Healthcare Data Analytics

Precision / Recall tradeoff und mehr als binäre Klassifikation

Dr. Michael Strobel

11.04.2022

Inhalt

Letzte Woche

- Umgang mit fehlenden Datenpunkten
- Daten Pipelines
- Binäre Klassifikationsaufgaben
- Konfusionsmatrix
- Precision / Recall

Diese Woche

- Precision / Recall tradeoff
- Multi-Klassen Klassifikation

Precision / Recall tradeoff

Erinnerung: Precision / Recall

Precision: Wie viele der erkannten Objekte sind relevant?

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

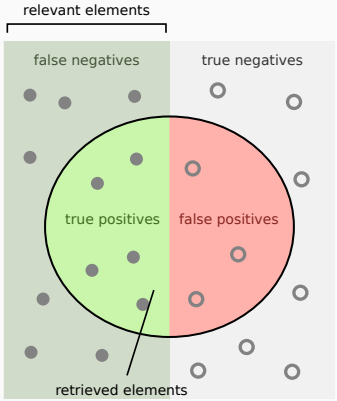
Recall: Wie viele der relevanten Objekte wurden erkannt?

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

Specificity: Wie viele der nicht relevanten Objekte wurden erkannt?

$$\text{Specificity} := \frac{TN}{TN + FP}$$

Precision Recall, Visualisierung



How many retrieved items are relevant?

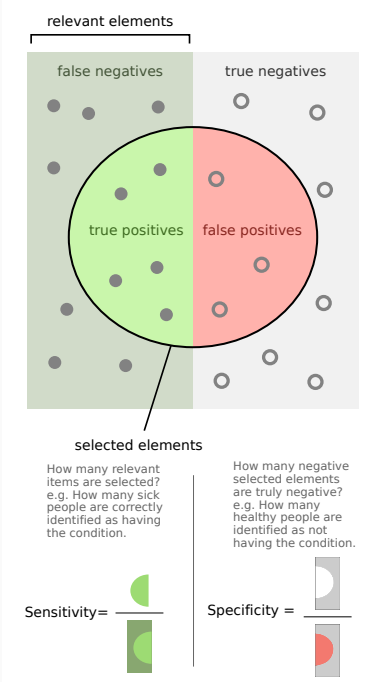
Precision = $\frac{\text{true positives}}{\text{true positives} + \text{false positives}}$

How many relevant items are retrieved?

Recall = $\frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$

Precision und Recall, Walber CC BY-SA 4.0

Specificity, Visualisierung



Precision / Recall tradeoff

- Manche Problemstellungen erfordern entweder hohe Precision oder hohen Recall
  - Jugendschutz Filter: hohe Precision / geringer Recall
  - Diebstahl Erkennung auf Video: geringe Precision / hoher Recall
- Die Erhöhung der Precision verringert den Recall
- Die Erhöhung des Recall verringert den Precision
- Dies nennen wir den *Precision / Recall tradeoff*

Wie entscheidet ein Klassifikator zu welcher Klasse eine Beobachtungseinheit gehört?

Binärer Klassifikator - Definition: Entscheidungsfunktion

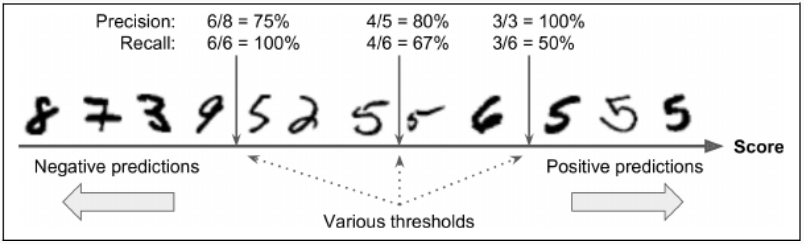
Gegeben sei eine Menge X von Input Daten. Wir nennen eine Funktion $f : X \to \mathbb{R}$ eine *Entscheidungsfunktion*.

