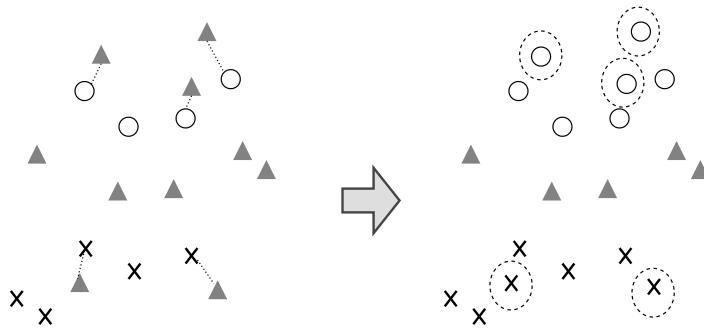


14. 少量データ学習



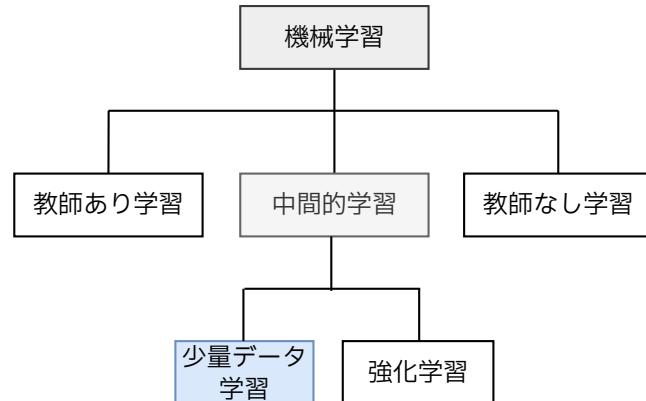
- 14.1 少量データ学習の位置付け
- 14.2 半教師あり学習
- 14.3 転移学習
- 14.4 自己教師あり学習



- 荒木雅弘：『Pythonではじめる機械学習』（森北出版, 2025年）
- スライドとコード

14.1 少量データ学習の位置付け (1/2)

- 問題の位置付け
 - 正解付きデータを使って損失関数の最小化を行う教師あり学習と、正解なしデータを使ってデータの確率分布を推定する教師なし学習の中間
 - 何らかの方法で（近似的に）得られたデータの確率分布を活用しながら、少量の正解付きデータを使ってモデルを構築する



14.1 少量データ学習の位置付け (2/2)

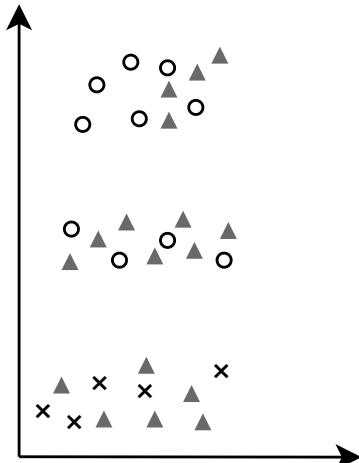
- 少量データ学習の問題設定
 - 対象タスクの正解なしデータは大量にあるが、同タスクの正解付きデータが少ない
 - 正解なしデータの確率分布を推定する半教師あり学習が有効
 - 類似タスクの正解付きデータは大量にあるが、対象タスクの正解付きデータが少ない
 - 類似タスクで学習したモデルを対象タスクに適用する転移学習が有効
 - 入力と同型式のデータは大量にあるが、対象タスクの正解付きデータが少ない
 - 入力データの確率分布を学習する自己教師あり学習と、ファインチューニングの組み合わせが有効

14.2 半教師あり学習

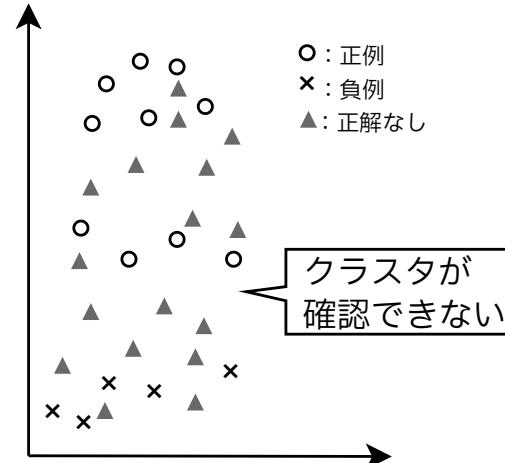
- 半教師あり学習の状況設定
 - 正解情報が一部の学習データにのみ与えられている
 - 例：tweet のP/N判定, 画像分類, 行動認識
 - データ自体はクローラ等で容易に収集できる
 - タグ付け作業にコストが掛かり, 大量の正解付きデータが得にくい

