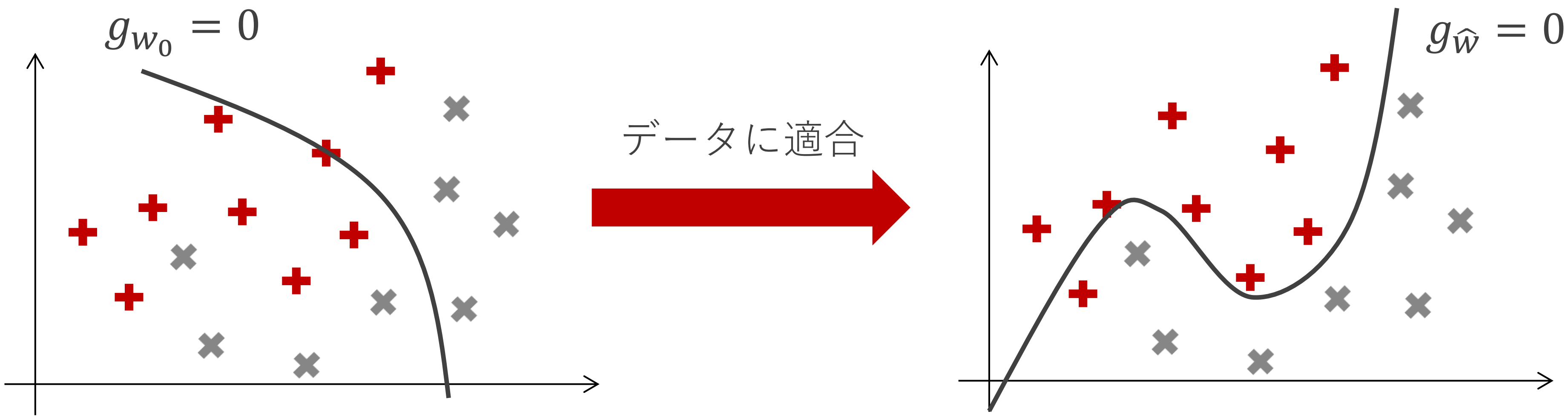
教師あり学習

機械学習の基本的な枠組みの一つ.

n個の入出力ペア $D = \{(x_i, y_i)\}_{i=1}^{n} \subset \mathcal{X} \times \mathcal{Y}$を観測する問題設定.

$x_i \in \mathcal{X} \subset \mathbb{R}^d$: ベクトル化された入力データ, $y_i \in \mathcal{Y} \subset \mathbb{R}$: 実数値化された出力データ.
例 (リンゴの認識): x_iは画像のRGB値, y_iはリンゴなら1, それ以外なら-1.

教師あり学習ではDから入出力の対応関係の仮説関数$g: \mathcal{X} \to \mathcal{Y}$を獲得する.



パラメータを持つ仮説関数 g_wをデータに適合することで実現.

期待損失最小化問題

$\mathcal{F}$: 仮説関数の集合 (多項式関数，ニューラルネットワークなど)

$l(g(x), y)$: 損失関数．仮説関数gとデータ(x, y)の誤差を測る．

例(ロジスティック損失): 2値の判別問題で使用 ($y \in \{-1,1\}$),

$$l(g(x), y) = \log(1 + \exp(-yg(x))).$$

期待損失最小化問題　教師あり学習で解きたい問題

$$\min_{g \in \mathcal{F}} \{\underbrace{\mathcal{L}(g) = \mathbb{E}_{(X,Y)}[l(g(X), Y)]}_{\text{期待損失関数}}\}.$$

(X, Y)は入力データとラベルを表す確率変数でデータ分布に従う．

期待損失$\mathcal{L}(g)$は仮説関数gで予測した場合に被る損失の期待値．

従って期待損失最小化により$\mathcal{F}$の中では平均的には最もデータに適合する仮説が得られる．

機械学習のアルゴリズム

期待値$\mathbb{E}_{(X,Y)}$は一般に厳密には取り扱えない.

そこで，教師あり学習では有限個の観測データを手掛かりに期待損失を近似的に最小化する仮説関数の獲得を試みる.

