

○ 分類モデルにおける代表的な評価指数

ここでは、TP、FP、FN、TNを利用した代表的な評価指標を4つ、紹介します。混合行列の表も再掲しますので、参照しながら読み進めてください。

■ 正解率 (Accuracy)

$$\text{正解率} = \frac{\text{TP} + \text{TN}}{\text{全体の数 (TP + FP + FN + TN)}}$$

	正解が「○」	正解が「×」
「○」と予想	TP	FP
「×」と予想	FN	TN

T: True (予想が当たっている)
F: False (予想が当たっていない)
P: Positive (「○」)
N: Negative (「×」)

正解率は、全体のデータ数のうち正しく分類できたデータ数の割合を計算できます。これは、一般にいう正解率や正答率と同じものです。

■ 再現率 (Recall)

$$\text{再現率} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

再現率は、実際にマルだったデータのうち、正しくマルとして分類できたデータの割合として計算されます。再現率が用いられるケースとしては「バツをマルとして分類しても問題なく、かつマルのものは確実にマルとして分類したい」といった状況下です。具体例としては、病気の診断が挙げられます。病気の診断においては、病気ではないデータ(バツ)を病気である(マル)と判断(偽陽性)してしまうことよりも、病気であるデータ(マル)を見逃してしまうこと(偽陰性)の方が危険です。そのため、再現率が重視されるのです。

■ 適合率 (Precision)

$$\text{適合率} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

適合率は、マルとして分類したデータのうち、正しく分類できたデータの割合として計算されます。適合率は再現率とは反対に「マルをバツとして分類してもよいが、バツのものは確実にバツとして分類したい」といった場合に用いられます。たとえばインターネットの検索システムです。この場合は、膨大な数のWebページの中から検索ワードに適合するページをなるべく絞り込んで表示することが求められるため、適合率が高いモデルがよいのです。

■ F値 (f-score)

$$\text{F値} = \frac{2 \times \text{再現率} \times \text{適合率}}{\text{再現率} + \text{適合率}}$$

再現率と適合率は分かりやすい指標ではありますが、これらを際限なく高めたモデルがよいかというと、必ずしもそうではありません。たとえば病気の診断においては、検査した人全員を病気であると分類してしまえば再現率は100%となりますが、この数値が実用的でないのは、言うまでもないことです。

そこで重要視されるのが**F値**です。実は再現率と適合率はトレードオフの関係にあり、一方が大きくなるともう一方は小さくなります。この再現率と適合率の平均(調和平均)を取ることで、よりよい評価指標となります。これをF値と呼ぶのです。

まとめ

□ 評価の際は適切な評価指標を用いる