DataScientist Society

モデルの精度評価

～AUC, Log Loss, RMSE, MAE, NDCG, MAP@n～

モデルの精度評価

➤ モデルの精度とは?

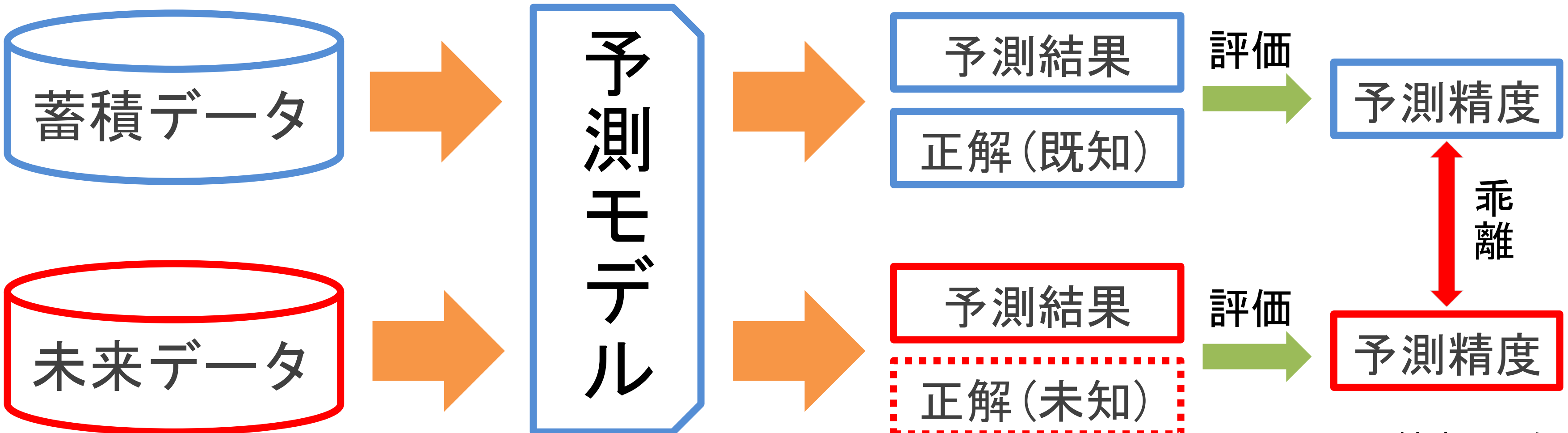


図1: 精度の評価

過去の蓄積データから予測モデルを構築し, 蓄積データに対するモデルの精度は, 評価指標(詳しくは後述)により測れる. しかし, 未知の未来データに対しては, 結果が出てからでないと確かめられない. また, 未来の結果が出たとしても, 蓄積データでの予測精度と未来データでの予測精度では, しばしば乖離が発生する. したがって,

モデルの精度 = 既知データでの予測精度 × 乖離幅

で考える必要がある.

モデルの精度評価

➢ 精度評価方法

　モデルが予測するものは, 個数, 分類, 順序などがあり, それらに対して, いくつかの評価指標が知られている. モデルの精度評価を正しく理解するには, 評価指標の正しい理解が必要不可欠である.

表1: 評価指標一覧

評価指標	予測対象	値域	見方
AUC	2群分類. 例:医療診断, 契約.	0 ～ 1	値が大きい方が良い
Log Loss	3群以上の分類. 例:運転行動.	0 ～ ∞	値が小さい方が良い
RMSE	個数. 例:売り上げ個数, 降水量.	0 ～ ∞	値が小さい方が良い
MAE	個数. 例:売り上げ個数, 降水量.	0 ～ ∞	値が小さい方が良い
NDCG	順序. 例:検索結果, 商品推薦.	0 ～ 1	値が大きい方が良い
MAP@n	(2群分類)^n×順序. 例:購入商品.	0～ $\sum_{i=1}^{n} 1/k$	値が大きい方が良い

モデルの精度評価

➢ 精度評価方法

　モデルが予測するものは, 個数, 分類, 順序などがあり, それらに対して, いくつかの評価指標が知られている. モデルの精度評価を正しく理解するには, 評価指標の正しい理解が必要不可欠である.

表1: 評価指標一覧

評価指標		値域	見方
AUC	2群分類	0 ～ 1	値が大きい方が良い
Log Loss	3群以上の分類. 例:運転行動.	0 ～ ∞	値が小さい方が良い
RMSE	個数. 例:売り上げ個数, 降水量.	0 ～ ∞	値が小さい方が良い
MAE	個数. 例:売り上げ個数, 降水量.	0 ～ ∞	値が小さい方が良い
NDCG	順序. 例:検索	0 ～ 1	値が大きい方が良い
MAP@n	(2群分類)^n×順序. 例:購入商品.	0 ～ $\sum_{i=1}^{n} 1/i$	値が大きい方が良い

今回の講座で
使用する評価指標

ここを見るだけ
では駄目！

モデルの精度評価(RMSE)

- RMSE (Root Mean Squared Error)

RMSEとは, N = 予測対象数, y_i = 実現値, $\hat{y}_i$ = 予測値として,

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

で定義される. 予測対象全てを誤差無しに当てることができれば, 値は0となる. 1ヶ所大きく外すと, 値はより大きくなる.

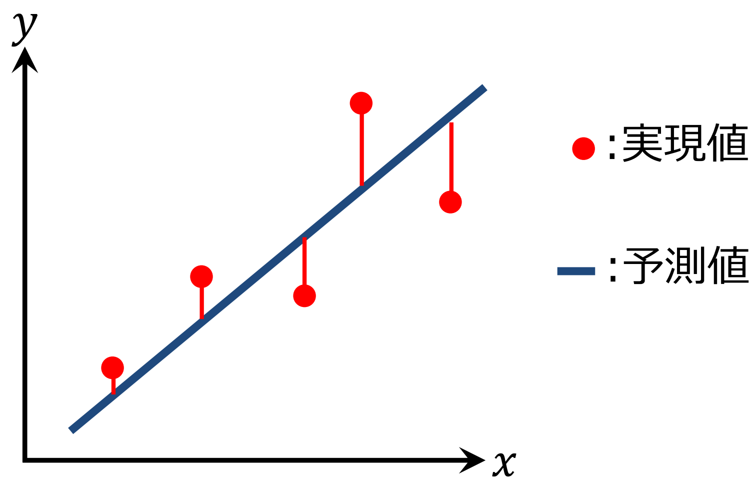


図2: Model

左図の赤い縦線の長さが, Modelの予測値と実現値の差を表しており,

$$y_i - \hat{y}_i$$

と対応している.

モデルの精度評価(分類)

➤ 分類の精度とは?

いま, 犬と猫の混合データがあり, データから犬か猫かを分類したい. 仮に, 体重で識別するモデルと, エサ代で識別するモデルを考える.

