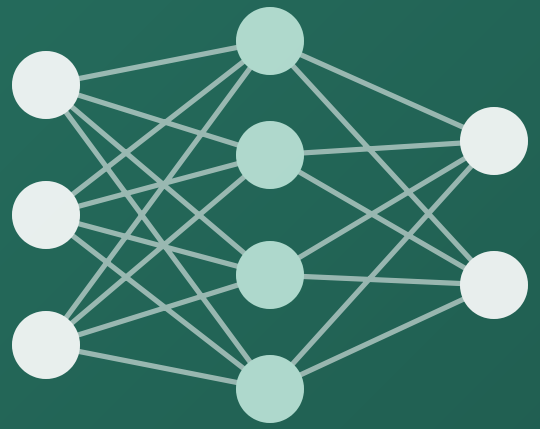
G検定対策講座　第 10 回

モデルの選択・評価

作ったAIの「良さ」を、どう測るか



今日学ぶこと

1. **モデルの良さとは**
   汎化性能・過学習・オッカムの剃刀
2. **検証手法**
   ホールドアウト・k分割交差検証
3. **回帰の評価指標**
   MSE・RMSE・MAE
4. **分類の評価指標**
   混同行列・適合率・再現率・F値・ROC/AUC
5. **情報量規準**
   AIC と BIC

