

# PERFORMANZ METRIKEN (bei Klassifizierungsproblemen)

## • THE CONFUSION MATRIX (Wahrheitsmatrix)

		LABEL (Target)	
		1	0
VORHERSAGE	1	TP	FP
	0	FN	TN

TP: Richtig klassifiziert als Positiv

FP: Falsch klassifiziert als Positiv

TN: Richtig klassifiziert als Negativ

FN: Falsch klassifiziert als Negativ

- Pandas's `crosstab()`

- sklearn's `metrics.confusion_matrix`

## • ACCURACY (Correct rate = Erfolgsrate) (Korrektklassifikationsrate)

$$\text{Accuracy} = \frac{\text{Richtig klassifizierte Beispiele}}{\text{Alle Beispiele}} = \frac{TP + TN}{TP + TN + FP + FN}$$

## • Skewed Classes : Daten sind nicht gleichmäßig verteilt

- $y=1$  hat Krankheit (% 1)

- $y=0$  hat Krankheit nicht (% 99)

Modell, das immer 0 vorhersagt : 99% accuracy!

(only 1% Error rate)

KEIN GUTES MAß !

- PRECISION (Positive Predictive Value - PPV)

$$\text{Precision} = \frac{\# \text{ richtig positive Vorhersagen}}{\# \text{ alle positive Vorhersagen}} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Y

	1	0
1	TP	FP
0	FN	TN

- Von allen positiven Vorhersagen, wie viele waren richtig?
- Wahrscheinlichkeit, dass ein positiv klassifiziertes Beispiel auch tatsächlich positiv ist.
- Accuracy of **positive predictions**.

- RECALL (True Positive Rate) (auch Sensitivität)

$$\text{Recall} = \frac{\# \text{ richtig positive Vorhersagen}}{\# \text{ alle positiven Beispiele}} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Y

	1	0
1	TP	FP
0	FN	TN

- Von allen positiven Beispielen, wie viele wurden richtig klassifiziert?
- Wahrscheinlichkeit, dass ein positives Beispiel tatsächlich als solches klassifiziert wird.
- Accuracy on **positive examples**.

BSP.

- $y=1$  hat Krankheit (% 1)
- $y=0$  hat Krankheit nicht (% 99)
- Algorithmus klassifiziert immer mit 0

	1	0
1	0	0
0	1	99

$$\text{Precision} = \frac{0}{0} (\text{NA})$$

$$\text{Recall} = \frac{0}{1}$$

$$\text{Accuracy} = \frac{99}{100}$$

BSP.

		$y=1$	$y=0$
$\hat{y}=1$	$y=1$	80	20
	$y=0$	80	820

$$\text{Precision} =$$

$$\text{Recall} =$$

• Bemerkung:

Wenn  $y=0$  die kleinere Klasse ist, folgende Werte beachten:

- Negative Predictive Value - NPV (Accuracy of negative predictions)

$$\frac{\# \text{ correct negative predictions}}{\# \text{ all negative predictions}}$$

	$y=1$	$y=0$
$\hat{y}=1$	TP	FP
$\hat{y}=0$	FN	TN

- True Negative Rate (Accuracy on negative examples)

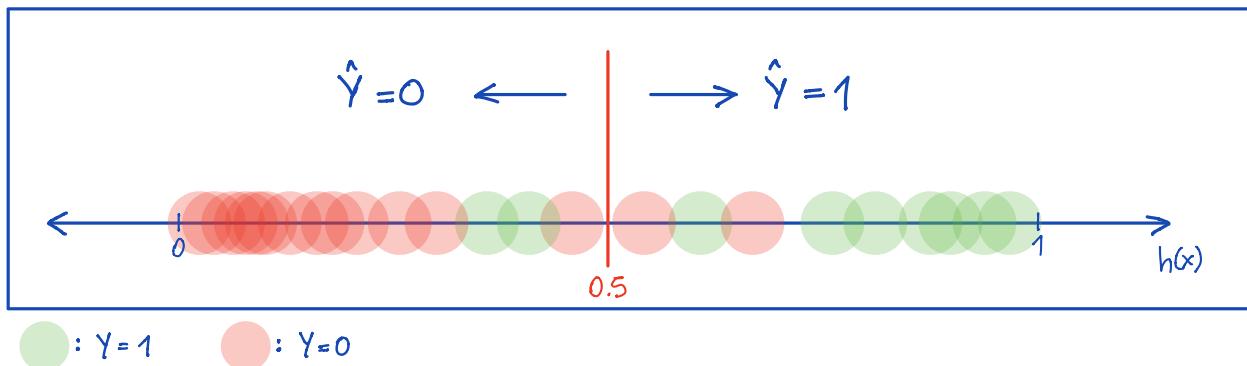
$$\frac{\# \text{ correct negative predictions}}{\# \text{ all negative examples}}$$

	$y=1$	$y=0$
$\hat{y}=1$	TP	FP
$\hat{y}=0$	FN	TN

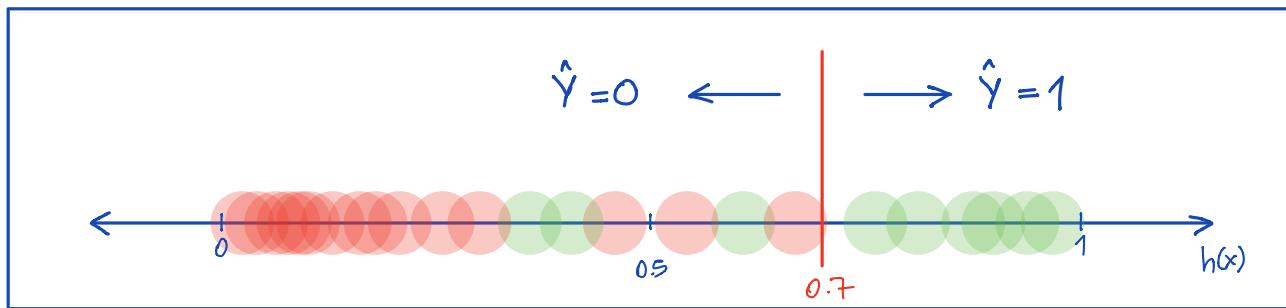
## Precision und Recall Trade-Off

- "Normalfall": Schwellenwert ist 0.5

TP	FP
FN	TN

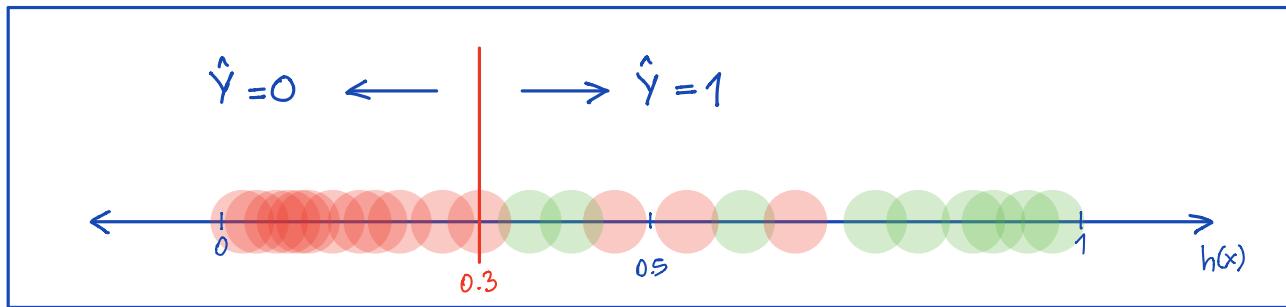


- Schwellenwert ist 0.7 (1, nur wenn ganz sicher!)



- Weniger rote Punkte rechts      P↑
- Mehr grüne Punkte links      R↓

- Schwellenwert ist 0.3 (Keine 1 verfehlten!)



- Mehr rote Punkte rechts      P↓
- Weniger grüne Punkte links      R↑

## • $F_1$ -Score (Harmonisches Mittel) (auch F-Maß)

Fasst Precision (P) und Recall (R) in einem Maß zusammen.

$$F_1\text{-Score} = \frac{2}{\frac{1}{P} + \frac{1}{R}} = 2 \cdot \frac{P \cdot R}{P + R}$$

Harmonic Mean

BSP.

Modell Nr.	Precision (P)	Recall (R)	Durchschnitt $(P+R)/2$	$F_1$ -Score
1	0.5	0.4	0.45	0.444
2	0.7	0.3	0.5	0.42
3	0.1	1.0	0.55	0.18
4	1	1	1	1 (max)
5	a	a	a	a

Je höher  
desto besser!

- Der gewöhnliche Durchschnitt  $(P+R)/2$  ist kein gutes Maß.
- Bei dem  $F_1$ -Score haben kleinere Werte größeren Einfluss auf das Ergebnis. Das  $F_1$ -Score wird nur dann hoch sein, wenn P und R beide hoch sind.

BSP

## Harmonisches Mittel - Intuition

- 1 km Strecke mit 60 km/h zurückgelegt.
- 1 km Strecke mit 30 km/h zurückgelegt.

Was ist die durchschnittliche Geschwindigkeit auf der gesamten 2km-Strecke?

- o Antwort:  $\frac{2}{\frac{1}{60} + \frac{1}{30}} = \frac{2}{\frac{3}{60}} = 2 \cdot \frac{60}{3} = 40 \text{ km/h}$
- o Warum?

60 km/h bedeutet 1 km in  $\frac{1}{60}$  Std.

30 km/h bedeutet 1 km in  $\frac{1}{30}$  Std.

---

$\Rightarrow$  Insgesamt 2 km in  $\frac{1}{60} + \frac{1}{30}$  Std.

$\Rightarrow$  1 km in  $\frac{\frac{1}{60} + \frac{1}{30}}{2}$  Std.

$\Rightarrow$   $\frac{2}{\frac{1}{60} + \frac{1}{30}}$  km in einer Stunde.

- Zusammenfassend :
  - Neben loss und accuracy (auf Trainings- und Validierungsmergen) auch Precision, Recall und  $F_1$ -Score während des Trainings im Auge behalten.  
→ gibt ein vollständigeres und wahrheitsgetreueres Bild von der Performance des Modells.

- Bei unangemessenen Datensätzen
    - Datenaugmentierung
    - Oversampling und Undersampling
- } Nachlesen !