

suzuki-navi's blog

2020-07-01

混同行列・適合率・再現率のまとめ

統計

このあたりの用語は試験の勉強をするたびに混乱して覚え直している気がする。あらためてまとめておく。

これまでに3つの試験勉強で出てきた。

- [AWS Certified Machine Learning – Specialty](#)
- [統計検定2級](#)
- [G検定](#)

混同行列、confusion matrix

↓予測 \ →真実	正	負
正	TP	FP
負	FN	TN

			予測値	正解
TP	True Positive	正しく正と判断	正	正
FP	False Positive	誤って正と判断	正	負
FN	False Negative	誤って負と判断	負	正
TN	True Negative	正しく負と判断	負	負

	がん検診での例
TP	がんだと正しく判定
FP	がんでないにも関わらずがんであると判定される誤検知
FN	がんであるにも関わらずがんではないと判定される見逃し
TN	がんではないと正しく判定

指標

正解率、精度、accuracy

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

再現率、感度、検出率、真陽性率、recall

$$Recall = \frac{TP}{TP + FN}$$

- 見逃しの損害が大きい場合に重要な指標
- 0.0はダメ、1.0に近いほどよい
- 検索の場合：マッチしてほしいアイテムの中で、実際にマッチした割合
- 反対は偽陰性率、False negative rate

適合率、precision

$$Precision = \frac{TP}{TP + FP}$$

- 見つけ出したもののうち正解の割合
- 0.0はダメ、1.0に近いほどよい

- 検索の場合：マッチしたアイテムの中で、本当にマッチしてほしいアイテムの割合

F値

再現率(Recall)と適合率(Precision)の調和平均

$$F = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}} \\ = \frac{2Recall Precision}{Recall + Precision}$$

- 0.0はダメ、1.0に近いほどよい
- 再現率と適合率はトレードオフの関係にあるため、両方をバランス良く見たい場合の指標として使える

調和平均は逆数の算術平均の逆数。例えば1と0.5の調和平均は0.67になる。

偽陽性率、False positive rate

$$\frac{FP}{TN + FP}$$

- 本当はnegativeなものの中で、検出してしまった割合
- 1.0はダメ、0.0に近いほどよい
- 第1種の過誤で α と一致

偽陰性率、False negative rate

$$\frac{FN}{TP + FN}$$

- 本当はpositiveなものの中で、検出できなかった割合
- 1.0はダメ、0.0に近いほどよい
- 再現率、recallの反対
- 第2種の過誤(β)

第1種の過誤と第2種の過誤

統計的仮説検定での話。

「対立仮説が正しい」「帰無仮説を棄却する」が上の議論でいう「正」「陽」、
「帰無仮説が正しい」「帰無仮説を棄却しない」が上の議論でいう「負」「陰」
に相当する。

↓検定結果\→真実	対立仮説が正しい	帰無仮説が正しい
帰無仮説を棄却する	検定結果は正しい	第1種の過誤(α)
帰無仮説を棄却しない	第2種の過誤(β)	検定結果は正しい

- 第1種の過誤
 - 有意水準（帰無仮説を棄却する基準となる確率、 α ）と一致
 - 偽陽性、FP、False positive
- 第2種の過誤
 - 偽陰性、FN、False negative

suzuki-navi 119日前

★+

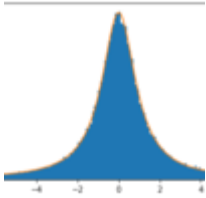
0

0

シェア

ツイート

関連記事



2020-08-23

正規分布に従っているが平均と分散が不明のn個の標本があるときのn+1個目の標本の分布

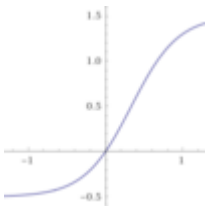
正規分布に従っているが平均と分散が不明の、サイズが の標本が...

$$\frac{1}{e^{-x}} =$$

2020-07-09

ディープラーニングの基本 - G検定の勉強ノート

G検定 2020#2 受験時の勉強ノートの全部で5ページのうちの3ペー...



2020-05-21

勾配降下法の自作アルゴリズム

前回の記事でよく知られている勾配降下法のアルゴリズムを書き...

2020-05-17

JavaとScalaにおける浮動小数点数の大小比較

浮動小数点数の数値の種類 2つのゼロ NaN ソート Java Scala Sc...

コメントを書く

« 人工知能の一般知識 - G検定の勉強ノート

Elasticsearch/Kibanaのインストールからc...
»

プロフィール



suzuki-navi

読者になる

1

検索

記事を検索

リンク

私の記事一覧

最新記事

ParquetファイルをJavaで生成する

AWS RDSのログファイルをダウンロードするPythonとRubyのスクリプト

PostgreSQLのテーブルやデータベースのデータサイズを調べるには

Parquet Toolsをインストール

AtCoder参戦日記 ARC104 — Scalaの限界を知る

もっと見る

月別アーカイブ

▽ 2020 (74)

2020 / 10 (17)

2020 / 9 (8)

2020 / 8 (5)

2020 / 7 (15)

2020 / 6 (16)

2020 / 5 (13)

カテゴリ

AWS (14)

Linuxコマンド (6)

AtCoder (4)

機械学習 (10)

GCP (4)

統計 (12)

Metabase (4)

ネットワーク (1)

これまでに書いたもの/してきたこと

はてなブログをはじめよう！

suzuki-naviさんは、はてなブログを使っています。あなたもはてなブログをはじめてみませんか？

はてなブログをはじめ（無料）

はてなブログとは

 suzuki-navi's blog

Powered by Hatena Blog | ブログを報告する