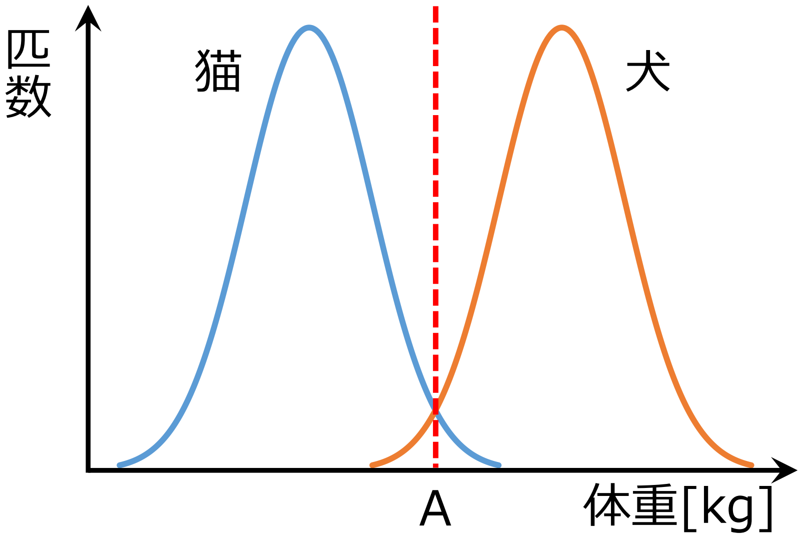


図3: 体重モデル

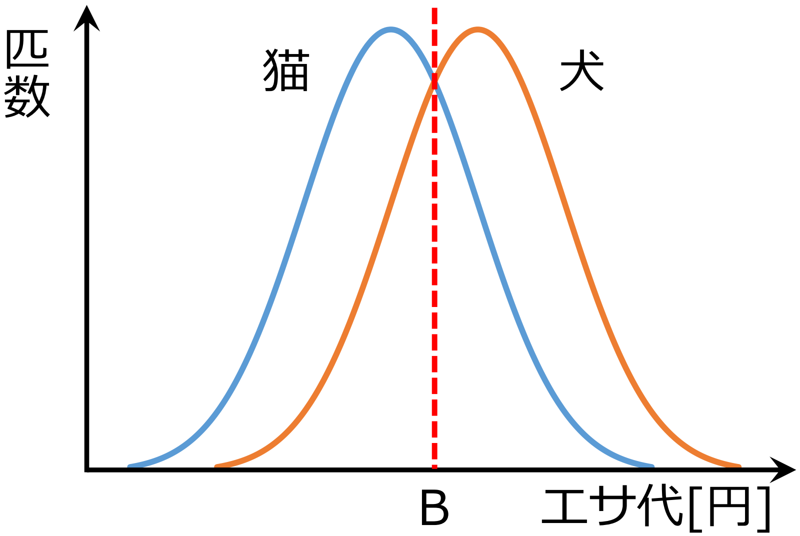


図4: エサ代モデル

図3, 4の場合, A, Bをそれぞれの閾値として分類する. なるべく2群の山が, 重ならないようになるモデルの方が識別が出来ている. そこで, 「重なり具合」から精度を評価することを考える.

モデルの精度評価(正解率と誤分類率)

- 正解率と誤分類率

「重なり具合」の指標の1つとして, 正解率と誤分類率が考えられる.

正解率 = (予測対象の内, 正しく分類した割合)

誤分類率 = (予測対象の内, 誤って分類した割合)

表2: 予測データ1

番号	真の分類	体重[kg]
1	犬	32.3
2	犬	25.4
3	犬	15.2
4	猫	10.1
5	犬	8.7
6	猫	7.5
7	犬	6.9
8	猫	4.8
9	猫	4.3
10	猫	3.2

予測
犬
10kg
猫

表2は10匹(犬5匹, 猫5匹)の予測対象データを体重の重い順に並べたものである. このとき, 10kgを閾値として**体重モデル(Model1)**を考えると,

正解率 = (3+4)/(5+5) = 0.7

誤分類率 = (1+2)/(5+5) = 0.3

と計算できる.

しかし, この正解率、誤分類率ではモデルの精度を上手く説明できない.

モデルの精度評価(正解率と誤分類率)

➢ 極端なデータの誤分類率

　極端に猫の件数が多い場合を考える. いま, 表2の猫の件数が, そのまま100倍の件数になるデータを表3とする. このとき, 犬5匹, 猫500匹に対して, Model1と, 体重に関係なく**全てを猫と分類するモデル(Model2)**を考える.

表3: 予測データ2

番号	真の分類	体重[kg]
1	犬	32.3
2	犬	25.4
3	犬	15.2
4～103	猫	10.1
104	犬	8.7
105～204	猫	7.5
205	犬	6.9
206～305	猫	4.8
306～405	猫	4.3
406～505	猫	3.2

Model1 予測　Model2 予測

犬　猫

10kg

猫　猫

・Model1の誤分類率:
(100+2)/(500+5) ≒ 0.20

・Model2の誤分類率:
(0+5)/(500+5) ≒ 0.01

　以上の結果より, 誤分類率を指標とすると, Model2の方が20倍精度の良いことになる.

→この場合は, 誤分類率は精度評価には使えない.

モデルの精度評価(混同行列)

➢ 混同行列(Confusion Matrix)

異なる数の2群を予測するモデルに対しての指標を考えるために, **混同行列**を定義する. 混同行列は下記の4種類の値を行列にしたものである.

True Positive(TP)　：犬を犬と予測した数
False Negative(FN)：犬を猫と予測した数
False Positive(FP)　：猫を犬と予測した数
True Negative(TN)　：猫を猫と予測した数

表4: 予測データ1(再掲)

番号	真の分類	体重[kg]
1	犬	32.3
2	犬	25.4
3	犬	15.2
4	猫	10.1
5	犬	8.7
6	猫	7.5
7	犬	6.9
8	猫	4.8
9	猫	4.3
10	猫	3.2

予測
犬
10kg
猫

表5: 予測データ1の混同行列

		予測の分類		
		犬	猫	合計
真の分類	犬	True Positive (TP) 3	False Negative (FN) 2	Positive (P) 5
	猫	False Positive (FP) 1	True Negative (TN) 4	Negative (N) 5

※ここでは、犬を予測したい対象(Positive)とする。

モデルの精度評価(適合率・再現率)

➢ 適合率と再現率

適合率

予測の分類が犬のうち、真の分類も犬である割合：TP/(TP+FP)=0.75

<u>犬の予測精度を評価</u>

再現率

真の分類が犬のうち、予測の分類も犬である割合：TP/(TP+FN)=0.6

<u>犬の取りこぼしを評価</u>

表4: 予測データ1(再掲)

番号	真の分類	体重[kg]	予測
1	犬	32.3	犬
2	犬	25.4	犬
3	犬	15.2	犬
4	猫	10.1	犬
5	犬	8.7	猫
6	猫	7.5	猫
7	犬	6.9	猫
8	猫	4.8	猫
9	猫	4.3	猫
10	猫	3.2	猫

10kg

表5: 予測データ1の混同行列

		予測の分類		
		犬	猫	合計
真の分類	犬	True Positive (TP) 3	False Negative (FN) 2	Positive (P) 5
	猫	False Positive (FP) 1	True Negative (TN) 4	Negative (N) 5

※ここでは、犬を予測したい対象(Positive)とする。

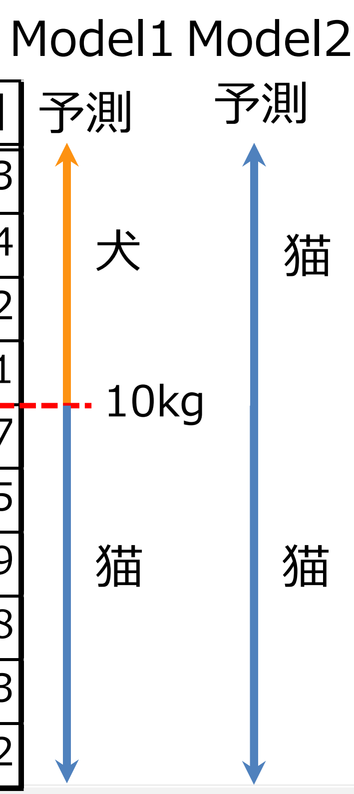
モデルの精度評価(適合率・再現率)

➢ 極端なデータの適合率と再現率

先ほどと同様に, 極端に猫の件数が多い場合を考える. 犬5匹, 猫500匹に対して, Model1と, **全てを猫と分類するモデル(Model2)**を考える.