14.2.1 半教師あり学習に適したデータ (1/2)

- 数値特徴の場合
 - 正解なしデータから得られる $p(\mathbf{x})$ に関する情報が $p(y | \mathbf{x})$ の推定に役立つこと



(a) 半教師あり学習に適したデータ

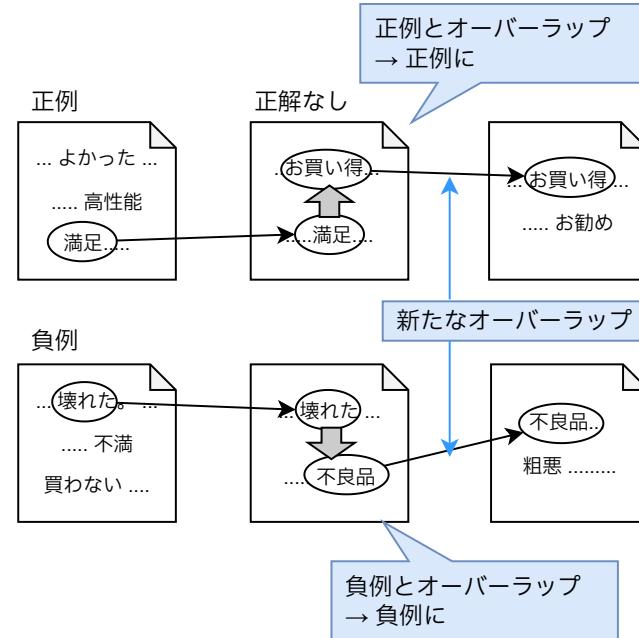


(b) 半教師あり学習に適さないデータ

- なめらかさ, 低密度分離, 多様体仮定などの仮定が満たされる場合に有効

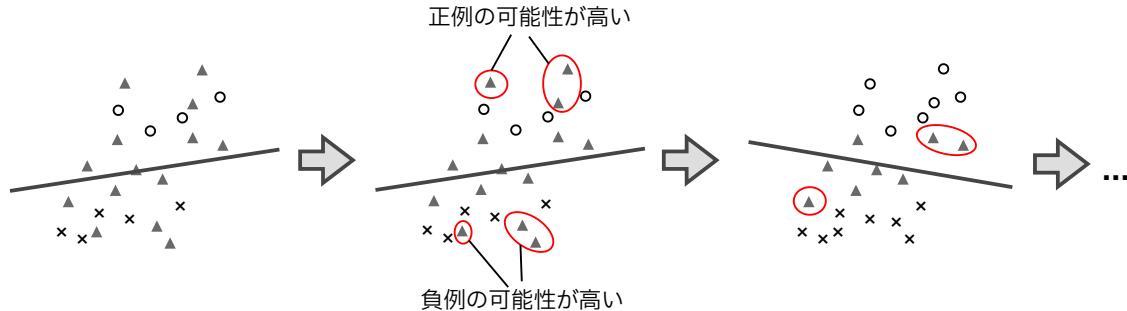
14.2.1 半教師あり学習に適したデータ (2/2)

- カテゴリ特徴の場合
 - 識別に役立つ特徴値の伝播が生じる場合
 - 例：文書のP/N判定



14.2.2 疑似ラベルの活用 (1/3)