Binärer Klassifikator

Gegeben seien eine Entscheidungsfunktion f und ein Schwellwert $T \in \mathbb{R}$. Ein *binärer Klassifikator mit Entscheidungsfunktion* ist eine Funktion $K_{f,T} : X \to \{0, 1\}$

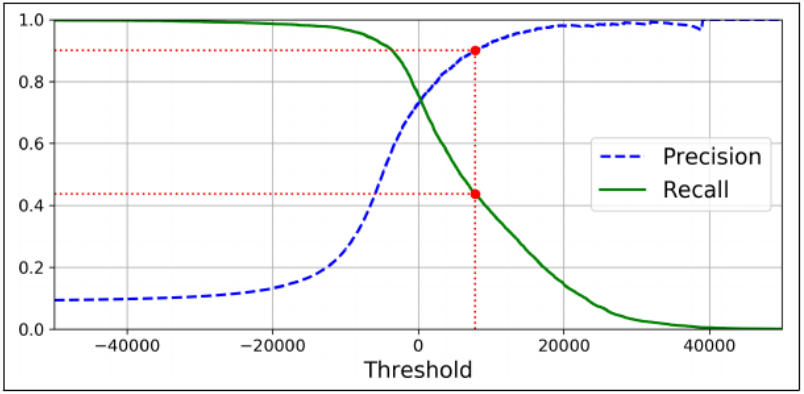
$$K_{f,T}(x) := \begin{cases} 1, \text{ falls } f(x) \geq T \\ 0, \text{ sonst} \end{cases}$$

Precision / Recall tradeoff, Beispiel



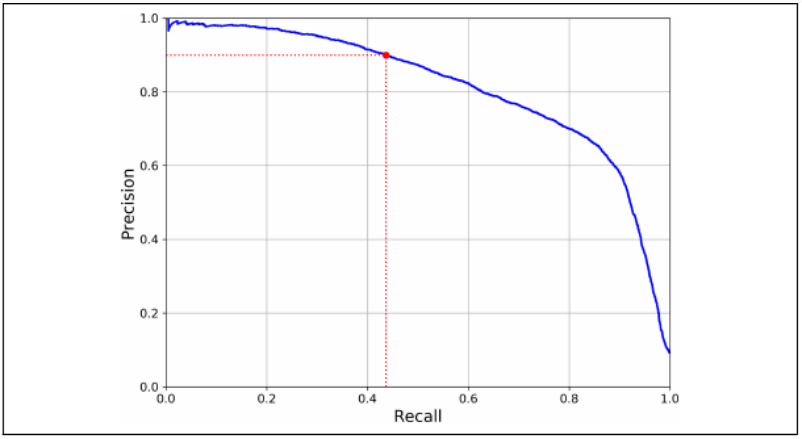
Géron, Aurélien. "Hands-on machine learning with scikit-learn and tensorflow"

Precision / Recall tradeoff, Plot Precision und Recall



Géron, Aurélien. "Hands-on machine learning with scikit-learn and tensorflow"

Precision / Recall tradeoff, Plot Precision gegen Recall

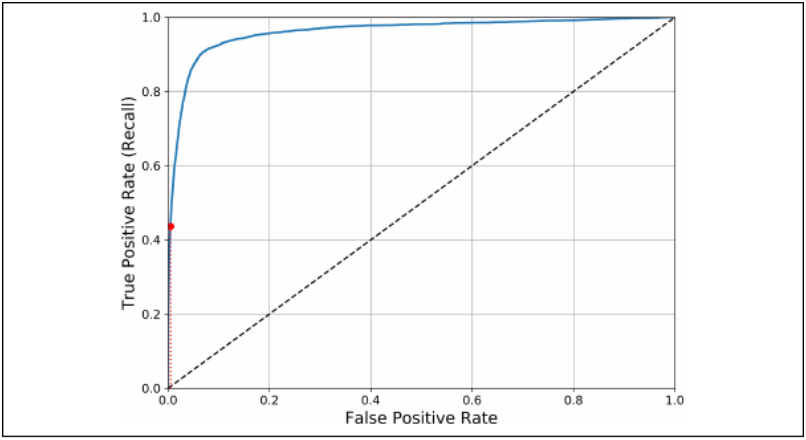


Géron, Aurélien. “Hands-on machine learning with scikit-learn and tensorflow”