すなわち機械学習におけるアルゴリズムとは次の写像$\mathcal{A}$のこと：

$$\mathcal{A}: \mathcal{X}\times\mathcal{Y} \supset \{(x_i, y_i)\}_{i=1,2,\ldots,n} \mapsto g \in \mathcal{F}.$$

アルゴリズムの例：

- **経験損失最小化法**　訓練データが定める経験損失の最小解を獲得.
- **確率的最適化法**　期待損失の最適化アルゴリズム.

※ 確率的最適化法は経験損失最小化問題にも適用可能.

経験損失最小化法

経験損失とは観測データ$D = \{(x_i, y_i)\}_{i=1}^n$による期待損失の近似.
経験損失を最小化することで間接的に期待損失の最小化を試みる.

経験損失最小化問題

$$\min_{g \in \mathcal{F}} \left\{ \mathcal{L}_n(g) = \frac{1}{n} \sum_{i=1}^{n} l(g(x_i), y_i) \right\}.$$

仮説関数gごとには$\mathcal{L}_n(g)$は高確率で$\mathcal{L}(g)$の近似値なので一見うまくいくように見える.
しかし，観測データから得られた仮説関数に対し近似がうまくできているとは限らない.
→ 過学習という問題が生じ得る.

過学習

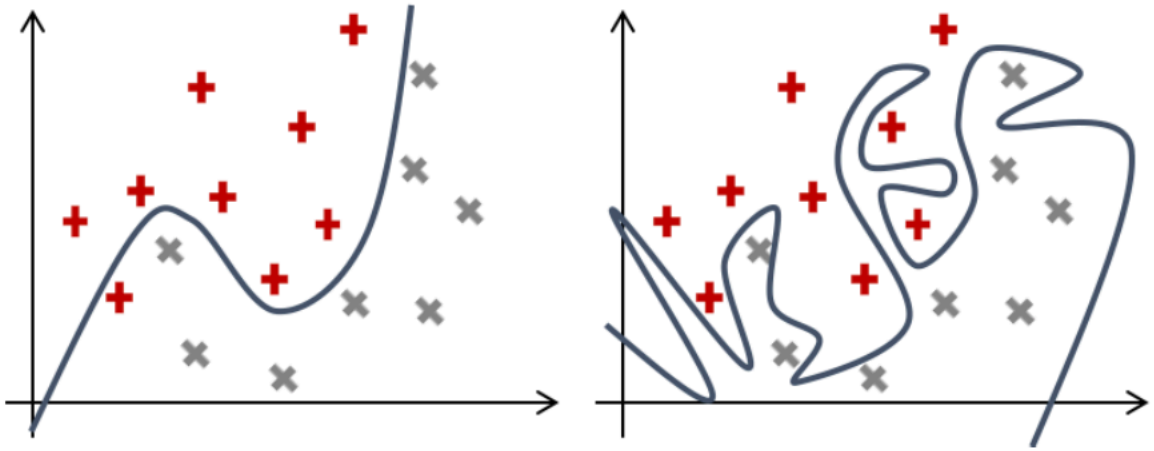
過学習　経験損失と期待損失のずれに起因し経験損失は小さくとも期待損失が大きくなる．

$$\left|\mathbb{E}_{(X,Y)}[l(g_{\hat{w}}(X),Y)] - \frac{1}{n}\sum_{i=1}^{n} l(g_{\hat{w}}(x_i),y_i)\right|.$$

（$\hat{w}$は有限の観測データにから得られたパラメータ）

モデルが複雑な仮説関数を含んでいると
過学習がおきやすくなる．
例えば経験損失は0でも期待損失が大いことは
あり得る．

過学習を防ぐ技法として正則化などがある．



適切な分類(左図)と過学習(右図)．

機械学習の理論

- 学習理論とは機械学習アルゴリズムの性能を評価する理論.
  すなわち獲得された仮説関数の期待損失を見積もる理論.
  機械学習アルゴリズム設計の指針を与える.

- 大規模機械学習問題に対する最適化手法の研究も重要な研究分野.
  大量の観測データに適合する仮説ほど高い汎化性能を持つ.
  確率的最適化などデータサイズにスケーラブルな最適化手法の重要性.

- 最適化手法の性質と汎化性能の間に潜む関係性の解明.

内容を数学的に正しく理解することを目指す.