Entropie und Information Gain

Christian Herta

Juni, 2014

Lernziele

- Wahrscheinlichkeit und Information
- Entropie
- *Conditional Entropy*
- *Information Gain*

Outline

Diskrete Ereignisse

Gegeben seien n exklusive Ereignisse (Nachrichten) i, die mit der Wahrscheinlichkeit $p(i)$ (kurz p_i) auftreten (empfangen werden):

$$\sum_{i}^{n} p_i = 1$$

Informationsgehalt (Überraschungswert)

Der Informationsgehalt von i hängt mit seiner Auftrittswahrscheinlichkeit p_i zusammen. Dabei gilt:

- Wenn das Ereignis mit Sicherheit ($p_i = 1$) auftritt, ist der Informationsgehalt 0.
- Je unwahrscheinlicher (überraschender) das Ereignis ist, desdo höher ist der Informationsgehalt.
- Der Informationsgehalt zweier unabhängiger Ereignisse soll additiv sein.

Dies ist mit folgender Definition des Informationsgehaltes I_i für das Ereignis i gegeben:

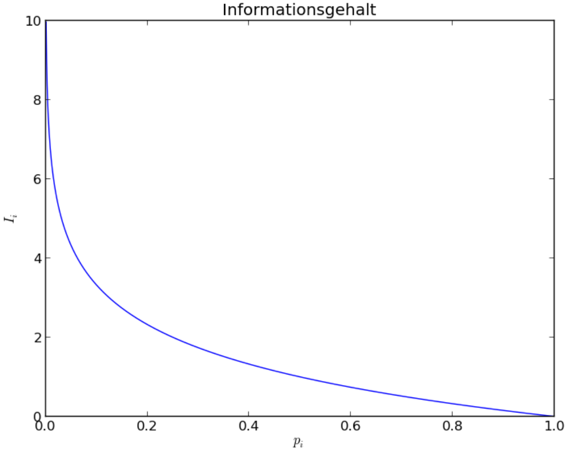
$$I_i = -\log_2 p_i$$

Einheit: bit

Informationsgehalt: Beispiele

- Münzwurf
  $p_{Kopf} = p_{Zahl} = 0.5 \Rightarrow I_{Kopf} = I_{Zahl} = 1 bit$
- Münzwurf: 2 x Zahl (oder Kopf)
  $p_{2xZahl} = 1/4 \Rightarrow I_{2xZahl} = \log(4) bit = 2 bit$
- Würfel
  $p_1 = \cdots = p_6 = 1/6 \Rightarrow I_1 = I_6 = \log_2(6) bit \approx 2.585 bit$
- 6 Richtige im Lotto (ohne Zusatzzahl)
  $p_{6\,Richtige} = \log_2(13983816) bit \approx 23.74 bit$

Informationsgehalt



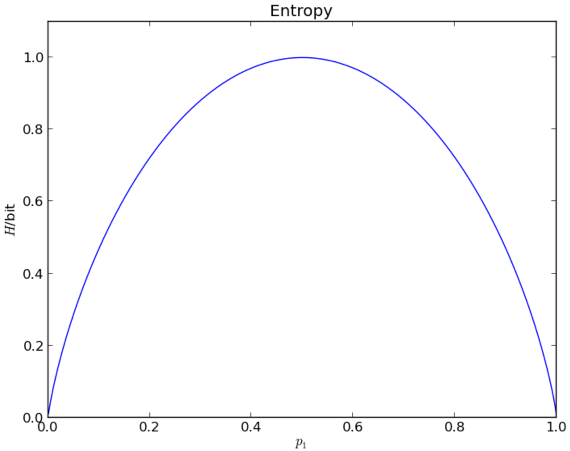
Entropy

Die Entropie ist der Erwartungswert des Informationsgehaltes:

$$H \equiv \mathbb{E}(I) = \sum_i p_i \cdot I_i = -\sum_i p_i \log_2 p_i$$

Entropie

Entropie bei zwei möglichen Ereignissen $\in 0, 1$



Outline

Bedingte Information *conditional information*

Der Informationsgehalt des Ergeignisses $y \in \mathcal{Y}$ bei bereits bekanntem (festen) Wert für $x \in \mathcal{X}$ ist:

$$I(y|x) \equiv -\log p(y|x)$$

Die Entropie bezüglich Y für festes $x \in \mathcal{X}$ ist somit:

$$H(Y|x) \equiv \sum_{y \in \mathcal{Y}} p(y|x) I(y|x) = -\sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x)$$

Bedingte Entropie *conditional entropy*

$$H(Y|X) \equiv \mathbb{E}_{\mathcal{X}}(H(Y|x \in \mathcal{X})) = \sum_{x \in \mathcal{X}} p(x) H(Y|x)$$

$$H(Y|X) = -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x) p(y|x) \log p(y|x)$$

Information Gain

Die Kenntniss, dass die Variablen X den konkreten Wert x annimmt, reduziert die Entropy von $H(Y)$ auf $H(Y|x)$.
Die "Information" (Verringerung der Unsicherheit) die hierdurch bezüglich Y gewonnen wurde, ist somit

$$I(Y; X = x) = H(Y) - H(Y|x)$$

Der durchschnittliche Informationsgewinn (*expected information gain* oder *mutual information*) bezüglich Y, der durch die Kenntniss von X erhalten wird, ist folglich

$$I(Y; X) = H(Y) - H(Y|X)$$

Literaturangaben

- T. Cover, J. Thomas, “Elements of Information Theory”, Wiley, 1991