表: 予測データ2(再掲)

番号	真の分類	体重[kg]
1	犬	32.3
2	犬	25.4
3	犬	15.2
4～103	猫	10.1
104	犬	8.7
105～204	猫	7.5
205	犬	6.9
206～305	猫	4.8
306～405	猫	4.3
406～505	猫	3.2



Model1

		予測の分類		
		犬	猫	合計
真の分類	犬	True Positive 3	False Negative 2	Positive 5
	猫	False Positive 100	True Negative 400	Negative 500

Model2

		予測の分類		
		犬	猫	合計
真の分類	犬	True Positive 0	False Negative 5	Positive 5
	猫	False Positive 0	True Negative 500	Negative 500

モデルの精度評価(適合率・再現率)

➢ 極端なデータの適合率と再現率

先ほどと同様に, 極端に猫の件数が多い場合を考える. 犬5匹, 猫500匹に対して, Model1と, **全てを猫と分類するモデル(Model2)**を考える.

	適合率	再現率
Model1	3/103 ≒ **0.02**	3/5 = **0.6**
Model2	算出不能(0/0)	0/5 = **0**

以上の結果より, 適合率は比較できず, 再現率はModel1が高くなる. 適合率, 再現率どちらを重視するかは, 問題により異なる.

Model1

		予測の分類		
		犬	猫	合計
真の分類	犬	True Positive 3	False Negative 2	Positive 5
	猫	False Positive 100	True Negative 400	Negative 500

Model2

		予測の分類		
		犬	猫	合計
真の分類	犬	True Positive 0	False Negative 5	Positive 5
	猫	False Positive 0	True Negative 500	Negative 500

モデルの精度評価(適合率・再現率)

➢ 適合率と再現率の例

適合率と再現率はどちらを使えばよいかは、分析目的に依存する.

例) クレジットカードのデフォルト予測
あなたがクレジットカードのデフォルトを予測するとき、どのような予測をするべきか？

再現率を重視する場合
・デフォルトする人の取りこぼしを評価.
・デフォルトする人をできるだけ検知したいときに使用.

適合率を重視する場合
・デフォルトする人の予測精度を評価.
・効率的にデフォルトする人を検知したいときに使用.

重視する指標を決め、条件付き指標(再現率が0.8を超える中でできるだけ適合率を上げるなど)を使用することもある.

モデルの精度評価(AUC)①

➢ 偽陽性率と真陽性率(引用: はじめてのパターン認識, 第3章)

適合率や再現率は**閾値を設定する必要があり**、また**TNが考慮されていない**. それらを考慮した指標であるAUCを紹介する. まずは下記のように**偽陽性率**と**真陽性率**を定義する.

真陽性率 = **犬の中**で犬と正しく分類した割合,
偽陽性率 = **猫の中**で犬と誤分類した割合

表4: 予測データ1(再掲)

番号	真の分類	体重[kg]
1	犬	32.3
2	犬	25.4
3	犬	15.2
4	猫	10.1
5	犬	8.7
6	猫	7.5
7	犬	6.9
8	猫	4.8
9	猫	4.3
10	猫	3.2

予測
犬
10kg
猫

表5: 予測データ1の混同行列

		予測の分類		
		犬	猫	合計
真の分類	犬	True Positive (TP) 3	False Negative (FN) 2	Positive (P) 5
	猫	False Positive (FP) 1	True Negative (TN) 4	Negative (N) 5

- 真陽性率: TP/P = 3/5 = 0.6
- 偽陽性率: FP/N = 1/5 = 0.2

モデルの精度評価(AUC)②

　偽陽性率(偽)と真陽性率(真)は, 両方とも片方の群のみから算出されるので, **2群の数が異なる場合でも有効な指標である**.
　閾値を+∞から下げていき, 偽陽性率が増えずに真陽性率が増えれば2群同士の「重なり合い」が少ない指標となる.

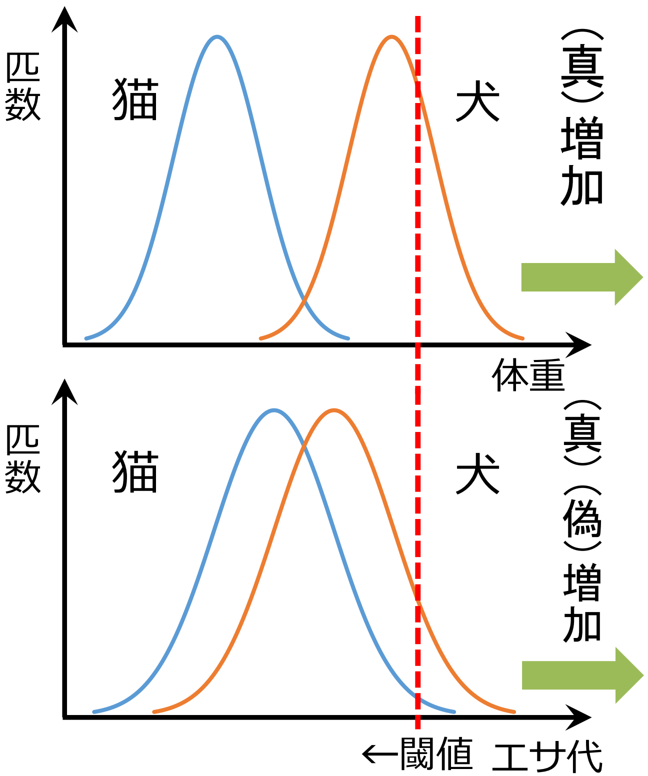


図5: 「重なり具合」偽陽性率と真陽性率

モデルの精度評価(AUC)②

偽陽性率(偽)と真陽性率(真)は, 両方とも片方の群のみから算出されるので, **2群の数が異なる場合でも有効な指標である**.

閾値を+∞から下げていき, 偽陽性率が増えずに真陽性率が増えれば2群同士の「重なり合い」が少ない指標となる.

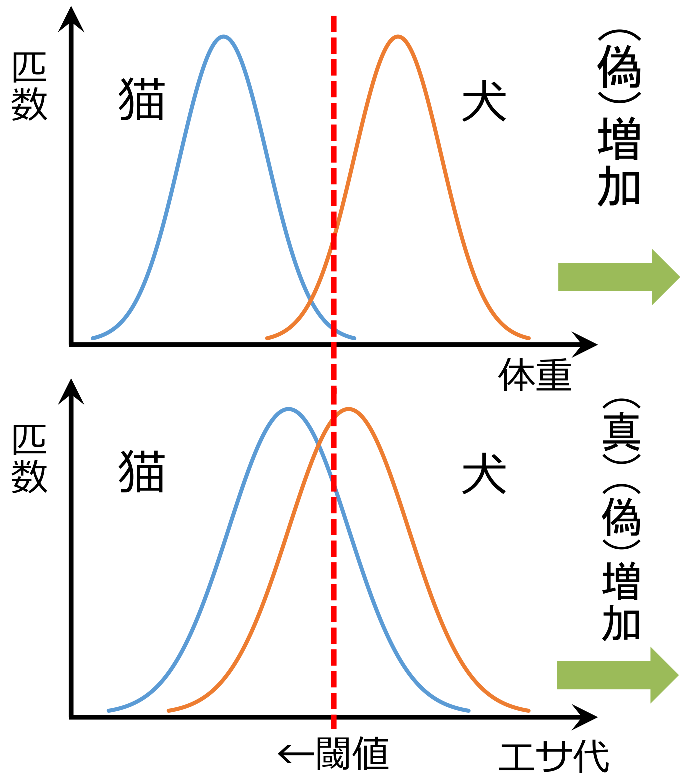


図5: 「重なり具合」偽陽性率と真陽性率

モデルの精度評価(AUC)②

偽陽性率(偽)と真陽性率(真)は, 両方とも片方の群のみから算出されるので, **2群の数が異なる場合でも有効な指標である**.