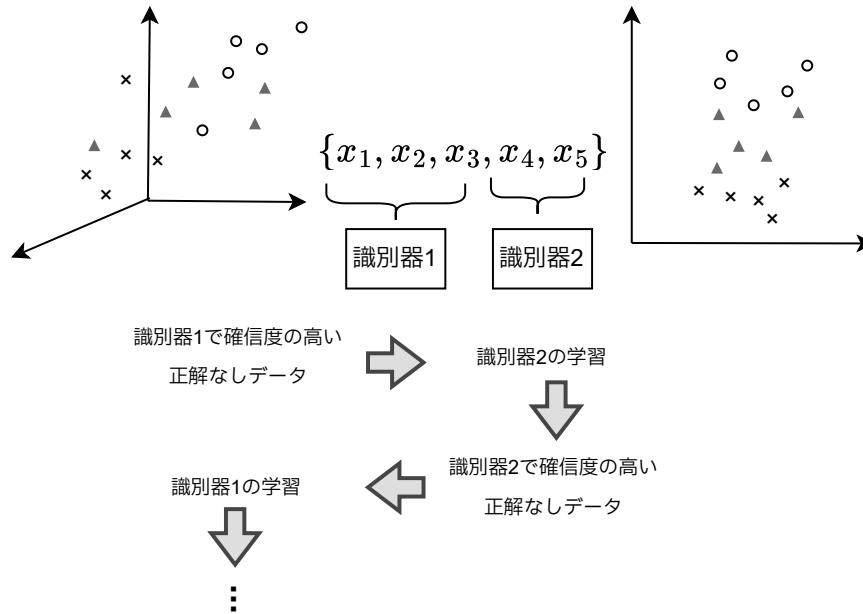
- 疑似ラベルとは
 - 正解なしデータに対して、現在の識別器で識別したラベル
- 自己学習
 1. 正解付きデータで初期識別器を作成し、正解なしデータを識別して確信度の高いものを正解付きデータとみなす
 2. 新しい正解付きデータを加えた識別器の再学習を、事前に決めた回数または正解なしデータがなくなるまで繰り返す



14.2.2 疑似ラベルの活用 (2/3)

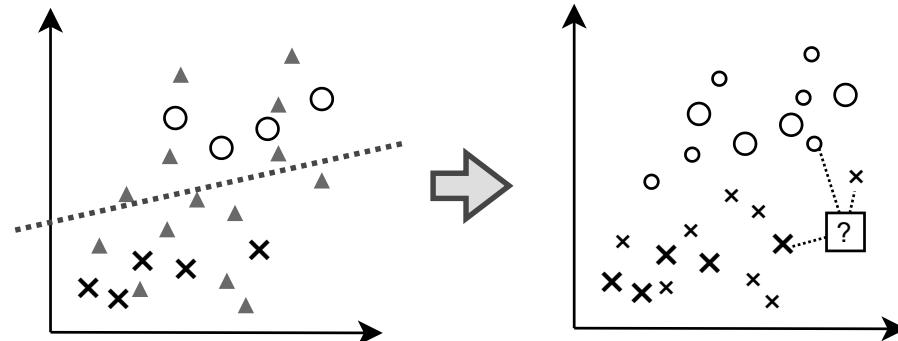
- 共訓練

- 特徴ベクトルを列で分割し、判断基準が異なる識別器を交互に用いる
- 片方の確信度が高いデータを、他方が正解付きデータとみなして学習



14.2.2 疑似ラベルの活用 (3/3)

- YATSI (Yet Another Two-Stage Idea) アルゴリズム
 - 正解付きデータで初期識別器を作成し、正解なしデータを識別
 - 正解付きデータと識別結果で異なる重みを用いた重み付き k-NN 識別器を作成
 - 繰り返し学習による誤りの増幅を避ける