そのモデル、
本当に使える？

訓練データで満点でも、新しいデータで30点なら役に立たない

いちばん大事な考え方

汎化性能

まだ見ぬデータに対する予測の力

だからテスト用のデータを
別に取っておく必要がある

対比 **訓練誤差 ／ 汎化誤差**

訓 訓練誤差

訓練データに対する誤差

モデルを複雑にすればいくらでも下がる

汎 汎化誤差

未知データに対する誤差

本当に大事なのはこちら

訓練誤差だけが低いモデル
＝ **過去問だけ得意な学生**

POINT 本番で点が取れなければ意味がない。**測るべきは汎化誤差**

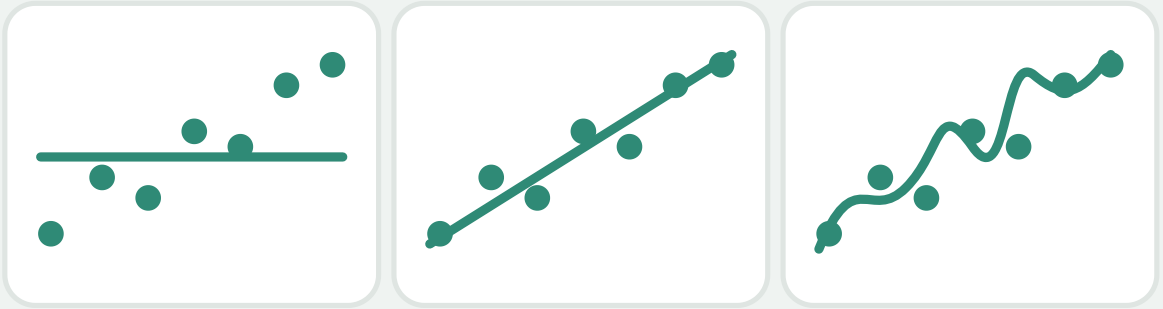
失敗は2種類

過学習と未学習

どちらも**汎化性能が低い**＝実用にならない

過学習の対策＝正則化・データ追加・単純化
未学習の対策＝複雑にする・特徴量を増やす

具体例 **同じデータに、3通りの当てはめ**



未学習	ちょうどよい	過学習
単純すぎて合わない	傾向をつかんでいる	点を追いかけすぎ

過学習＝訓練データのたまたまのノイズまで覚える

未学習＝訓練データすら学べていない

POINT 練習問題は満点だが本番は30点／練習問題でも50点——どちらもダメ

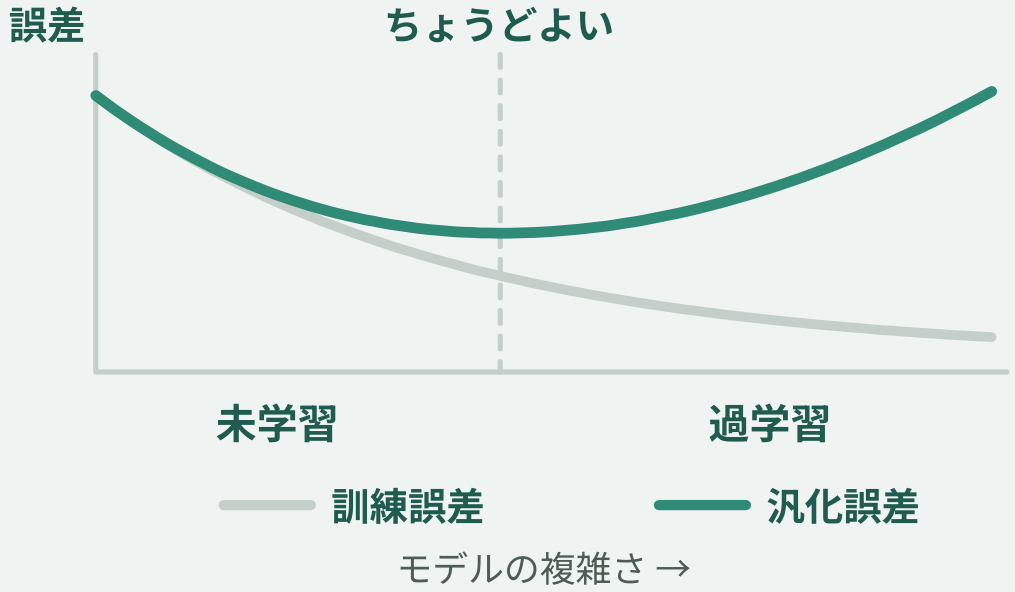
14世紀 ウィリアム・オッカム

オッカムの剃刀

同じ説明力なら、**より単純なほうがよい**

あとで出てくる**情報量規準**の思想的な背景になっている

具体例 複雑さと誤差の関係



複雑さが足りないと**未学習**

複雑にしすぎると**過学習**

ちょうどいい点を探すのが**モデル選択**

POINT 必要以上に複雑なモデルは避ける、というのが**オッカムの剃刀**

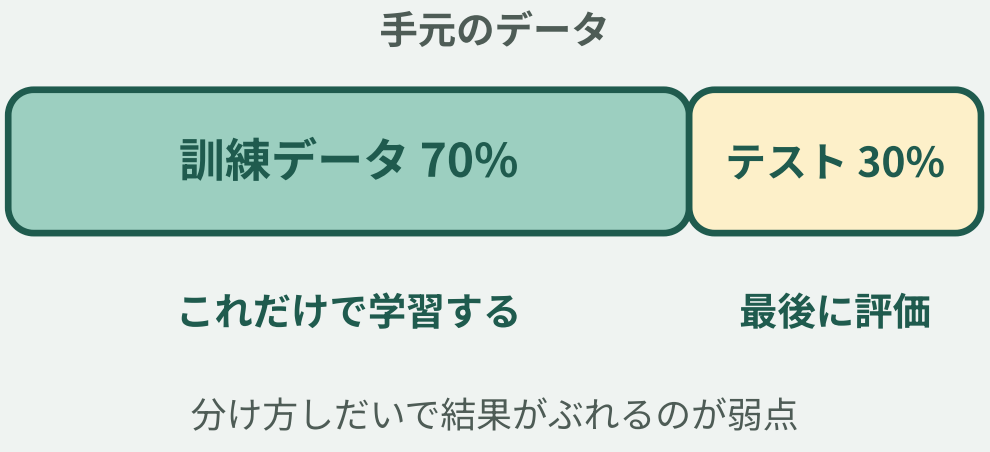
いちばんシンプル

ホールドアウト検証

データを**訓練用とテスト用に分ける**だけ

よくある割合は
7:3 や **8:2**

対比 利点 ／ 欠点



◯ 利点

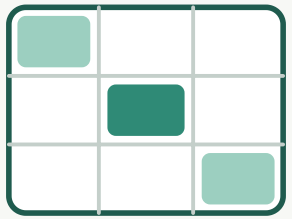
学習が1回だけ

計算が速い

× 欠点

分け方で結果がぶれる。運の要素がある

POINT データが少ないときは、次の**k分割交差検証**が向く