閾値を+∞から下げていき, 偽陽性率が増えずに真陽性率が増えれば2群同士の「重なり合い」が少ない指標となる.

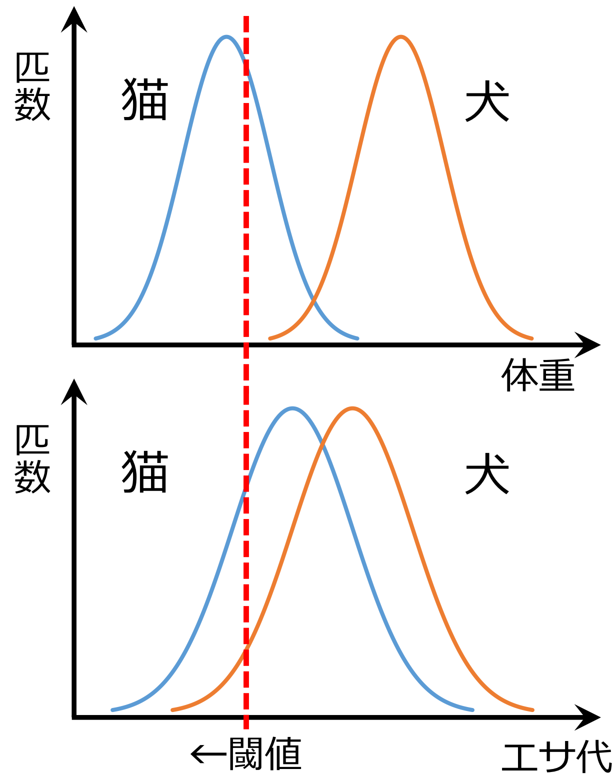


図5: 「重なり具合」偽陽性率と真陽性率

モデルの精度評価(AUC)②

偽陽性率(偽)と真陽性率(真)は, 両方とも片方の群のみから算出されるので, **2群の数が異なる場合でも有効な指標である**.

閾値を+∞から下げていき, 偽陽性率が増えずに真陽性率が増えれば2群同士の「重なり合い」が少ない指標となる.

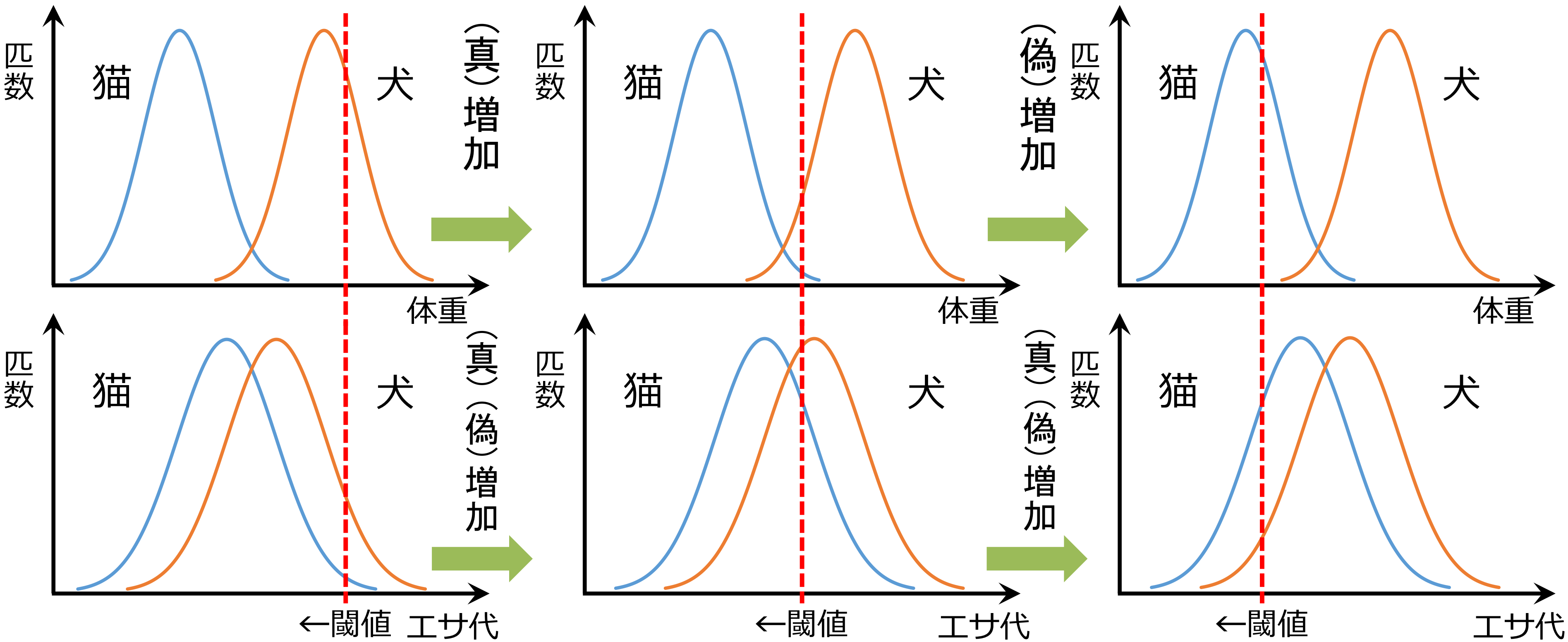


図5: 「重なり具合」偽陽性率と真陽性率

モデルの精度評価(AUC)③

- ROC曲線の作成

　先程の偽陽性率と真陽性率の変化をグラフで表したものを, ROC曲線(Receiver Operating Characteristic curve)と呼び, 横軸に偽陽性率, 縦軸に真陽性率をとり, 閾値を+∞から下げたときの変化を線分でつないだ曲線である.

表6: 予測データ1(再掲)

番号	真の分類	体重[kg]
1	犬	32.3
2	犬	25.4
3	犬	15.2
4	猫	10.1
5	犬	8.7
6	猫	7.5
7	犬	6.9
8	猫	4.8
9	猫	4.3
10	猫	3.2

表7: ROC曲線準備

番号	真の分類	体重[kg]	偽陽性率	真陽性率
1	犬	32.3	0.0	0.2
2	犬	25.4	0.0	0.4
3	犬	15.2	0.0	0.6
4	猫	10.1	0.2	0.6
5	犬	8.7	0.2	0.8
6	猫	7.5	0.4	0.8
7	犬	6.9	0.4	1.0
8	猫	4.8	0.6	1.0
9	猫	4.3	0.8	1.0
10	猫	3.2	1.0	1.0

モデルの精度評価(AUC)③

➢ ROC曲線の作成

先程の偽陽性率と真陽性率の変化をグラフで表したものを, ROC曲線(Receiver Operating Characteristic curve)と呼び, 横軸に偽陽性率, 縦軸に真陽性率をとり, 閾値を+∞から下げたときの変化を線分でつないだ曲線である.

表6: 予測データ1(再掲)

番号	真の分類	体重[kg]
1	犬	32.3
2	犬	25.4
3	犬	15.2
4	猫	10.1
5	犬	8.7
6	猫	7.5
7	犬	6.9
8	猫	4.8
9	猫	4.3
10	猫	3.2

偽陽性率

$= 0/5 = 0$

真陽性率

$= 0/5 = 0$

表7: ROC曲線準備

番号	真の分類	体重[kg]	偽陽性率	真陽性率
1	犬	32.3	0.0	0.2
2	犬	25.4	0.0	0.4
3	犬	15.2	0.0	0.6
4	猫	10.1	0.2	0.6
5	犬	8.7	0.2	0.8
6	猫	7.5	0.4	0.8
7	犬	6.9	0.4	1.0
8	猫	4.8	0.6	1.0
9	猫	4.3	0.8	1.0
10	猫	3.2	1.0	1.0

モデルの精度評価(AUC)③

➢ ROC曲線の作成

先程の偽陽性率と真陽性率の変化をグラフで表したものを, ROC曲線(Receiver Operating Characteristic curve)と呼び, 横軸に偽陽性率, 縦軸に真陽性率をとり, 閾値を+∞から下げたときの変化を線分でつないだ曲線である.

表6: 予測データ1(再掲)

番号	真の分類	体重[kg]
1	犬	32.3
2	犬	25.4
3	犬	15.2
4	猫	10.1
5	犬	8.7
6	猫	7.5
7	犬	6.9
8	猫	4.8
9	猫	4.3
10	猫	3.2

犬

偽陽性率
= 0/5 = 0

真陽性率
= 1/5 = 0.2

表7: ROC曲線準備

番号	真の分類	体重[kg]	偽陽性率	真陽性率
1	犬	32.3	0.0	0.2
2	犬	25.4	0.0	0.4
3	犬	15.2	0.0	0.6
4	猫	10.1	0.2	0.6
5	犬	8.7	0.2	0.8
6	猫	7.5	0.4	0.8
7	犬	6.9	0.4	1.0
8	猫	4.8	0.6	1.0
9	猫	4.3	0.8	1.0
10	猫	3.2	1.0	1.0

DataScientist Society

モデルの精度評価(AUC)③

➤ ROC曲線の作成

先程の偽陽性率と真陽性率の変化をグラフで表したものを, ROC曲線(Receiver Operating Characteristic curve)と呼び, 横軸に偽陽性率, 縦軸に真陽性率をとり, 閾値を+∞から下げたときの変化を線分でつないだ曲線である.

表6: 予測データ1(再掲)

番号	真の分類	体重[kg]
1	犬	32.3
2	犬	25.4
3	犬	15.2
4	猫	10.1
5	犬	8.7
6	猫	7.5
7	犬	6.9
8	猫	4.8
9	猫	4.3
10	猫	3.2

犬

偽陽性率
= 0/5 = 0

真陽性率
= 2/5 = 0.4

表7: ROC曲線準備

番号	真の分類	体重[kg]	偽陽性率	真陽性率
1	犬	32.3	0.0	0.2
2	犬	25.4	0.0	0.4
3	犬	15.2	0.0	0.6
4	猫	10.1	0.2	0.6
5	犬	8.7	0.2	0.8
6	猫	7.5	0.4	0.8
7	犬	6.9	0.4	1.0
8	猫	4.8	0.6	1.0
9	猫	4.3	0.8	1.0
10	猫	3.2	1.0	1.0

モデルの精度評価(AUC)③

➢ ROC曲線の作成

先程の偽陽性率と真陽性率の変化をグラフで表したものを, ROC曲線(Receiver Operating Characteristic curve)と呼び, 横軸に偽陽性率, 縦軸に真陽性率をとり, 閾値を+∞から下げたときの変化を線分でつないだ曲線である.

表6: 予測データ1(再掲)

番号	真の分類	体重[kg]
1	犬	32.3
2	犬	25.4
3	犬	15.2
4	猫	10.1
5	犬	8.7
6	猫	7.5
7	犬	6.9
8	猫	4.8
9	猫	4.3
10	猫	3.2

犬

偽陽性率

= 0/5 = 0

真陽性率

= 3/5 = 0.6

表7: ROC曲線準備

番号	真の分類	体重[kg]	偽陽性率	真陽性率
1	犬	32.3	0.0	0.2
2	犬	25.4	0.0	0.4
3	犬	15.2	0.0	0.6
4	猫	10.1	0.2	0.6
5	犬	8.7	0.2	0.8
6	猫	7.5	0.4	0.8
7	犬	6.9	0.4	1.0
8	猫	4.8	0.6	1.0
9	猫	4.3	0.8	1.0
10	猫	3.2	1.0	1.0

モデルの精度評価(AUC)③

- ROC曲線の作成

　先程の偽陽性率と真陽性率の変化をグラフで表したものを，ROC曲線(Receiver Operating Characteristic curve)と呼び，横軸に偽陽性率，縦軸に真陽性率をとり，閾値を+∞から下げたときの変化を線分でつないだ曲線である．

表6: 予測データ1(再掲)

番号	真の分類	体重[kg]
1	犬	32.3
2	犬	25.4
3	犬	15.2
4	猫	10.1
5	犬	8.7
6	猫	7.5
7	犬	6.9
8	猫	4.8
9	猫	4.3
10	猫	3.2

犬

偽陽性率 = 1/5 = 0.2

真陽性率 = 3/5 = 0.6

表7: ROC曲線準備

番号	真の分類	体重[kg]	偽陽性率	真陽性率
1	犬	32.3	0.0	0.2
2	犬	25.4	0.0	0.4
3	犬	15.2	0.0	0.6
4	猫	10.1	0.2	0.6
5	犬	8.7	0.2	0.8
6	猫	7.5	0.4	0.8
7	犬	6.9	0.4	1.0
8	猫	4.8	0.6	1.0
9	猫	4.3	0.8	1.0
10	猫	3.2	1.0	1.0

モデルの精度評価(AUC)③

- ROC曲線の作成

　先程の偽陽性率と真陽性率の変化をグラフで表したものを, ROC曲線(Receiver Operating Characteristic curve)と呼び, 横軸に偽陽性率, 縦軸に真陽性率をとり, 閾値を+∞から下げたときの変化を線分でつないだ曲線である.

表6: 予測データ1(再掲)

番号	真の分類	体重[kg]
1	犬	32.3
2	犬	25.4
3	犬	15.2
4	猫	10.1
5	犬	8.7
6	猫	7.5
7	犬	6.9
8	猫	4.8
9	猫	4.3
10	猫	3.2

犬

偽陽性率
= 1/5 = 0.2

真陽性率
= 4/5 = 0.8

表7: ROC曲線準備

番号	真の分類	体重[kg]	偽陽性率	真陽性率
1	犬	32.3	0.0	0.2
2	犬	25.4	0.0	0.4
3	犬	15.2	0.0	0.6
4	猫	10.1	0.2	0.6
5	犬	8.7	0.2	0.8
6	猫	7.5	0.4	0.8
7	犬	6.9	0.4	1.0
8	猫	4.8	0.6	1.0
9	猫	4.3	0.8	1.0
10	猫	3.2	1.0	1.0

モデルの精度評価(AUC)③

➤ ROC曲線の作成

先程の偽陽性率と真陽性率の変化をグラフで表したものを，ROC曲線(Receiver Operating Characteristic curve)と呼び，横軸に偽陽性率，縦軸に真陽性率をとり，閾値を+∞から下げたときの変化を線分でつないだ曲線である．

表6: 予測データ1(再掲)

番号	真の分類	体重[kg]
1	犬	32.3
2	犬	25.4
3	犬	15.2
4	猫	10.1
5	犬	8.7
6	猫	7.5
7	犬	6.9
8	猫	4.8
9	猫	4.3
10	猫	3.2

犬

偽陽性率

= 2/5 = 0.4

真陽性率

= 4/5 = 0.8

表7: ROC曲線準備

番号	真の分類	体重[kg]	偽陽性率	真陽性率
1	犬	32.3	0.0	0.2
2	犬	25.4	0.0	0.4
3	犬	15.2	0.0	0.6
4	猫	10.1	0.2	0.6
5	犬	8.7	0.2	0.8
6	猫	7.5	0.4	0.8
7	犬	6.9	0.4	1.0
8	猫	4.8	0.6	1.0
9	猫	4.3	0.8	1.0
10	猫	3.2	1.0	1.0

モデルの精度評価(AUC)③

- ROC曲線の作成

　先程の偽陽性率と真陽性率の変化をグラフで表したものを，ROC曲線(Receiver Operating Characteristic curve)と呼び，横軸に偽陽性率，縦軸に真陽性率をとり，閾値を+∞から下げたときの変化を線分でつないだ曲線である．

表6: 予測データ1(再掲)

番号	真の分類	体重[kg]
1	犬	32.3
2	犬	25.4
3	犬	15.2
4	猫	10.1
5	犬	8.7
6	猫	7.5
7	犬	6.9
8	猫	4.8
9	猫	4.3
10	猫	3.2

犬

偽陽性率
= 2/5 = 0.4

真陽性率
= 5/5 = 1.0

表7: ROC曲線準備

番号	真の分類	体重[kg]	偽陽性率	真陽性率
1	犬	32.3	0.0	0.2
2	犬	25.4	0.0	0.4
3	犬	15.2	0.0	0.6
4	猫	10.1	0.2	0.6
5	犬	8.7	0.2	0.8
6	猫	7.5	0.4	0.8
7	犬	6.9	0.4	1.0
8	猫	4.8	0.6	1.0
9	猫	4.3	0.8	1.0
10	猫	3.2	1.0	1.0

モデルの精度評価(AUC)③

- ROC曲線の作成

　先程の偽陽性率と真陽性率の変化をグラフで表したものを, ROC曲線(Receiver Operating Characteristic curve)と呼び, 横軸に偽陽性率, 縦軸に真陽性率をとり, 閾値を+∞から下げたときの変化を線分でつないだ曲線である.

表6: 予測データ1(再掲)

番号	真の分類	体重[kg]
1	犬	32.3
2	犬	25.4
3	犬	15.2
4	猫	10.1
5	犬	8.7
6	猫	7.5
7	犬	6.9
8	猫	4.8
9	猫	4.3
10	猫	3.2

犬

偽陽性率
= 3/5 = 0.6

真陽性率
= 5/5 = 1.0

表7: ROC曲線準備

番号	真の分類	体重[kg]	偽陽性率	真陽性率
1	犬	32.3	0.0	0.2
2	犬	25.4	0.0	0.4
3	犬	15.2	0.0	0.6
4	猫	10.1	0.2	0.6
5	犬	8.7	0.2	0.8
6	猫	7.5	0.4	0.8
7	犬	6.9	0.4	1.0
8	猫	4.8	0.6	1.0
9	猫	4.3	0.8	1.0
10	猫	3.2	1.0	1.0

モデルの精度評価(AUC)③

➢ ROC曲線の作成

先程の偽陽性率と真陽性率の変化をグラフで表したものを, ROC曲線(Receiver Operating Characteristic curve)と呼び, 横軸に偽陽性率, 縦軸に真陽性率をとり, 閾値を+∞から下げたときの変化を線分でつないだ曲線である.

表6: 予測データ1(再掲)

番号	真の分類	体重[kg]
1	犬	32.3
2	犬	25.4
3	犬	15.2
4	猫	10.1
5	犬	8.7
6	猫	7.5
7	犬	6.9
8	猫	4.8
9	猫	4.3
10	猫	3.2

犬

偽陽性率
= 4/5 = 0.8

真陽性率
= 5/5 = 1.0

表7: ROC曲線準備

番号	真の分類	体重[kg]	偽陽性率	真陽性率
1	犬	32.3	0.0	0.2
2	犬	25.4	0.0	0.4
3	犬	15.2	0.0	0.6
4	猫	10.1	0.2	0.6
5	犬	8.7	0.2	0.8
6	猫	7.5	0.4	0.8
7	犬	6.9	0.4	1.0
8	猫	4.8	0.6	1.0
9	猫	4.3	0.8	1.0
10	猫	3.2	1.0	1.0

モデルの精度評価(AUC)③

- ROC曲線の作成

　先程の偽陽性率と真陽性率の変化をグラフで表したものを, ROC曲線(Receiver Operating Characteristic curve)と呼び, 横軸に偽陽性率, 縦軸に真陽性率をとり, 閾値を+∞から下げたときの変化を線分でつないだ曲線である.

表6: 予測データ1(再掲)

番号	真の分類	体重[kg]
1	犬	32.3
2	犬	25.4
3	犬	15.2
4	猫	10.1
5	犬	8.7
6	猫	7.5
7	犬	6.9
8	猫	4.8
9	猫	4.3
10	猫	3.2

犬

偽陽性率
= 5/5 = 1.0

真陽性率
= 5/5 = 1.0

表7: ROC曲線準備

番号	真の分類	体重[kg]	偽陽性率	真陽性率
1	犬	32.3	0.0	0.2
2	犬	25.4	0.0	0.4
3	犬	15.2	0.0	0.6
4	猫	10.1	0.2	0.6
5	犬	8.7	0.2	0.8
6	猫	7.5	0.4	0.8
7	犬	6.9	0.4	1.0
8	猫	4.8	0.6	1.0
9	猫	4.3	0.8	1.0
10	猫	3.2	1.0	1.0

モデルの精度評価(AUC)④

　表7をプロットすると図6の様に階段の形の曲線が出来る. また, 2群が重なることなく完全に分類された場合は, 図7の様になり, ランダム分類された場合は, 図8の様になる.

➢ AUCの定義

ROC曲線下の面積をAUC値とする.

したがって, 図6のAUCは,

$$AUC = 0.2 \times 0.6 + 0.2 \times 0.8 + 0.6 \times 1.0 = 0.88$$

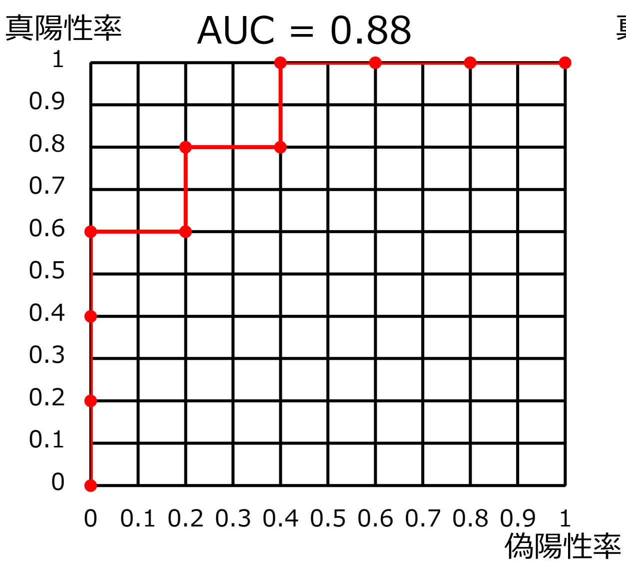


図6: ROC曲線(予測データ1)

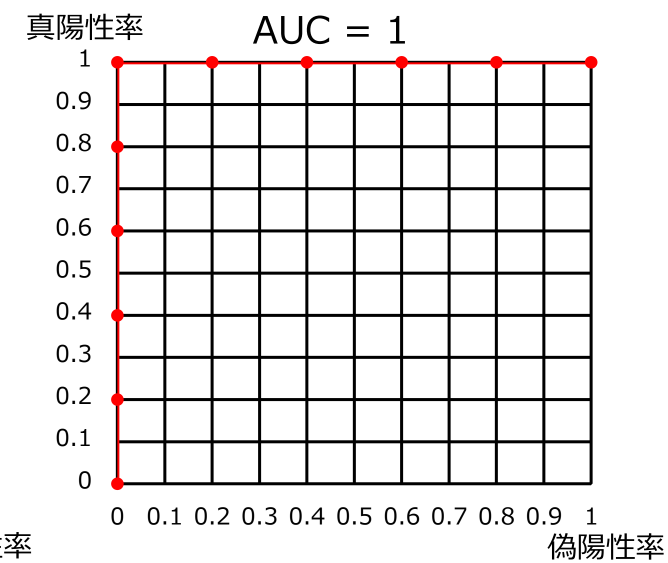


図7: ROC曲線(完全分類)

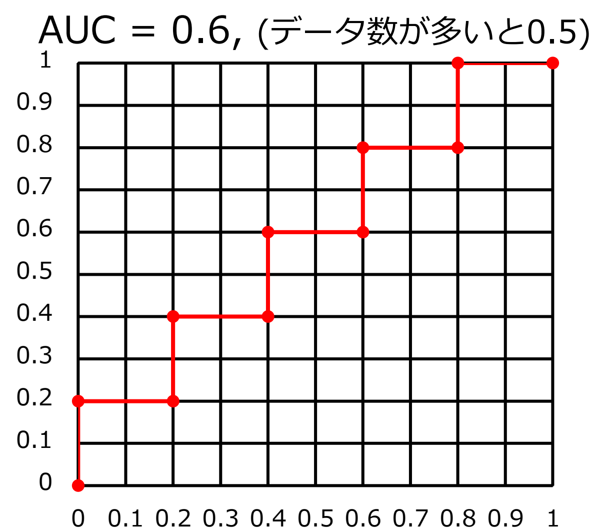


図8: ROC曲線(ランダム分類)

モデルの精度評価(AUC)⑤

- AUCの比較1

ここで, 先ほど誤分類率では出来なかった, モデルの精度評価に戻る. 極端に猫に偏った予測データ2のROC曲線は, 予測データ1と一致する. したがって, AUCの値も**一致する**ので, 群の偏りによらず安定な精度評価が出来ている.

表8: 予測データ2(再掲)

番号	真の分類	体重[kg]
1	犬	32.3
2	犬	25.4
3	犬	15.2
4～103	猫	10.1
104	犬	8.7
105～204	猫	7.5
205	犬	6.9
206～305	猫	4.8
306～405	猫	4.3
406～505	猫	3.2

計算例:

偽陽性率
= 100/500
= 0.2

真陽性率
= 3/5
= 0.6

表9: ROC曲線準備

番号	真の分類	体重[kg]	偽陽性率	真陽性率
1	犬	32.3	0.0	0.2
2	犬	25.4	0.0	0.4
3	犬	15.2	0.0	0.6
4～103	猫	10.1	0.2	0.6
104	犬	8.7	0.2	0.8
105～204	猫	7.5	0.4	0.8
205	犬	6.9	0.4	1.0
206～305	猫	4.8	0.6	1.0
306～405	猫	4.3	0.8	1.0
406～505	猫	3.2	1.0	1.0

モデルの精度評価(AUC)⑥

➢ AUCの比較2

　次に, 全てを猫と分類するモデルの精度評価を行う. いま, 全てを猫と分類するモデルのAUCは, 予測データ1も予測データ2も0.5である. なぜならば, 表10の様に猫フラグを全てに立てているのと同じことで, 全てのデータが同時に閾値を超えるので, (0,0)→(1,1)となる.

表10: 猫フラグ

番号	真の分類	猫フラグ
1	犬	1
2	犬	1
3	犬	1
4～103	猫	1
104	犬	1
105～204	猫	1
205	犬	1
206～305	猫	1
306～405	猫	1
406～505	猫	1

表11: ROC曲線準備

閾値	偽陽性率	真陽性率
+∞ ～ 1	0.0	0.0
1未満	1.0	1.0

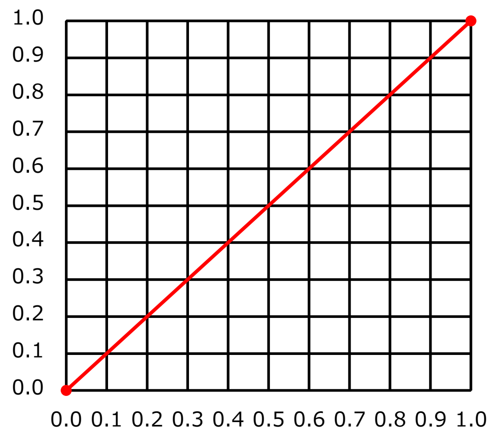


図9: ROC曲線(全猫)

AUC = 0.5

モデルの精度評価(AUC)⑦

➢ AUCの比較3

誤分類率とAUCの値を, 予測データ1と予測データ2ごとにまとめる.

表12: 精度評価比較(予測データ1)

予測データ1		モデル	
		体重モデル	全て猫と分類するモデル
指標	誤分類率	0.3	0.5
	AUC	0.88	0.5

表13: 精度評価比較(予測データ2)

予測データ2		モデル	
		体重モデル	全て猫と分類するモデル
指標	誤分類率	0.2	0.01
	AUC	0.88	0.5

表12, 表13を比べると, AUCは安定して精度評価を行っていることが分かる. したがって, 2群分類モデルの精度評価はAUCが有用である.

データの偏りに左右されないことを要求する背景の1つとして, 医療診断が挙げられる. 健康な人の群と, 健康でない人の群に分類する場合に, どうしても健康でない人の群は少なくなる. このとき, 医療診断の精度を正しく評価する上でAUCが使われている.

モデルの精度評価(AUC)⑦

➤ AUCの比較3

誤分類率とAUCの値を, 予測データ1と予測データ2ごとにまとめる.

表12: 精度評価比較(予測データ1)

予測データ1		モデル	
		体重モデル	全て猫と分類するモデル
指標	誤分類率	0.3	0.5
	AUC	0.88	0.5

表13: 精度評価比較(予測データ2)

予測データ2		モデル	
		体重モデル	全て猫と分類するモデル
指標	誤分類率	0.2	0.01
	AUC	0.88	0.5

表12, 表13を比べると, AUCは安定して精度評価を行っていることが分かる. したがって, 2群分類モデルの精度評価はAUCが有用である.

データの偏りに左右されないことを要求する背景の1つとして, 医療診断が挙げられる. 健康な人の群と, 健康でない人の群に分類する場合に, どうしても健康でない人の群は少なくなる. このとき, 医療診断の精度を正しく評価する上でAUCが使われている.

モデルの精度評価(AUC)⑧

- AUCの性質(引用: はじめてのパターン認識, 章末問題3.3)

  AUCの値がランダム分類の0.5を下回った, AUCの値を上げるにはどうすれば良いか?
  → 現在の結果と逆の分類にすれば良い(犬→猫, 猫→犬). 2群分類において, 全てを不正解にするには, 全てを当てないといけないからである.

- AUCの計算

  Python, Rともにいつくつかに AUCを計算するパッケージがある.

  ✓ 第2回のサンプルプログラム内で計算する

～補足資料～

補足資料(混同行列)①

- 混同行列(引用: はじめてのパターン認識, 第3章)
　2群分類では, 対象xが1つの群に属しているか否かを考えれば良く, 属していると予測する集合をp(陽性:positive), 属していないと予測する集合をn(陰性:negative)とする.

表14: 混同行列(Confusion Matrix)

		予測の分類		
		p	n	合計
真の分類	p^*	True Positive (TP)	False Negative (FN)	P = TP + FN
	n^*	False Positive (FP)	True Negative (TN)	N = FP + TN

補足資料(混同行列)②

➢ 記号の整理と評価値(引用: はじめてのパターン認識, 第3章)

表15: 記号の定義1

記号	定義
p^*	実際は群に属している集合.
n^*	実際は群に属していない集合.
P	p^*の件数.
N	n^*の件数.

表16: 記号の定義2

記号	定義
TP	p^*をp^*と予測した件数.
FN	p^*をn^*と予測した件数.
TN	n^*をn^*と予測した件数.
FP	n^*をp^*と予測した件数.

表17: 良く使用される評価値

名前	意味	定義
偽陽性率	$\boldsymbol{n^*}$**の中**でp^*と誤予測した割合.	FP/N
真陽性率	$\boldsymbol{p^*}$**の中**で正しくp^*と予測した割合.	TP/P
適合率(Pr)	p^*と予測した中で実際にp^*である割合.	TP/(TP+FP)
再現率(R)	予測対象外も含めた$p^*(=:p_{all}^*)$のうち, p^*と予測できた割合.	TP/P_all
正確度	正しく予測できた割合.	(TP+TN)/(P+N)
F値(F尺度)	適合率と再現率の調和平均(詳しくは補足資料1).	2/(1/Pr+1/R)

補足資料(適合率と再現率, ROC曲線)

➢ 適合率と再現率の関係

適合率と再現率は二律背反の関係(トレードオフ)であることが分かる.

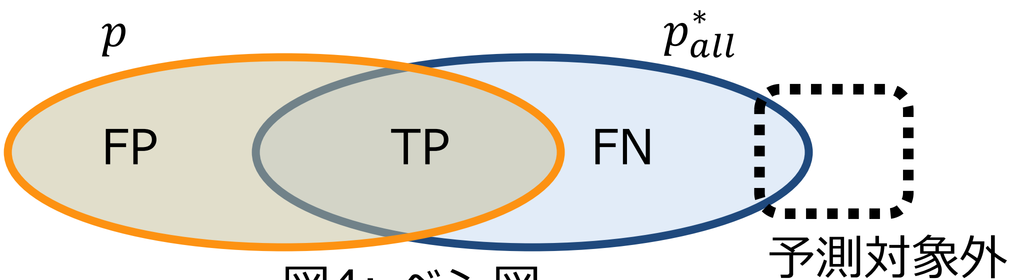


図4: ベン図

適合率:=TP/(TP+FP)
再現率:=TP/p^*_{all}

p^*_{all}の件数は定数なので, 再現率を上げるため, TPを増やそうとpを大きくするとFPも増える可能性があり, 結果的に適合率は下がる.

➢ ROC曲線の作成

例えば, Model1が2群(p^*:契約, n^*:非契約)に対してp^*に属する確率を出すとする. このとき, Xを予測するデータ集合とすると, $x \in X$に対して,

$$\mathrm{Model1}(x) = y_1 \quad (y_1:\text{スコア})$$

とかける. いま, 契約スコアの降順でデータを並べ, 上のデータから順に閾値を超える様に, 閾値を下げていく, 各閾値ごとに閾値を超えるデータをp^*(契約)として偽陽性率と真陽性率を計算し, 横軸を偽陽性率, 縦軸を真陽性率として, グラフにしたものがROC曲線である.

補足資料(MAE)

- MAE (Mean Absolute Error)

  MAEとは, N = 予測対象数, y_i = 実現値, $\hat{y}_i$ = 予測値として,

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

  で定義される. 予測対象全てを誤差無しに当てることができれば, スコアは0となり, モデルの誤差の平均を推し量れる.

補足資料(NDCG)

- NDCG (Normalized Discounted Cumulated Gain)

NDCGとは, 何かワードをWEB検索する例を考えると, k = 表示する検索結果数, rel_i = 表示i番目のWEBページ満足度として,

$$\mathrm{NDCG}_k = \frac{\mathrm{DCG}_k}{\mathrm{IDCG}_k}, \quad \mathrm{DCG}_k = \sum_{i=1}^{k} \frac{2^{rel_i} - 1}{\log_2(i+1)},$$

$$\mathrm{IDCG}_k = \mathrm{DCG}_k\text{最大スコア(満足度の高い順に検索結果を表示)}$$

で定義される. 満足度の高い順に検索結果を表示できれば, 最大スコアの1となり, 0に近づくほど精度が悪くなる.

✓ 具体例

表2: 検索結果表示順序と満足度

WEBページ	i : 表示順序	rel_i : 満足度
A	1	5
B	3	4
C	2	3

$$\mathrm{DCG}_k = \frac{2^5-1}{\log_2(1+1)} + \frac{2^4-1}{\log_2(3+1)} + \frac{2^3-1}{\log_2(2+1)} ≒ 42.916$$

$$\mathrm{IDCG}_k = \frac{2^5-1}{\log_2(1+1)} + \frac{2^4-1}{\log_2(2+1)} + \frac{2^3-1}{\log_2(3+1)} ≒ 43.964$$

$$\mathrm{NDCG}_k = \frac{\mathrm{DCG}_k}{\mathrm{IDCG}_k} ≒ 0.976$$

補足資料(MAP@n)

- MAP@n (Mean Average Precision)

  MAP@nとは, 最大n個の商品推薦の例を考えると,
  U = 予測対象人数, $p_u(k)$ = ユーザuにおいてk番目までの当たり確率,
  h = 予測した商品数, m = ユーザuが実際に購入した商品数として,

$$\mathrm{MAP@}n = \frac{1}{U}\sum_{u=1}^{U}\left(\frac{1}{\min(m,n)}\sum_{k=1}^{\min(h,n)} P_u(k)\right)$$

m=0のとき,

$$\frac{1}{\min(m,n)}\sum_{k=1}^{\min(h,n)} P_u(k) = 0$$

で定義される. 最大スコアは $\sum_{k=1}^{n} 1/k$(標準化された場合は1).

補足資料(Log Loss)

➢ Log Loss

Log Lossとは, 3群以上の分類に用いられることが多く, N = 予測対象数, M = 分類する群数, y_{ij} = i番目の対象が真にj番目の群のとき1, そうでないとき0, p_{ij} = i番目の対象がj番目の群である予測確率として,

$$\mathrm{LogLoss} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M} y_{ij}\log(p_{ij})$$

で定義される. 0に近いほど良いスコアである.

引用, 参考文献

[1] 平井有三：はじめてのパターン認識, 2012年, 森北出版株式会社
[2] 石田基広：Rで学ぶデータ・プログラミング入門, 2012年, 共立出版