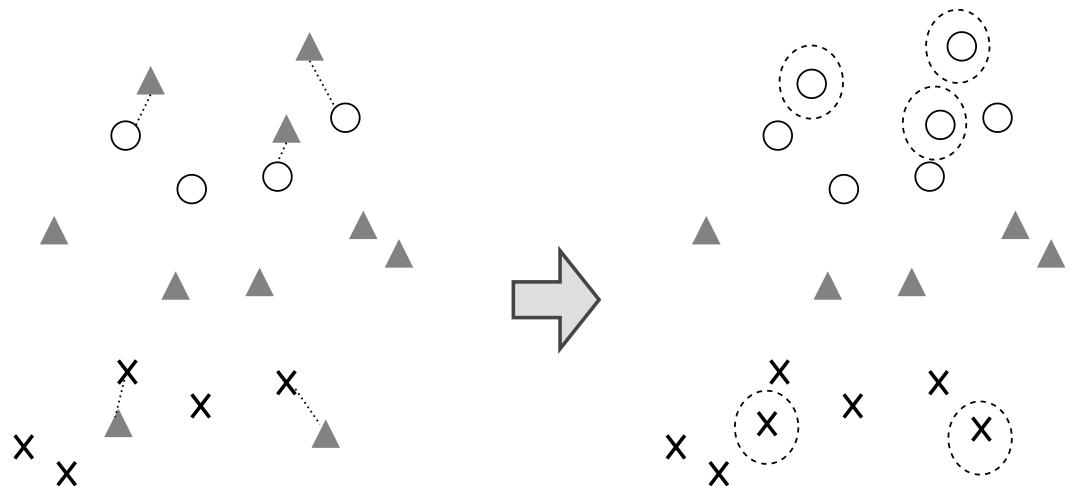


正解付きデータで作った
識別器で全データを識別

・正解付きデータ : 1
・識別後の正解なしデータ : $\alpha < 1$
の重みでk-NN

14.2.3 データの分布の活用 (1/2)

- ラベル伝搬法のアイデア
 - 特徴空間上のデータをノードとみなし、類似度に基づいたグラフ構造を構築する
 - 近くのノードは同じクラスになりやすいという仮定で、正解なしデータの予測を行う



14.2.3 データの分布の活用 (2/2)

- ラベル伝搬法の処理手順
 - ノード間の類似度計算はガウシアンカーネルや、 k-NN 法が使われる
 - 評価関数を最小化するように、予測値を伝播させて計算

$$J(\mathbf{f}) = \alpha \sum_{i < j} w_{ij} (f_i - f_j)^2 + (1 - \alpha) \sum_{i=1}^N (y_i - f_i)^2$$

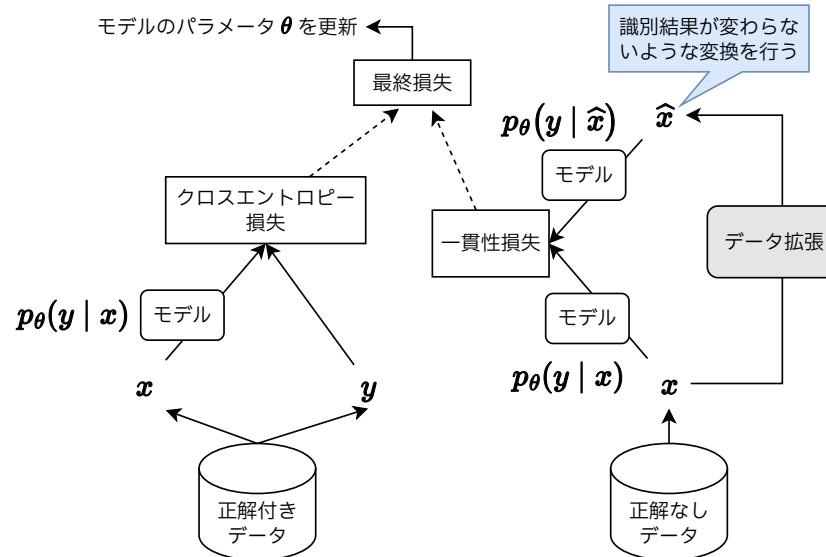
- f_i : i 番目のノードの予測値
- y_i : i 番目のノードの正解情報 (1 (正例), -1 (負例), 0 (正解なし))
- w_{ij} : i 番目のノードと j 番目のノードの類似度
- α ($0 \leq \alpha \leq 1$) は類似度の影響度を表すパラメータ
 - $\alpha = 0$ のときは正解情報は変更されず、 $\alpha = 1$ のときはすべてのノードが変更対象

14.2.4 整合性の活用 (1/2)

