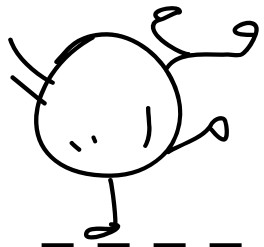


Evaluierung

Vorlesung 186.844

09.01.2016

TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

Überblick

I. No Free Lunch Theorem

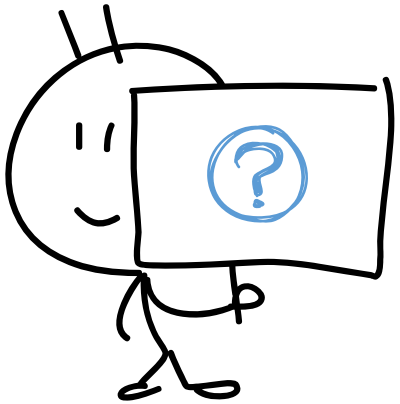
II. Quantitative Evaluierung

I. No Free Lunch Theorem

Einleitung

… eine Frage die sich wahrscheinlich jeder von euch ein oder mehrmals gestellt hat:

Welcher Klassifikator, Trainingsalgorithmus oder welche Methode ist **DIE BESTE**?



Einleitung

Methode **bevorzugt wegen**:

- ihrer geringen Komplexität oder
- ihrer Fähigkeit Vorwissen (Priors) zu berücksichtigen

ABER: Es gibt Mustererkennungsprobleme bei denen die obigen oder ähnliche Eigenschaften **nicht relevant** sind oder gleich für alle zu vergleichenden Methoden sind.

No Free Lunch Theorem

Für das **allgemeine Problem** der Mustererkennung, ohne Annahmen über Art der Muster, Verteilungen, Vorwissen, etc., beantwortet das „*No Free Lunch Theorem*“ folgende Fragen:

Gibt es irgendwelche Gründe einen Klassifikator oder Trainings-algorithmus einem anderen vorzuziehen?

Gibt es einen Algorithmus oder eine Methode die generell besser ist als der Zufall?

Nein!

No Free Lunch Theorem

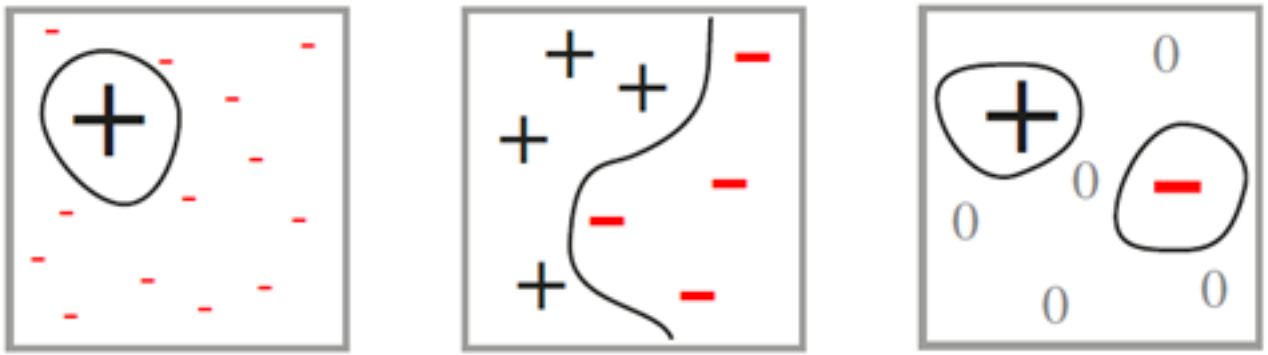
„No Free Lunch Theorem“ anders formuliert: Es gibt **KEINE kontextunabhängige** oder **anwendungsunabhängige** Gründe einen bestimmten Klassifikator, Trainingsalgorithmus oder eine Methode zu bevorzugen.

