

# 情報リテラシー(第14回 後期)ハンドアウト

---

## AIモデルの評価—学習・推論・評価指標を理解する—

---

### 1. 今日のねらい

- 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
  - 第12回のMNISTモデルを評価できる
- 

### 2. 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

**第12回（MNIST）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

---

### 3. 評価指標①：正解率（Accuracy）

**定義:** 全体で何%正解したか

**計算式:**  $\text{正解率} = \text{正解数} \div \text{全体の数}$

**例（MNIST）:** テストデータ10,000個、正解9,800個 →  $\text{正解率} = 9,800 \div 10,000 = 98\%$

**特徴:** 最も基本的な指標／全体の性能を見る／分かりやすい

---

### 4. 正解率の限界

データが偏っている場合、正解率だけでは良いモデルか判断できない

→ 詳細な評価には適合率・再現率が必要

---

### 5. 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:**  $\text{適合率} = \text{正しく予測した数} \div \text{その数字だと予測した全体}$

**例（数字7）:** 「7だ」と予測1,050回、そのうち実際に7が1,000回 →  $\text{適合率} = 1,000 \div 1,050 = 95.2\%$

**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

---

### 6. 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) :** 実際の7が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をどれだけ見つけれられたか／取りこぼしの少なさ

---

## 7. 適合率と再現率の違い

**適合率 (Precision) :** 予測した側から見る／予測は正しいか？／間違った予測を減らす

**再現率 (Recall) :** 実際の側から見る／見逃していないか？／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／再現率重視 → 病気の人を見逃さない

**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／バランスが重要

---

## 8. 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

**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

**分析:** 「7」を「1」と間違いやすい／「7」を「8」と間違いやすい

**使い分け:** 正解率 → 全体の性能／適合率 → 予測の信頼性／再現率 → 見逃しの少なさ／混同行列 → 間違いのパターン

---

## 9. 過学習と未学習

**過学習 (Overfitting) :** 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 (Underfitting) :** 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

**第12回のMNISTは? :** 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)

---

## 10. 今日の実習

**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:** 正解率、適合率、再現率、混同行列

---

## 11. まとめ (キーワード)

訓練データ／テストデータ／推論／正解率／適合率／再現率／混同行列／過学習／未学習