ROC Curve / AUC Score

- receiver operating characteristic (ROC) ist ein weiteres Mittel zur Visualisierung der Performance von binären Klassifikatoren
- True Positive Rate (= Recall) gegen 1-Specificity (Falsch Positiv Rate (FPR))
- Die Fläche unter der ROC Kurve wird als *area under the curve (AUC)* bezeichnet

ROC Curve / AUC Score, cont'd



Géron, Aurélien. "Hands-on machine learning with scikit-learn and tensorflow"

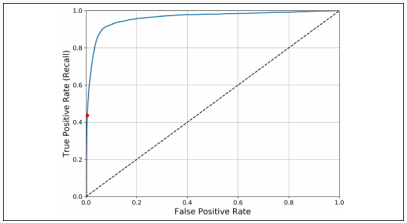
ROC Curve / AUC Score, Metrik

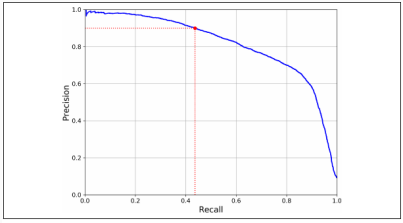
- Ein fehlerfreier binärer Klassifikator hat einen AUC von 1
- Pures Raten hat ein AUC von 0.5
- Werte von ≤ 0.5 sind Anzeichen eines schlechten binären Klassifikators

Wann P-R und wann ROC Kurve?

ROC und Precision / Recall Kurve sind sehr ähnlich. Wann sollten sie was verwenden?

- Verwenden Sie die P-R Kurve wenn die positiv Klasse selten ist
- Verwenden Sie die P-R Kurve wenn mehr an den falsch positiven Klassifikationen interessiert sind
- Sonst verwenden sie die ROC Kurve





Klassifikation mehrerer Klassen

Typen von Klassifikation mehrerer Klassen

- Mehrklassen Klassifikatoren können per Konstruktion mehr als zwei Klassen unterscheiden z.B.
  - Decision Trees
  - Random Forests
  - SGD Klassifikatoren
- Einige Klassifikatoren sind rein binär
  - Support Vector Machine Klassifikator
  - Logistische Regression Klassifikator
- Binäre Klassifikatoren können über “Umwege” mehr als Zwei Klassen unterscheiden

Mehrklassenerkennung über Wahrscheinlichkeiten

Gegeben seien N verschiedene Klassen $K := \{K_1, ..., K_N\}$

Intuition

- Viele Machine Learning Algorithmen berechnen Wahrscheinlichkeiten dass eine Beobachtungseinheit $X = (x_1, ..., x_n)$ zu einer Klasse $y \in K$ gehört
- Die Klassifikation erfolgt dann über die größte Wahrscheinlichkeit

Etwas formaler

- Formaler gesprochen berechnen die Algorithmen die Wahrscheinlichkeiten $P(y|X) = P(y|x_1, ..., x_n)$ mit $y \in K$
- Damit ist die Klassifikation definiert über $\hat{y} := \underset{y \in K}{\text{argmax}} P(y|X)$

Beispiel

```
In [0]: probabilities = [0.0, 0.2, 0.5, 0.3]

In [1]: np.argmax(probabilities)

Out[1]: 2
```