Dieses Theorem erinnert uns beim Design eines Musterer-kennungssystems auf die wesentlichen Dinge zu achten:

- ✓Vorwissen
- ✓Datenverteilung
- ✓Größe des Trainingsdatensatzes
- ✓Kosten oder Gewinn der jeweiligen Entscheidung

No Free Lunch Theorem

… mögliche Musterkennungssysteme



Legende (Problemräume)

\+ … Generalisierungsfähigkeit höher als Durchschnitt

\- … Generalisierungsfähigkeit niedriger als Durchschnitt

0 … durchschnittliche Generalisierugnsfähigkeit

Größe der Symbole … beschreibt die Größe der Abweichung vom Durchschnitt

No Free Lunch Theorem

… unmögliche Musterkennungssysteme

Legende (Problemräume)

\+ … Generalisierungsfähigkeit höher als Durchschnitt

\- … Generalisierungsfähigkeit niedriger als Durchschnitt

0 … durchschnittliche Generalisierugnsfähigkeit

Größe der Symbole … beschreibt die Größe der Abweichung vom Durchschnitt

Schlussfolgerung

Es gibt **kein** universell einsetzbares, bestes Mustererkennungsssystem.

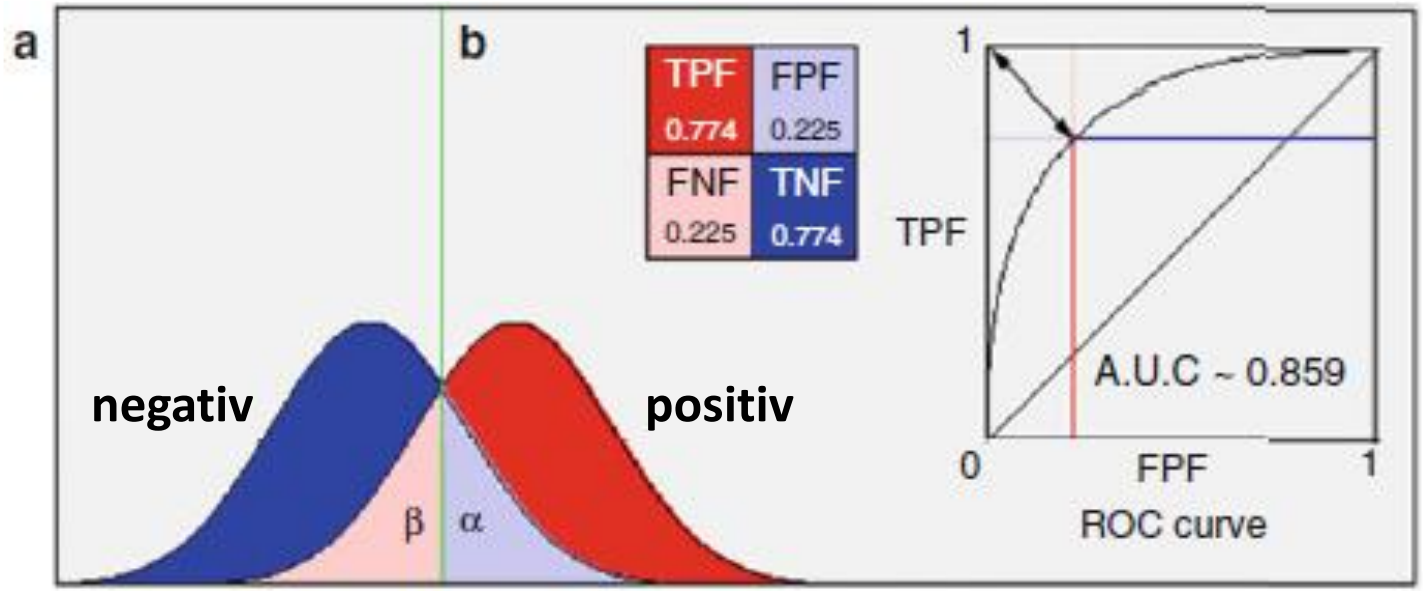
Was wir tun können:

Evaluieren wie gut ein Mustererkennungssystem für ein bestimmtes Musterkennungsproblem geeignet ist.

