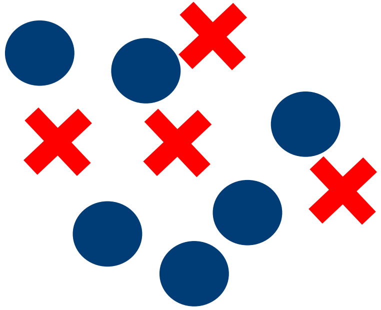
Entropie

Ein Maß für den Informationsgehalt

INNOVET KI B³
Künstliche Intelligenz in die berufliche Bildung bringen

INNOVET
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung
bibb Bundesinstitut für Berufsbildung
Gefördert als InnoVET-Projekt aus Mitteln des Bundesministeriums für Bildung und Forschung.

Rückblick: Gini-Index als Maß für „Unordnung“



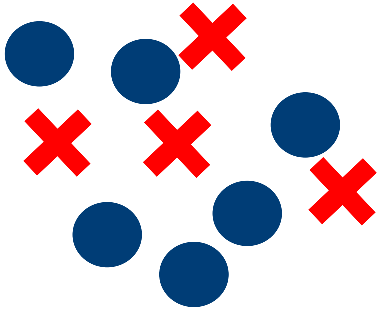
Verteilung der Klassen:

x 0.4 ● 0.6

Gezogen:		Klassif. als:		Wahrscheinlichkeit:	
x	0.4	x	0.4	0.4 * 0.4 = 0.16	
x	0.4	●	0.6	0.4 * 0.6 = 0.24	falsch
●	0.6	x	0.4	0.6 * 0.4 = 0.24	
●	0.6	●	0.6	0.6 * 0.6 = 0.36	

Gini-Index: 0.24 + 0.24 = 0.48

Ziel: ein weiteres Maß für Unordnung/Unsicherheit



Idee:

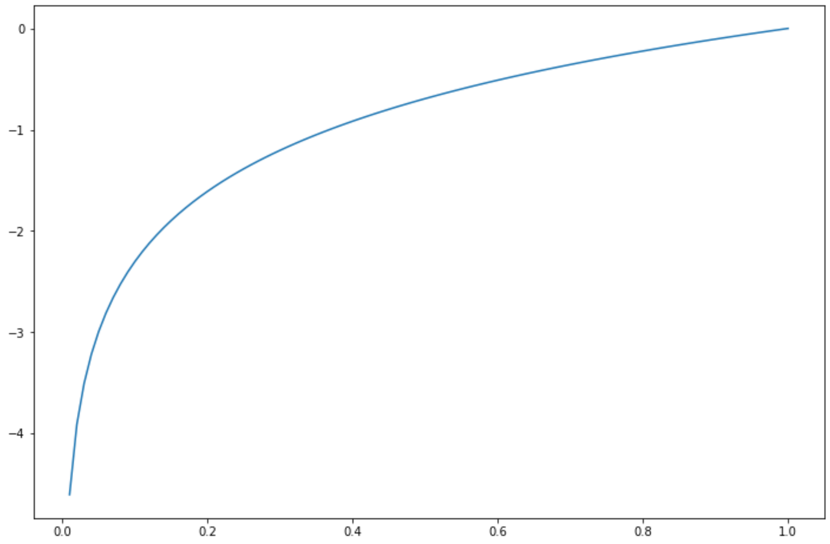
- Wahrscheinlichkeitsverteilung in “Unsicherheitsmaß“ umwandeln
- Klassenwahrscheinlichkeiten (relative Häufigkeiten) immer zwischen 0 und 1
- Gut: hohe Klassenwahrscheinlichkeiten
- Gewünscht: eine Transformation, die
  - hohe Klassenwahrscheinlichkeiten zu Werten um die 0 transformiert
  - niedrige Klassenwahrscheinlichkeiten zu hohen Werten transformiert

Logarithmus

Hier: natürlicher Logarithmus, oft bezeichnet als ln

- Wandelt Werte zwischen 0 und 1 in Werte zwischen –Unendlich und 1 um
- Werte nahe Null ergeben sehr kleine negative Zahlen
- Werte nahe eins ergeben Zahlen nahe Null

⇒ Vorzeichen umdrehen!