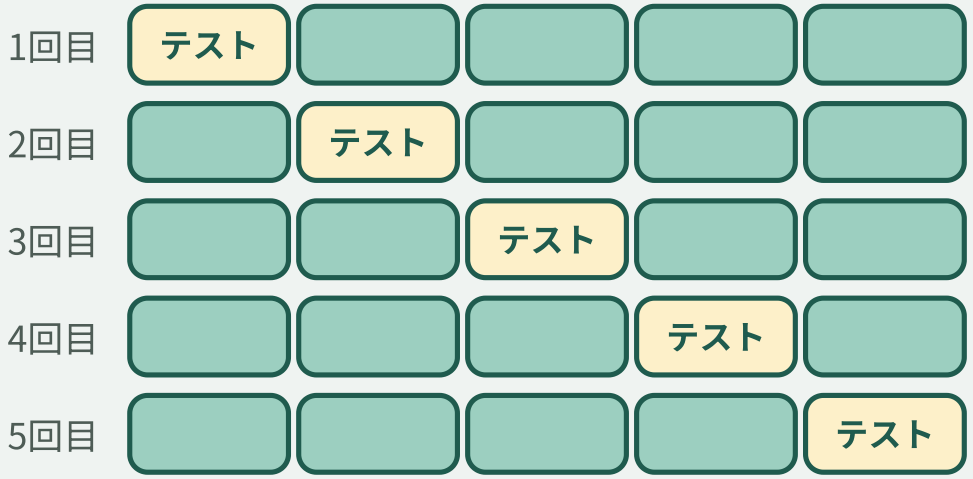
データが少ないとき

k分割交差検証

データを**k個に分け、k回くり返す**

よくあるkは**5**や**10**

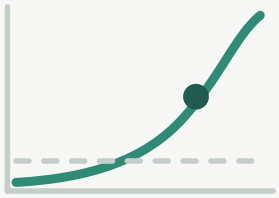
具体例 毎回ひとつをテストに使う



k回くり返して、結果を平均する

- 全データを訓練にもテストにも使える
- 分け方の偏りが平均され結果が安定する
- 欠点はk回学習するので時間がかかる

POINT データが少ないときの**定番**。ホールドアウトの弱点をつぶす手法



連続値を予測する問題

回帰の評価指標

予測と実測の**ズレの測り方**が3通り

大きな外れを避けたい → **RMSE**
外れ値が多い → **MAE**

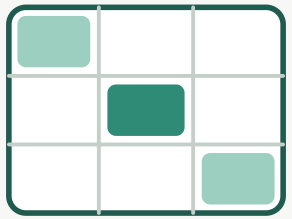
具体例 MSE ／ RMSE ／ MAE

MSE
平均二乗誤差
差を**2乗**して平均
外れ値に敏感

RMSE
MSEの**平方根**
単位が元データと同じ
直感的に分かりやすい

MAE
平均絶対誤差
差の**絶対値**を平均
外れ値に強い

POINT **2乗する＝大きな誤差を強く罰する。**ここが MSE・RMSE と MAE の分かれ目



分類の評価の土台

混同行列

分類の結果を**4つに整理**した表

頭の**真・偽**＝当たったかどうか
後ろの**陽性・陰性**＝
どう判定したか

具体例 4つのマス

		予測 陽性	予測 陰性
実際	陽性	**真陽性** 正しく陽性	**偽陰性** 見逃し
実際	陰性	**偽陽性** 冤罪	**真陰性** 正しく陰性

偽陽性＝陰性なのに陽性と判定
いわゆる**冤罪**

偽陰性＝陽性なのに陰性と判定
いわゆる**見逃し**

POINT これから見る指標は、**すべてこの4つの値から計算する**

3つの指標

正解率・適合率・再現率

どのマスを使うかが、それぞれ違う

適合率と再現率は**トレードオフ**
片方を上げると、もう片方が下がる関係

具体例 式と、意味

正解率

（真陽性＋真陰性）
÷ 全体
偏ったデータに弱い

適合率

真陽性 ÷（真陽性＋偽陽性）
陽性と判定したうち
本当に陽性だった割合
高いほど冤罪が少ない

再現率

真陽性 ÷（真陽性＋偽陰性）
実際に陽性のうち
拾えた割合
高いほど見逃しが少ない

POINT **適合率＝冤罪の少なさ／再現率＝見逃しの少なさ**

計算例題 ①

100人を検査した結果

この4つの数字から、指標を出していく

実際の試験でも
この形で数字が与えられる

具体例 **混同行列に置く**

		予測	
		陽性	陰性
実際	陽性	真陽性 **40 人**	偽陰性 **10 人**
実際	陰性	偽陽性 **5 人**	真陰性 **45 人**

陽性と判定したのは
40 ＋ 5 ＝ 45人

実際に陽性だったのは
40 ＋ 10 ＝ 50人

POINT まず**どのマスがいくつか**を、確実に読み取る

計算例題 ②

正解率を出す

正しく当てた**真陽性＋真陰性**を、全体で割る

全体の何割を
正しく判定できたか

具体例 色のついたマスを使う

		予測：陽性	予測：陰性
実際	陽性	真陽性 **40 人**	偽陰性 **10 人**
実際	陰性	偽陽性 **5 人**	真陰性 **45 人**

正解率

$(40 + 45) \div 100$

0.85

全体の 85% は正しく判定できた

POINT いちばん直感的な指標。ただし**偏ったデータでは信用できない**

計算例題 ③

適合率を出す

陽性と判定した人のうち、本当に陽性だった割合

> 分母は**タテ1列**
> （陽性と判定した合計）

具体例 陽性と判定した列を使う

		予測：陽性	予測：陰性
実際 陽性		真陽性 **40 人**	偽陰性 **10 人**
実際 陰性		偽陽性 **5 人**	真陰性 **45 人**

適合率

$40 \div (40 + 5)$

約 0.89

陽性と判定したうち
約9割が本当に陽性

POINT 高いほど**冤罪が少ない**。この例では冤罪は少なめ

計算例題 ④

再現率を出す

実際に陽性だった人のうち、拾えた割合

分母は**ヨコ1行**
（実際に陽性だった合計）

具体例 **実際に陽性だった行を使う**

		予測：陽性	予測：陰性
実際	陽性	真陽性 **40 人**	偽陰性 **10 人**
実際	陰性	偽陽性 **5 人**	真陰性 **45 人**

再現率

$40 \div (40 + 10)$

0.8