II. Quantitative Evaluierung

Für zwei Klassen …

- Klasse w_1: negativ (links von der Entscheidungsgrenze)
- Klasse w_2: positive (rechts von der Entscheidungsgrenze)

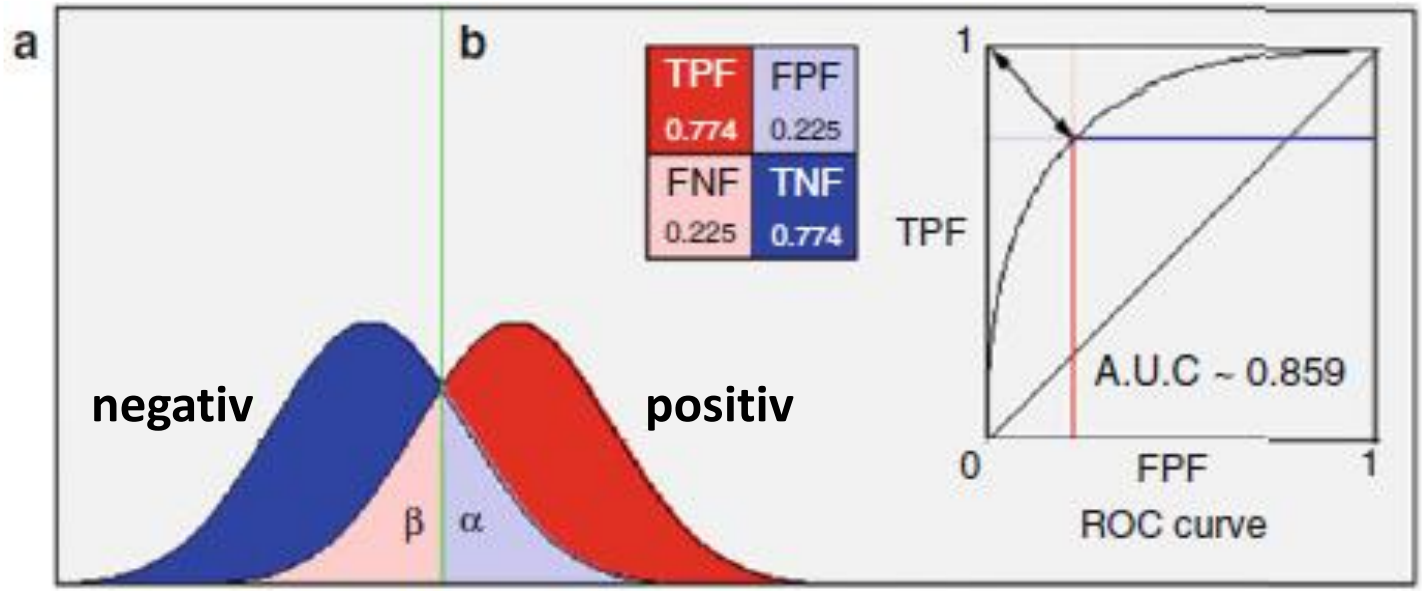


[Quelle: Dougherty 2013]

Für zwei Klassen …

Mögliche Ergebnisse:

- **TP** = True Positive
- **FN** = False Negative
- **TN** = True Negativ
- **FP** = False Positiv



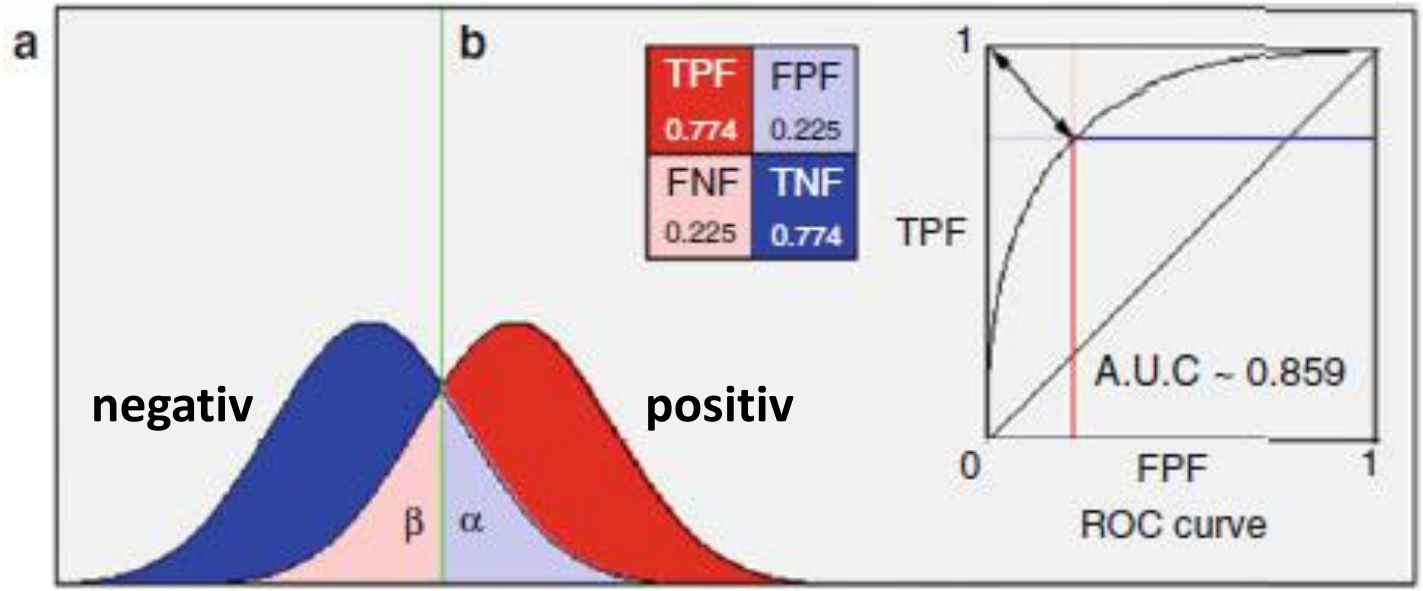
[Quelle: Dougherty 2013]

WH: ROC-Kurve

ROC = Receiver Operating Characteristic

TPF = True Positive Fraction

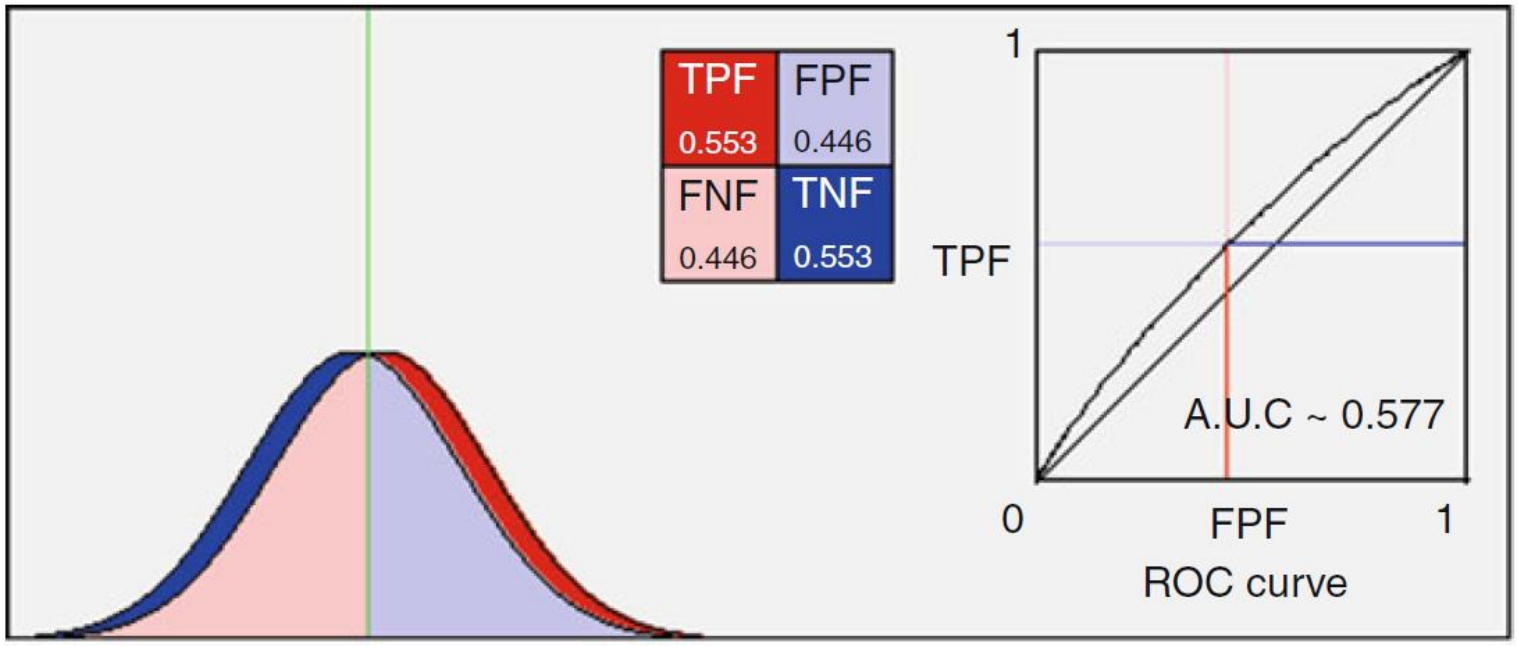
FPF = False Positive Fraction



[Quelle: Dougherty 2013]

Area Under Curve

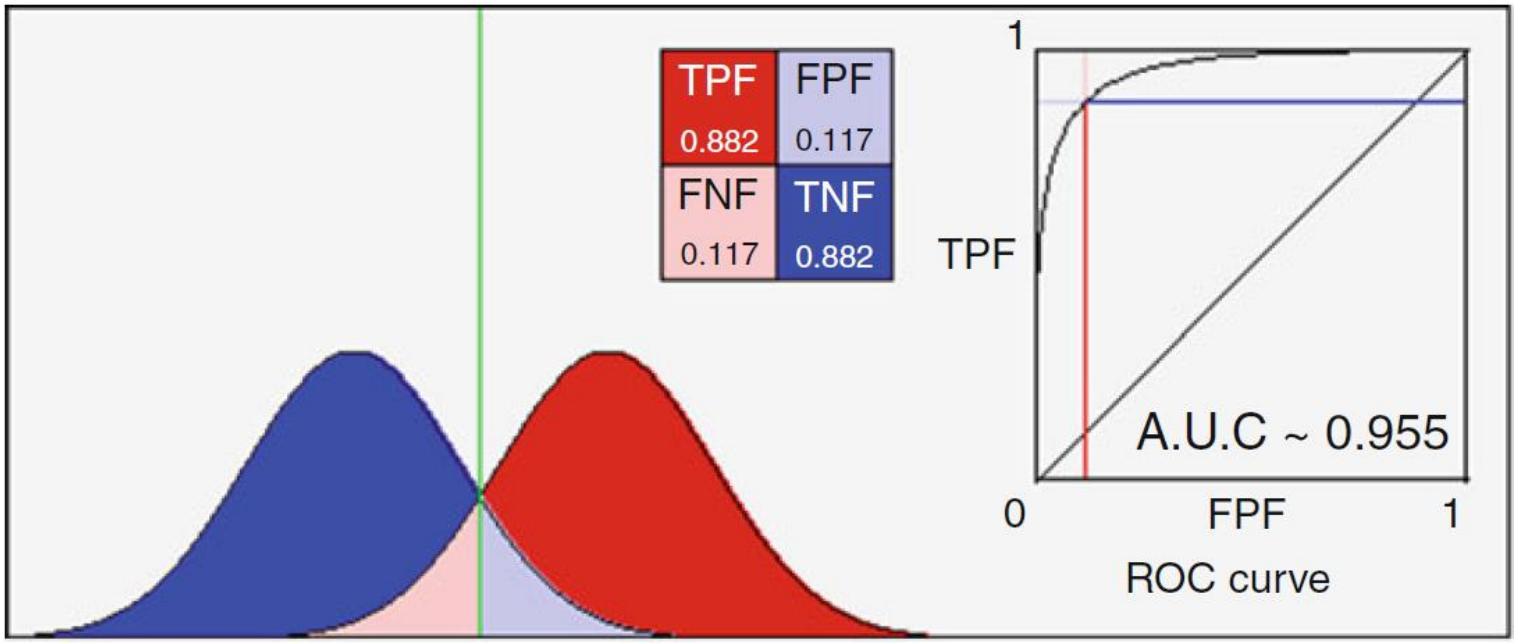
- AUC ≈ 0,6→ (fast) vollständiger Überlapp der beiden Klassen (Kurven)



[Quelle: Dougherty 2013]

Area Under Curve

- AUC nahe an 1 → Klassen überlappen kaum



[Quelle: Dougherty 2013]