Strategien für binäre Klassifikatoren zur Mehrklassenerkennung

Sei N die Anzahl der zu unterscheidenden Klassen $K_1, ..., K_N$

Generell gibt es zwei Strategien

- one-versus-rest (OvR)
  - Beantwortet Frage ob Beobachtungseinheit zu einer festen Klasse $K_n, n = 1...N$ gehört.
  - Somit muss man N Klassifikatoren trainieren und auswerten
- one-versus-one (OvO)
  - Beantwortet Frage ob Beobachtungseinheit zu einer festen Klasse K_n oder festen Klasse K_m mit $n, m = 1...N$ gehört.
  - Somit müssen $N \cdot (N-1)/2$ Klassifikatoren trainiert werden

Mehrere Klassen mit binären Klassifikatoren

Beispiel: Handschrift Erkennung der Zahlen 0 bis 9

- OvR: 10 Klassifikatoren
- OvO: 45 Klassifikatoren

Die meisten binären Klassifikatoren sollten mit OvR genutzt werden, eine Ausnahme ist die Support Vector Machine.

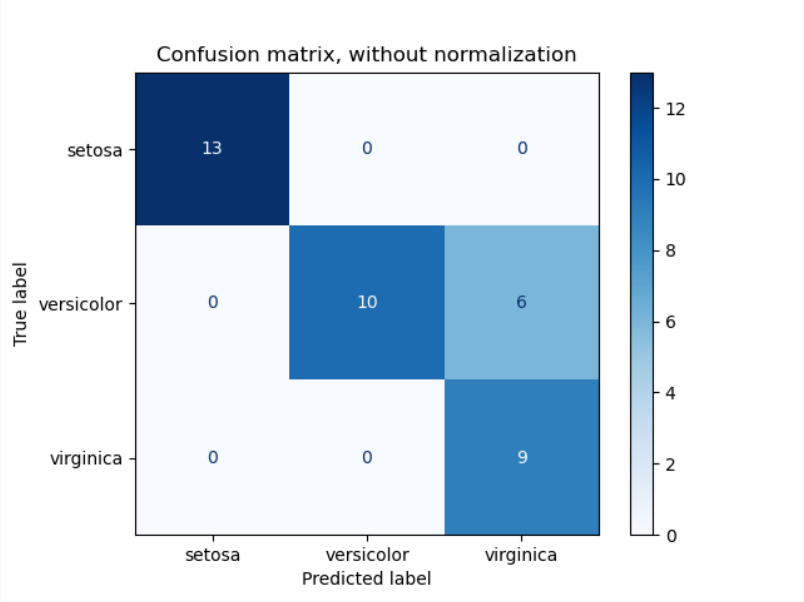
Metrik für mehrklassen Klassifikatoren

Analog zur binären Klassifikation lässt sich auch für mehrere Klassen eine Konfusionsmatrix definieren.

Definition: Mehrklassen Konfusionsmatrix

Gegeben seien $N \in \mathbb{N}$ verschiedene Klassen. Wir definieren *mehrklassen Konfusionsmatrix* $C \in \mathbb{N}^{N \times N}$ mit $C_{i,j}, 1 \leq i, j \leq N$ die Anzahl der Beobachtungseinheit die deren wahre Klasse i ist aber vom Klassifikator als j klassifiziert wurde.

Beispiel: Konfusionsmatrix



scikit-learn Dokumentation, https://scikit-learn.org/stable/auto_examples/model_selection/plot_confusion_matrix.htm

Precision / Recall für mehrere Labels

Precision und Recall sind nur für binäre Klassifikatoren definiert.

Um dies für mehrere Klassen zu generalisieren gibt es mehrere Ansätze:

- Mikro: Summe der echten Positiven, falschen Negativen und falschen Positiven über alle Klassen.
- Makro: Berechnung der Metrik für jedes Label und Ermittlung ihres ungewichteten Mittelwerts.
- Gewichtet: Wie Makro nur, dass diese nach der Häufigkeit des Auftretens einer Klasse gewichtet ins Ergebnis eingehen.

Nutzen Sie “gewichtet” für asymmetrisch verteilte Datensätze, also wenn eine Klasse signifikant weniger auftritt als eine andere Klasse.

Regression

Ausblick Regression

Kommende Vorlesung behandeln wir Regression: