

情報リテラシー（第14回 後期）

AIモデルの評価－学習・推論・評価指標を理解する－

今日のねらい

- AIの学習プロセスを理解できる
- 推論の意味を理解できる
- 評価指標（正解率・適合率・再現率・混同行列）を理解できる
- 過学習と未学習の違いを知る

AIの学習プロセス

データの分割

訓練データ

- AIモデルを学習させるデータ
- 例：手書き数字60,000枚

テストデータ

- 評価に使うデータ
- 例：手書き数字10,000枚
- **学習には使わない**

なぜ分ける？ 学習に使ったデータで評価すると正しく評価できない

推論とは 定義

学習したモデルで新しいデータを予測すること

MNIST（エムニスト）での推論

1. 学習: 60,000枚の手書き数字で学習
2. 推論: 新しい手書き数字を見せる
3. 予測: 「これは7だ」と予測

日常の例

- 顔認識: 学習した顔を認識
- 音声認識: 学習した言葉を認識
- スパムメール判定: 学習したパターンで判定

推論の性能を測るのが「評価」

なぜ評価が必要か

AIモデルの良し悪しを判断するため

- このモデルは信頼できるか？
- どれくらい正確か？
- どんな間違いをするか？
- 実用に耐えるか？

評価なしでは

- モデルの性能が分からぬ
- 改善の方向が分からぬ
- 実用化できない

→ **評価指標が必要**

評価指標①：正解率（Accuracy）

定義

全体で何%正解したか

計算式

正解率 = 正解数 ÷ 全体の数

例 (MNIST)

- テストデータ: 10,000個
- 正解: 9,800個
- 正解率 = $9,800 \div 10,000 = \text{98\%}$

特徴

- 最も基本的な指標
- 全体の性能を見る
- 分かりやすい

正解率の限界

問題点：データが偏っている場合

例：病気診断

- 健康な人: 990人
- 病気の人: 10人
- 合計: 1,000人

モデルA: 全員を「健康」と予測

- 正解率 = $990 \div 1,000 = 99\%$ ← 高い！
- でも病気の人10人を全員見逃し ← 最悪！
→ 正解率だけでは良いモデルか判断できない

MNISTの場合も

- 全体で98%正解（でも数字7は28個見逃している）
- 数字ごとの性能が見えない

評価指標②：適合率（Precision）

定義

「この数字だ」と予測したとき、実際に正しい割合

計算式

適合率 = 正しく予測した数 ÷ その数字だと予測した全体

例（数字7の場合）

- 「7だ」と予測: 1,050回
- そのうち実際に7: 1,000回
- 適合率 = $1,000 \div 1,050 = 95.2\%$

意味

- **予測の信頼性**
- 「7だ」と言ったとき、どれくらい信用できるか
- 間違った予測の少なさ

評価指標③：再現率（Recall）

定義

実際のその数字を、どれだけ正しく見つけられたか

計算式

再現率 = 正しく予測した数 ÷ 実際のその数字の数

例（数字7の場合）

- 実際の7: 1,028個
- 正しく「7だ」と予測: 1,000個
- 再現率 = $1,000 \div 1,028 = 97.3\%$

意味

- 見逃しの少なさ
- 実際の7をどれだけ見つけられたか
- 取りこぼしの少なさ

適合率と再現率の違い

	適合率 (Precision)	再現率 (Recall)
視点	予測した側から	実際の側から
質問	予測は正しいか？	見逃していないか？
重視	間違った予測を減らす	見逃しを減らす

具体例（病気診断）

適合率重視:

- 「病気だ」と診断したら確実に病気
- 誤診を減らす

再現率重視:

- 病気の人を見逃さない
- 見逃しを減らす

トレードオフ

- 適合率を上げる → 再現率が下がる / 再現率を上げる → 適合率が下がる

評価指標④：混同行列 (Confusion Matrix)

定義

どの数字をどの数字と間違えたか

例 (一部)

実際 \ 予測	0	1	7	8
0	970	0	2	8
1	0	1125	5	5
7	0	8	1000	20
8	5	3	10	956

読み方

- **行:** 実際の数字
- **列:** 予測した数字
- **対角線:** 正解
- **それ以外:** 間違い

評価指標のまとめ

指標	何を測るか	使い分け
正解率	全体の正解率	全体の性能
適合率	予測の信頼性	間違いを減らしたい
再現率	見逃しの少なさ	取りこぼしを減らしたい
混同行列	間違いのパターン	どう間違えるか知りたい

使い分けの例

スパムメール判定:

- 適合率重視: 大事なメールをスパムと判定しない
- 再現率重視: スパムを見逃さない

病気診断:

- 適合率重視: 誤診を減らす
- 再現率重視: 病気を見逃さない

過学習と未学習

過学習

訓練データに合わせすぎ

- 訓練データ: 99%正解
- テストデータ: 70%正解
- 新しいデータで性能が落ちる

原因: モデルが複雑すぎる、データが少ない

未学習

学習が不足

- 訓練データ: 70%正解
- テストデータ: 68%正解
- そもそも性能が低い

原因: モデルが単純すぎる、学習が不足

MNISTは？(第12回)

- 訓練データ: 約99%
- テストデータ: 約98%
- **良いモデル！** (過学習していない)

モデル

- **入力層**: 784ノード (28×28ピクセル)
- **隠れ層1**: 128ノード
- **隠れ層2**: 64ノード
- **出力層**: 10ノード (0-9の数字)

結果

- **正解率**: 約98%
- **学習時間**: 5エポック
- **パラメータ数**: 109,386個

今日の実習の流れ

Part 1: 第12回の復習

- MNISTモデルの学習を実行

Part 2: 評価指標の計算

1. 正解率の確認
2. 数字7の適合率と再現率を計算
3. 全ての数字の適合率と再現率
4. 混同行列の表示
5. 間違えやすい組み合わせの分析

まとめ

データサイエンス・AIの授業回で学んだこと

- **AIの基礎:** 機械学習、深層学習、生成AI
- **倫理的課題:** バイアス、プライバシー、著作権
- **実践:** 線形回帰、MNIST、NotebookLM/Copilot
- **評価:** 正解率、適合率、再現率、混同行列

今後の学習

- より専門的な科目へ
- AIを使いこなす力を養う
- 社会でのAI活用を考える

振り返り: Teamsに今日の感想を書いてください