Confusion-Matrix

- für zwei Klassen

Predicted			
Actual	Positive	Negative	Total
Positive	TP	FN	p
Negative	FP	TN	n
Total	p'	n'	N

Ground Truth

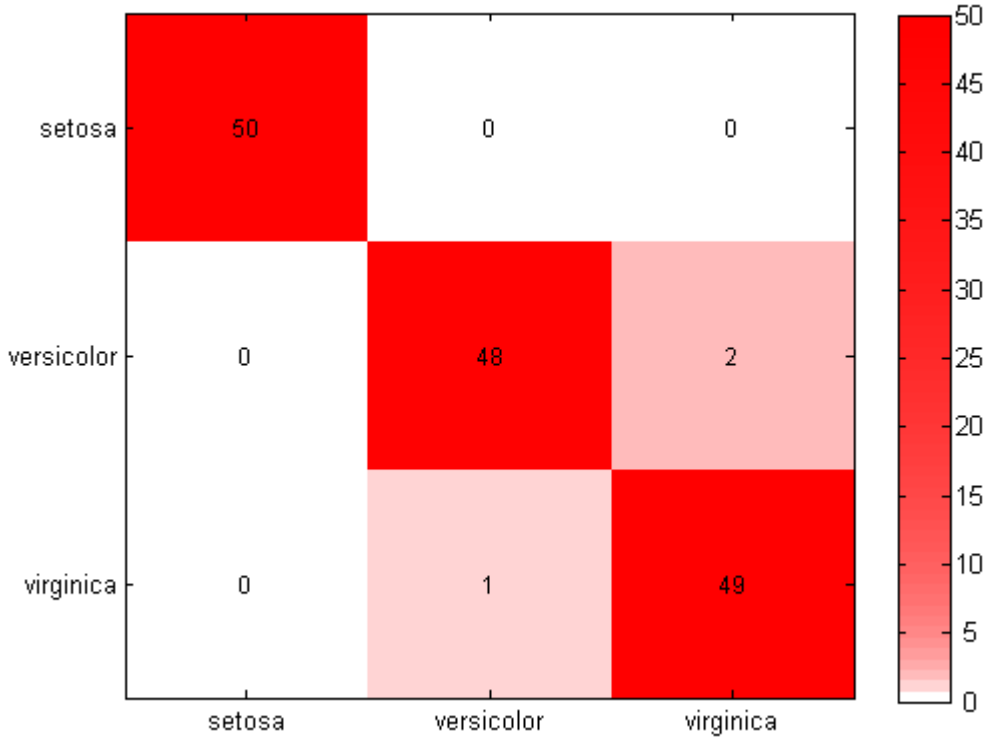
Anzahl der positiv klassifizierten Elemente

Anzahl der negativ klassifizierten Elemente

[Quelle: Dougherty 2013]

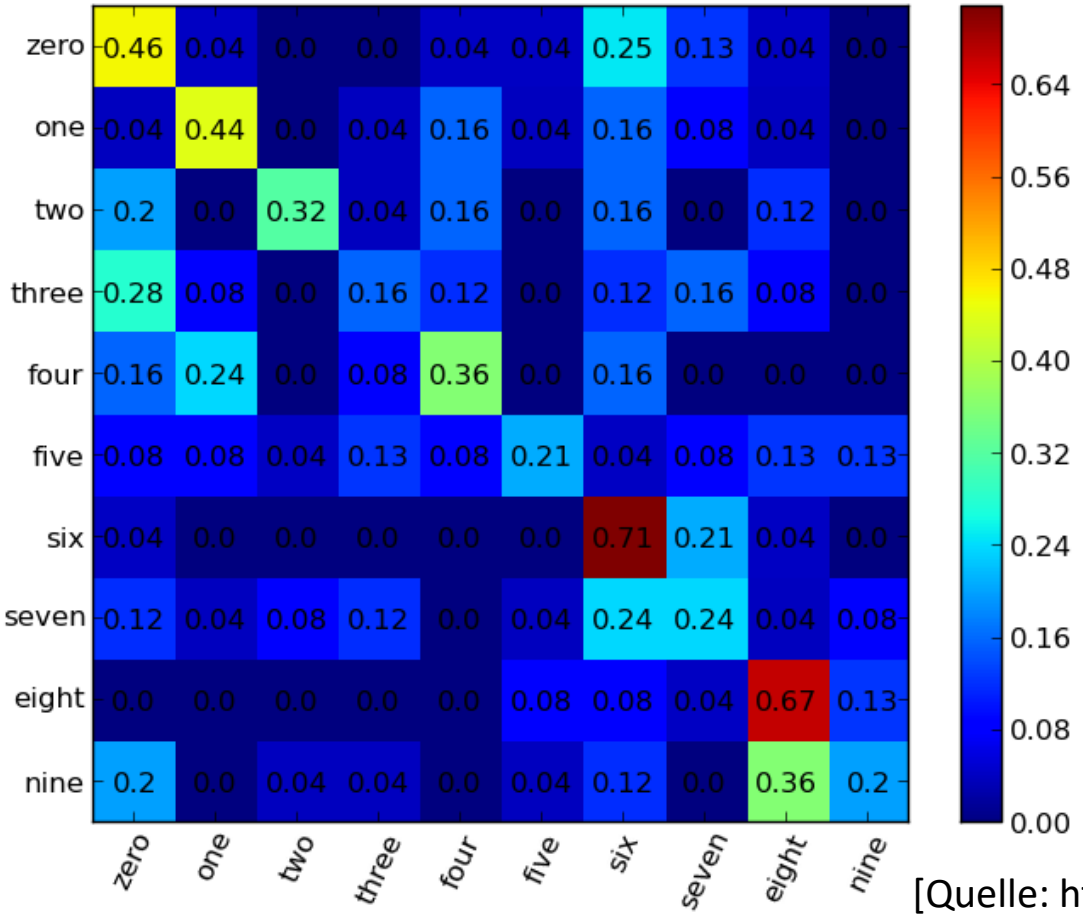
Confusion-Matrix

- für mehr als zwei Klassen
- Idealfall: alle Werte außer jene in der Diagonale sind „0“



Confusion-Matrix

- für mehr als zwei Klassen



[Quelle: http://stackoverflow.com]

Precision

Die **Precision** P_i gibt an wieviel Prozent der Muster die als Klasse i klassifiziert wurden auch tatsächlich der Klasse i angehören (Ground-Truth). Die „Precision“ der 1. Klasse eines binären Klassifikations-problems wird folgendermaßen bestimmt:

$$P_1 = \frac{C(1,1)}{C(1,1)+C(2,1)} \text{ wobei}$$

C die Confusion-Matrix ist und ihre Elemente $C(i,j)$ entsprechen der Anzahl an Mustern welche das Klassenlabel i haben (Ground-Truth) und als j klassifiziert wurden.

Anmerkung: Falls die 1. Klasse die „positive“ bzw. die „negative“ Klasse wäre, könnte man die „Precision“ auch wie folgt anschreiben: $\frac{TP}{TP+FP} = \frac{TN}{TN+FN}$

Recall

Recall R_i gibt an wieviel Prozent der Klasse i auch als Klasse i klassifiziert wurden. In einem binären Klassifikationsproblem wird „Recall“ folgendermaßen bestimmt:

$$R_1 = \frac{C(1,1)}{C(1,1)+C(1,2)} \text{ wobei}$$

C die Confusion-Matrix ist und ihre Elemente $C(i,j)$ entsprechen der Anzahl an Mustern welche das Klassenlabel i haben (Ground-Truth) und als j klassifiziert wurden.

Anmerkung: Falls die 1. Klasse die „positive“ bzw. die „negative“ Klasse wäre, könnte man „Recall“ auch wie folgt anschreiben: $\frac{TP}{TP+FN} = \frac{TN}{TN+FP}$

Overall Accuracy

Overall Accuracy A gibt (klassenübergreifend) an wieviel Prozent der Muster richtig klassifiziert wurden. Für M Klassen und N Muster berechnet sich die „Overall Accurarcy“ wie folgt:

$$A = \frac{1}{N}\sum_{i=1}^{M} C(i,i) \text{ wobei}$$

C die Confusion-Matrix ist und die Elemente $C(i,i)$ liegen auf ihrer Diagonale.

Anmerkung: Der Rest auf 100% entspricht der Fehlklassifikation in Prozent.