残りの 2割を見逃している

POINT 病気の診断では、この**見逃しが命取り**になる

試験の頻出ポイント

F値と、不均衡データの罠

F値＝適合率と再現率をひとつにまとめた指標

どちらか片方が低いと
F値は一気に下がる

具体例 **1,000件中、不正が10件しかないデータ**

罠 「全部が正常」と予測すると

正解率は 99%

一見すごい数字に見える

実 でも中身は

不正を1件も見つけていない

再現率は 0

POINT **不均衡なデータで正解率は信用できない。F値や再現率を見る**

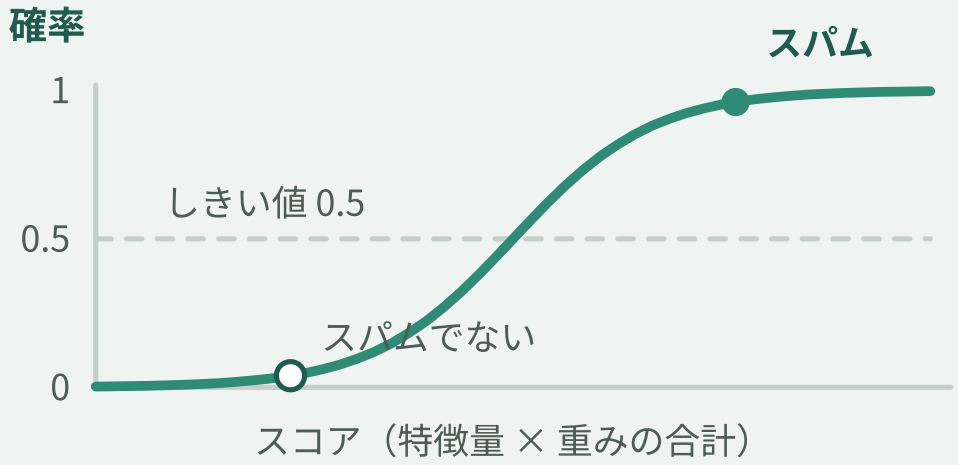
ROCの前ふり

閾値（しきい値）の問題

分類モデルの出力は、じつは**陽性である確率**

閾値ごとに性能が変わるので
モデルどうしを比べにくい

対比 **閾値を厳しく ／ 緩く**



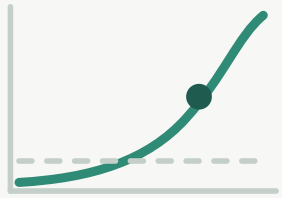
厳 厳しくする

見逃しが増え
再現率が下がる
冤罪は減り**適合率は上がる**

緩 緩くする

見逃しが減り
再現率は上がる
冤罪が増え**適合率は下がる**

POINT この「比べにくい」問題を解くのが、次の**ROC曲線とAUC**



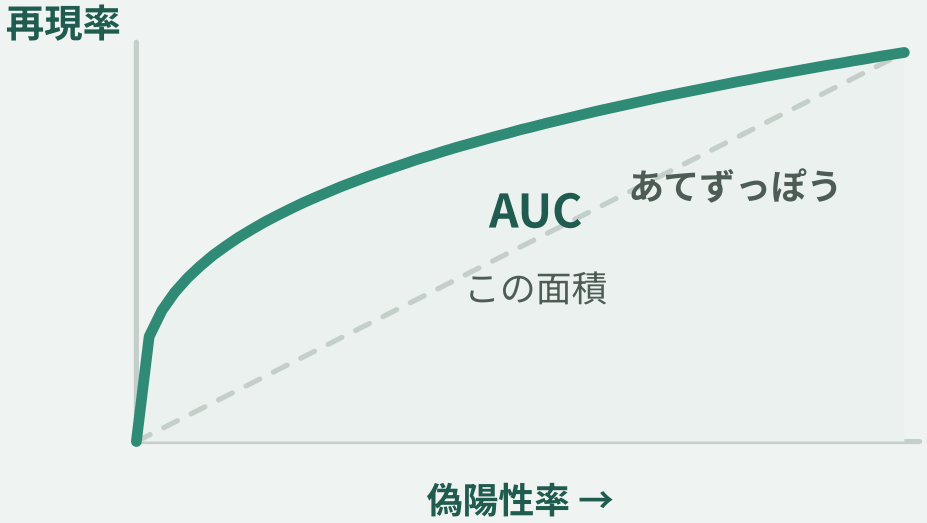
閾値によらない指標

ROC曲線とAUC

閾値を**端から端まで動かして**、一本の曲線にする

医療AI・不正検知で
よく使われる

具体例 曲線の下の面積が AUC



左上に張り出すほど良いモデル

0.5 ＝ あてずっぽうと同じ

1 に近いほど優秀

**0.8〜0.9 で実用的に良い
というのが目安**

POINT 閾値を気にせず、**モデルの素の実力**を1つの数字で比べられる

AIC と BIC

情報量規準

オッカムの剃刀を点数にしたもの

当てはまりにプラス点
複雑さにマイナス点
値が小さいほど良いモデル

対比 赤池（AIC）／ベイズ（BIC）

A 赤池情報量規準 AIC

当てはまり重視

複雑さへの罰はゆるめ

B ベイズ情報量規準 BIC

単純さ重視

複雑なモデルへの罰が強い

BICのほうが
より単純なモデルを
選ぶ

POINT **値が小さいほど良いモデル**。ちょうどいい複雑さを自動で選ぶ仕組み