- データ拡張
 - 入力特徴に対して識別結果が変化しないことが期待される変換を行うことでデータを水増しする
 - 画像認識では、回転・平行移動・拡大縮小・色調変換などがよく使われる
 - 自然言語データに対しては、自動翻訳してもとに戻したり、重要な意味を担っていないとみなされる単語を近い意味の単語に置換する手法などがある
- データ拡張の問題点
 - もとの正解付きデータが少ない場合には、データ拡張結果のバリエーションもそれほど多くなく、効果があまりみられない場合もある

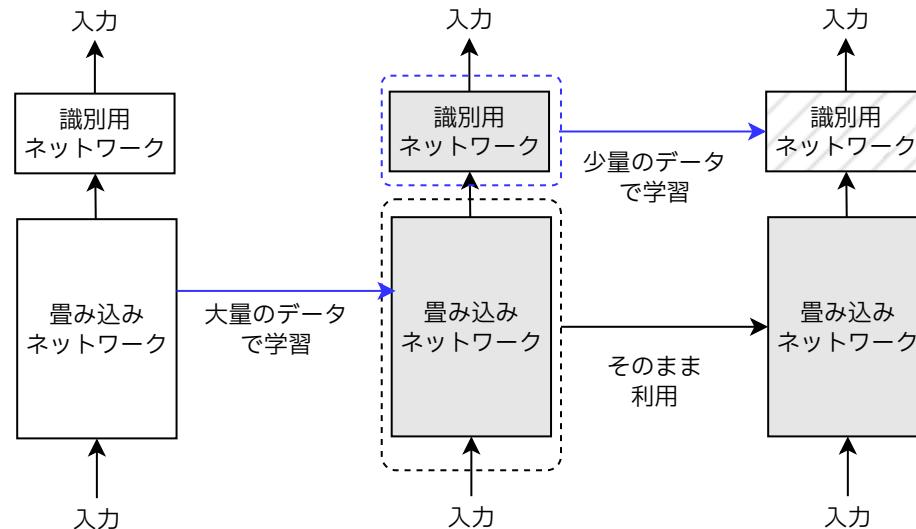
14.2.4 整合性の活用 (2/2)

- 正解なしデータの整合性の活用
 - データ拡張を大量の正解なしデータに対して行い、データ拡張前後の識別結果が変わらないような項を、損失関数に加える



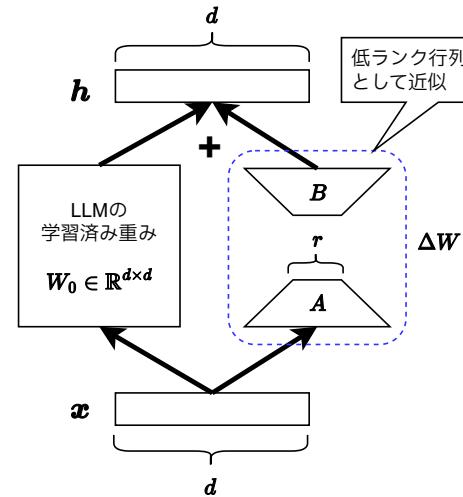
14.3 転移学習 (1/2)

- ・ 置み込みネットワークの転移学習
 - 大量のデータで学習済みの CNN の特徴抽出部分を利用
 - 対象タスクの少量のデータで識別用ネットワークのみ学習
 - 特徴抽出部分を追加学習する場合もある



14.3 転移学習 (2/2)

- 大規模言語モデル (LLM) の転移学習
 - LLM は一般的なタスクに関しては追加の学習は不要だが、医療分野や法律分野など、特定タスクに適応させる場合はファインチューニングを行う
 - LoRA (Low-Rank Adaptation)
 - LLM の重み行列に対する更新の情報を別の低ランク行列で近似する

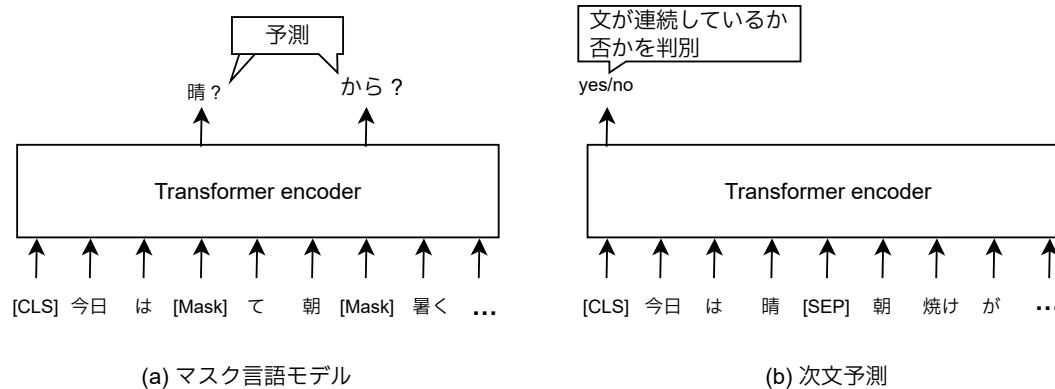


14.4 自己教師あり学習 (1/4)

- 自己教師あり学習とは
 - データの一部を隠して、元のデータを予測するタスクを自動的に設定
 - さまざまな応用が可能な特徴ベクトルを学習
 - 例：画像の一部を隠して元の画像を予測、文章の一部を隠して元の文章を予測
 - 学習したモデルは、少量の正解付きデータでファインチューニング

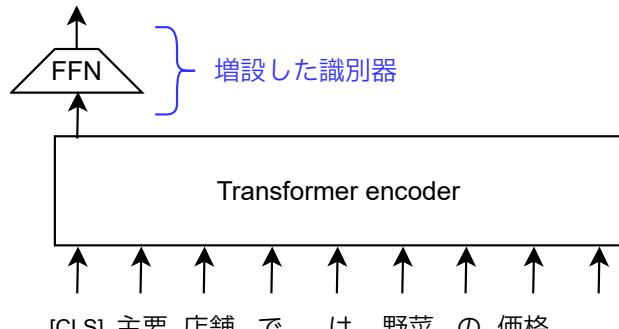
14.4 自己教師あり学習 (2/4)

- BERT の自己教師あり学習
 - Transformer の encoderのみを用いる
 - 2種の自己教師あり学習を事前学習として行う
 - 一部の入力単語を [Mask] に置換して、置換前の単語を予測
 - 2つの文が連続したものかどうかを判定（あまり有効ではないとされ、現在では使われていない）

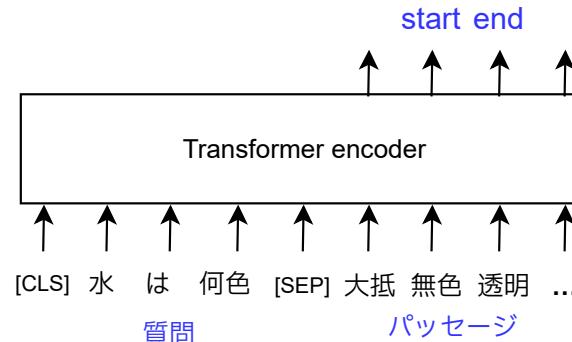


14.4 自己教師あり学習 (3/4)

- BERT のファインチューニング
 - 単純な識別問題など
 - [CLS] の出力を特徴ベクトルとする識別器を学習
 - 系列を扱う場合
 - 入出力をタスクに合わせてパラメータを再調整



(a) 文書分類など



(b) 機械読解

14.4 自己教師あり学習 (4/4)

- GPT の事前学習
 - Transformer の decoder のみを用いる
 - 自己回帰モデルとして、系列の次の単語を予測するタスクで学習
 - 単語の確率分布から計算されるクロスエントロピー損失を最小化
- GPT のファインチューニング
 - 質問応答データなどで、入力系列の続きを予測するモデルから、適切な応答を生成するモデルへとパラメータを再調整
 - 不適切な応答を生成しないように、強化学習を用いる場合もある

まとめ

- 少量データ学習は、正解付きデータが少ない場合の学習手法
- 少量データ学習の代表的な手法
 - 半教師あり学習
 - 正解なしデータの分布や整合性を活用して、識別器の性能を向上
 - 転移学習
 - 類似タスクで学習したモデルを対象タスクに適用
 - 自己教師あり学習
 - データの一部を隠して元のデータを予測するタスクで事前学習し、少量の正解付きデータでファインチューニング