Der Überraschungswert einzelner Ereignisse

- Engl.: „Surprisal“
- Auch: Informationsgehalt
- Ereignis mit Wahrscheinlichkeit P
- Überraschungswert des Ereignisses
  -ln(P)

  (es variiert, zu welcher Basis der Logarithmus genommen wird, hier: Basis e, also natürlicher Logarithmus)

- Ereignis: Eisstand geschlossen
- P(geschlossen) = 20/34 = 0.588 (58.5%)
  -ln(20/34) = 0.531

- Ereignis: Eisstand offen
- P(offen) = 14/34 = 0.412
  -ln(14/34) = 0.887

Der Überraschungswert eines Ereignisses ist der negative Logarithmus der Ereigniswahrscheinlichkeit.

Im Bereich Maschinelles Lernen wird meist der natürliche Logarithmus verwendet.

Der Überraschungswert extremerer Ereignisse

- Schale mit Losen, davon 99% Nieten
- Also Wahrscheinlichkeit, eine Niete zu ziehen: 0.99
- -ln(0.99) = 0.01
- Niete: wenig Überraschung

- Schale mit Losen, davon 1% Gewinne
- Also Wahrscheinlichkeit, zu gewinnen: 0.01
- -ln(0.01) = 4.6
- Gewinn: große Überraschung

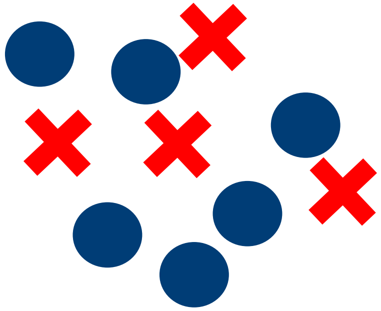
Wie kombiniert man dies zu einem einzigen Maß für die ganze Schale?

Mittelwert ½ * (0.01 + 4.6) = 2.35 ist “unfair“: die Niete ist viel wahrscheinlicher!

⇒ Wir **gewichten** mit den Ereigniswahrscheinlichkeiten!

⇒ 0.99 * 0.01 + 0.01 * 4.6 = 0.0559

(Shannon-)Entropie



	Klassen-wahrsch.	mal Überr.-wert	
x	0.4	* - ln(0.4)	= - 0.4 * -0.9163 = + 0.3665
●	0.6	* - ln(0.6)	= - 0.6 * -0.5108 = + 0.3065
			+ 0.6730

Entropie als Maß für Informationsgehalt (Claude Shannon, 1948)

Die Entropie im Bereich der Informationstheorie wurde von Claude Shannon eingeführt.

Sie gibt ursprünglich an, wie stark ein zufälliges Signal maximal komprimiert werden kann.

Im Bereich Machine Learning wird die Entropie als Maß für Unsicherheit (Überraschung, Unordnung) verwendet.

Die Entropie ist die gewichtete Summe der Überraschungswerte aller Ereignisse.

Je höher sie ist, desto unsicherer/unordentlicher ist die Wahrscheinlichkeitsverteilung der Ereignisse.

Question Set: Entropie und Wahrscheinlichkeitsverteilung

Sort the paragraphs: Sortieren nach Entropie

INNOVET
KI B³
Künstliche Intelligenz in die berufliche Bildung bringen

Dr. Antje Schweitzer

Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung

Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung
Institut für Software Engineering

IHK Industrie- und Handelskammer Reutlingen
Reutlingen | Tübingen | Zollernalb

IHK Region Stuttgart

IHK Industrie- und Handelskammer Karlsruhe

LMU Ludwig-Maximilians-Universität München

INNOVET

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

bibb Bundesinstitut für Berufsbildung

Gefördert als InnoVET-Projekt aus Mitteln des Bundesministeriums für Bildung und Forschung.

Lizenzbestimmungen

“Entropie" von Antje Schweitzer, KI B³ / Uni Stuttgart

Das Werk - mit Ausnahme der folgenden Elemente:

- Logos der Verbundpartner und des Förderprogramms

ist lizenziert unter: