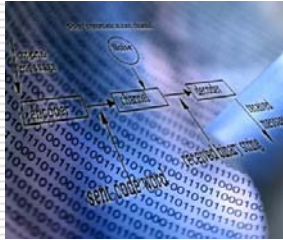


Kapitel 4: Bedingte Entropie



Beiweis: Entropiegrenzen **ETH**

- Zusammenfassung: Es gilt

$$0 \leq H(X) \leq \log_2 |\mathcal{X}|$$

1. Aufbau: Umformung in
2. **Jensen-Ungleichung** für Erwartungswerte
3. Definition von Konvexität/Konkavität (x'')
4. Tangente an konkave Funktion immer oberhalb oder unterhalb des Graphen
5. Entropie ist streng konkav
6. Erwartungswert linearisieren und Tangenteigenschaft nutzen
7. -> Jensen-Ungleichung bewiesen

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

2

Beweis **ETH**

- Beweis: Wir verwenden die Methode der Lagrange-Multiplikatoren und stellen folgende Bedingung auf:

$$E = -\sum_{i=1}^N p_X(x_i) \log_2 p_X(x_i) \rightarrow \min \quad \left| \sum_{i=1}^N p_X(x_i) = 1 \right.$$

- Mit Hilfe des Lagrange-Multiplikators λ erhalten wir

$$E = -\sum_{i=1}^N p_X(x_i) \log_2 p_X(x_i) + \lambda \left(1 - \sum_{i=1}^N p_X(x_i) \right)$$

$$\rightarrow \nabla E = 0$$

$$\rightarrow \frac{\partial E}{\partial p_X(x_i)} = -\log_2 p_X(x_i) - \log e - \lambda$$

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

3

Beweis **ETH**

$$\rightarrow \log p_X(x_i) = -\log e + \lambda \text{ gilt für alle } p_X(x_i)$$

$$\rightarrow p_X(x_i) = p_X(x_j) \quad \forall i, j: 1 \dots N$$

$$\text{da } \sum_{i=1}^N p_X(x_i) = 1 \rightarrow p_X(x_i) = \frac{1}{N}$$

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

4

Bedingte Entropie **ETH**

- Das vorherige Theorem kann durch mehrfache Anwendung direkt verallgemeinert werden

$$H(X_1 X_2 \dots X_N) \leq \sum_{i=1}^N H(X_i)$$

- Ebenso kann die bedingt Entropie definiert werden
- **Definition:** Die bedingte Entropie von X , gegeben Y , ist

$$H(X|Y) = H(XY) - H(Y)$$



$H(X|Y)$ ist also die restliche Unsicherheit über X , wenn Y bekannt ist

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

5

Bedingte Entropie **ETH**

- **Definition:** Die **gegenseitige Information**, die X über Y gibt (sowie symmetrisch Y über X), ist

$$I(X;Y) := H(X) + H(Y) - H(XY)$$

- $I(X;Y)$ ist die **Reduktion** der Unsicherheit über X , wenn man Y erfährt

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

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

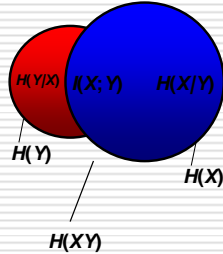
Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

6

Bedingte Entropie

ETH



Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

7

Bedingte Entropie

ETH

- **Theorem:** Es gilt:

$$0 \leq H(X|Y) \leq H(X)$$
- Mit Gleichheit links genau dann, wenn X durch Y vollständig bestimmt ist
- Die rechte Ungleichheit ist äquivalent zu

$$I(X;Y) \geq 0$$
- Also, mit Gleichheit rechts genau dann, wenn X und Y statistisch unabhängig sind
- Alle Definitionen und Aussagen gelten auch für Listen von Zufallsvariablen, wobei z.B.

$$H(RS|TUV) = H(RSTUV) - H(TUV)$$

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

8

Bedingte Entropie

ETH

- Durch wiederholte Anwendung der Definition erhalten wir die **Kettenregel** für Entropien:

$$H(X_1 \dots X_N) = \sum_{i=1}^N H(X_i | X_1 \dots X_{i-1})$$

- Oder auch

$$H(X_1 \dots X_N | Y) = \sum_{i=1}^N H(X_i | X_1 \dots X_{i-1} Y)$$



Man vergleiche dies mit den Formeln für die bedingten Wahrscheinlichkeiten zu Beginn von Modul 2 und wende den Logarithmus an.

$$P(X_1 \dots X_N) = \prod_{i=1}^N P(X_i | X_1 \dots X_{i-1})$$

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

9

Reihenfolge der Variablen

ETH

- Wichtig ist, zu verstehen, dass die Reihenfolge, in der die Zufallsvariablen abgespalten werden, **egal** ist!

$$H(XYZ) = H(X) + H(Y|X) + H(Z|XY)$$

$$H(XYZ) = H(X) + H(Z|X) + H(Y|XZ)$$

$$H(XYZ) = H(Y) + H(X|Y) + H(Z|XY)$$

$$H(XYZ) = H(Y) + H(Z|Y) + H(X|YZ)$$

$$H(XYZ) = H(Z) + H(X|Z) + H(Y|XZ)$$

$$H(XYZ) = H(Z) + H(Y|Z) + H(X|YZ)$$

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

10

Alternative Herleitung

ETH

- Die bedingte Entropie kann auch wie folgt hergeleitet werden
- Für ein Ereignis A mit Wahrscheinlichkeit $P(A)$ können wir $H(X|A)$ als die Entropie der bedingten Verteilung $p_{X|A}$ definieren

$$H(X|A) = - \sum_x p_{X|A}(x) \log p_{X|A}(x)$$

- Seien X und Y zwei Zufallsvariablen. Die **bedingte** Entropie $H(X|Y=y)$ ist

$$H(X|Y=y) = - \sum_x p_{X|Y}(x, y) \log p_{X|Y}(x, y)$$

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

11

Alternative Herleitung

ETH

- $H(X|Y)$ ergibt sich durch gewichtete Mittelwertbildung (Erwartungswert), also

$$\begin{aligned} H(X|Y) &= \sum_y H(X|Y=y) p_Y(y) \\ &= E[-\log p_{X|Y}(X, Y)] \end{aligned}$$

- Die Ungleichung $H(Y|X) \leq H(Y)$ bedeutet, dass zusätzliche Information die Unsicherheit **niemals** erhöhen kann



Dennoch kann die Entropie $H(X|Y=y)$ durchaus lokal grösser werden als $H(X)$ (siehe Beispiel)

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

12

Bedingte Entropie (1)

ETH

- Zwei Zufallsvariablen haben folgende Verteilungen:

$y \backslash x$	1	2	3	4
1	1/8	1/16	1/32	1/32
2	1/16	1/8	1/32	1/32
3	1/16	1/16	1/16	1/16
4	1/4	0	0	0

$$P_X(x_i) = \sum_{j=1}^4 P_{XY}(x_i, y_j) = \frac{x_i}{2} \quad \begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline & 1/2 & 1/4 & 1/8 & 1/8 \end{array}$$

ebenso

$$P_Y(y_j) = \frac{y_j}{4} \quad \begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline & 1/4 & 1/4 & 1/4 & 1/4 \end{array}$$

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

13

Bedingte Entropie (2)

ETH

$$\begin{array}{c|cccc} x_i & 1 & 2 & 3 & 4 \\ \hline & 1/2 & 1/4 & 1/8 & 1/8 \end{array}, \quad \begin{array}{c|cccc} y_j & 1 & 2 & 3 & 4 \\ \hline & 1/4 & 1/4 & 1/4 & 1/4 \end{array}$$

daraus folgt:

$$H(X) = \frac{7}{4} \text{ Bits}, \quad H(Y) = 2 \text{ Bits}$$

$$\begin{aligned} H(X|Y) &= \sum_{j=1}^4 P_Y(y_j) \cdot H(X|Y=y_j) \\ &= \frac{1}{4} \cdot H(X|Y=1) + \frac{1}{4} \cdot H(X|Y=2) \\ &= \frac{1}{4} \cdot H(X|Y=3) + \frac{1}{4} \cdot H(X|Y=4) \end{aligned}$$

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

14

Bedingte Entropie (3)

ETH

$$H(X|Y=1) = - \sum_{i=1}^4 P_{X|Y}(x_i, 1) \cdot \log(P_{X|Y}(x_i, 1))$$

$$\text{mit } P_{X|Y}(x_i, 1) = \frac{P_{XY}(x_i, 1)}{P_Y(1)}$$

$$x_i = 1 \Rightarrow \frac{1}{2}, x_i = 2 \Rightarrow \frac{1}{4}, x_i = 3 \Rightarrow \frac{1}{8}, x_i = 4 \Rightarrow \frac{1}{8}$$

$$\rightarrow H(X|Y=1) = H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right) = \frac{7}{4}$$

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

15

Bedingte Entropie (4)

ETH

$$\begin{aligned} H(X|Y) &= \sum_{j=1}^4 P_Y(y_j) \cdot H(X|Y=y_j) \\ &= \frac{1}{4} \cdot H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right) + \frac{1}{4} \cdot H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) \\ &= \frac{1}{4} \cdot H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) + \frac{1}{4} \cdot H(1, 0, 0, 0) \\ &= \frac{1}{4} \cdot \frac{7}{4} + \frac{1}{4} \cdot \frac{7}{4} + \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot 0 = \frac{11}{8} \text{ Bits} \end{aligned}$$

$$H(Y|X) = \frac{13}{8} \text{ Bits und } H(X|Y) = \frac{11}{8} \text{ Bits}$$

$$H(Y|X) \neq H(X|Y), \text{ jedoch gilt immer:}$$

$$H(X) - H(X|Y) = H(Y) - H(Y|X) = I(X, Y) = \frac{3}{8} \text{ Bits}$$

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

16

Bedingte binäre Entropie (1)

ETH

- Selen X und Y zwei binäre Zufallsvariablen $\{0, 1\}$

$$P_{XY}(0, 1) = P_{XY}(1, 0) = P_{XY}(0, 0) = \frac{1}{3}, P_{XY}(1, 1) = 0$$

$$H(X|Y=0) = - \sum_{x \in \{0, 1\}} P_{X|Y}(x, 0) \cdot \log(P_{X|Y}(x, 0))$$

$$P_{X|Y}(x, 0) = \frac{P_{XY}(x, 0)}{P_Y(0)}$$

$$P_{X|Y}(0, 0) = \frac{1/3}{2/3} = \frac{1}{2}, P_{X|Y}(1, 0) = \frac{1/3}{2/3} = \frac{1}{2}$$

$$H(X|Y=0) = - \frac{1}{2} \log\left(\frac{1}{2}\right) - \frac{1}{2} \log\left(\frac{1}{2}\right) = \frac{1}{2} + \frac{1}{2} = 1 \text{ Bit}$$

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

17

Bedingte binäre Entropie (2)

ETH

$$H(X) = - \sum_{x \in \{0, 1\}} P_X(x) \cdot \log(P_X(x)) = 0.918$$

$$P_X(0) = \frac{2}{3}, P_X(1) = \frac{1}{3}$$

$$H(X|Y=1) \Rightarrow P_{X|Y}(x, 1) = \frac{P_{XY}(x, 1)}{P_Y(1)}$$

$$P_{X|Y}(0, 1) = \frac{1/3}{1/3} = 1, P_{X|Y}(1, 1) = 0$$

$$\rightarrow H(X|Y=1) = 0$$

$$\rightarrow H(X|Y=1) < H(X) < H(X|Y=0)$$

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

18

Bedingte binäre Entropie (3) **ETH**

$$\begin{aligned} \rightarrow H(X|Y) &= P_Y(0) \cdot H(X|Y=0) + P_Y(1) \cdot H(X|Y=1) \\ &= \frac{2}{3} \cdot 1 \text{ Bit} + \frac{1}{3} \cdot 0 \text{ Bits} \\ &= \frac{2}{3} \text{ Bit} < H(X) \end{aligned}$$

- Die Entropie kann lokal ansteigen, fällt jedoch im Durchschnitt immer

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

19

Mehrere Variablen **ETH**

- Entsprechend gilt für 3 Zufallsvariablen

$$H(X|YZ) = - \sum_z H(X|Y, Z=z) p_Z(z)$$

- wobei

$$H(X|Y, Z=z) = - \sum_{(x,y)} p_{X|YZ}(x, y, z) \log p_{X|YZ}(x, y, z)$$

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

20

Bedingte Information **ETH**

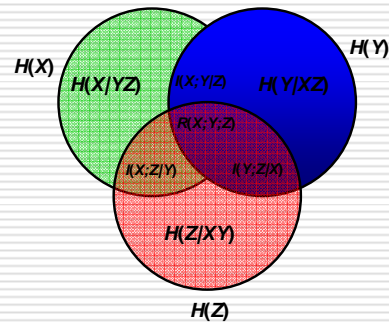
- Wir erweitern unser Diagramm um eine weitere Variable
- Wir erhalten 7 Bereiche
- Jeder Bereich entspricht einer Kombination von 7 Verbundentropien $H(X)$, $H(Y)$, $H(Z)$, $H(XY)$, $H(XZ)$, $H(YZ)$, $H(XYZ)$
- 3 der Bereiche haben bereits eine Interpretation
- Bedingte Entropien $H(X|YZ)$, $H(Y|XZ)$, $H(Z|XY)$
- Wir wollen die übrigen Bereiche kennen lernen
- Siehe Bild

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

21

Bild **ETH**



Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

22

Interpretation **ETH**

- Die bedingten Informationen $I(X; Y|Z)$, $I(Y; Z|X)$, $I(X; Z|Y)$ können wie folgt interpretiert werden:
 - Definition:** Die bedingte, gegenseitige Information, die X über Y gibt, gegeben Z, ist $I(X; Y|Z) := H(XZ) + H(YZ) - H(XYZ) - H(Z)$
 - Oder auch $I(X; Y|Z) := H(X|Z) - H(X|YZ)$
 - $I(X; Z|Y)$ ist also die Reduktion der Unsicherheit über X, wenn man Y erfährt, wobei aber Z schon bekannt ist
 - Man interpretiere $R(X; Y; Z)$ (kann negativ sein)
- $$R(X; Y; Z) := H(X) + H(Y) + H(Z) - H(XY) - H(XZ) - H(YZ) + H(XYZ)$$

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

23

Interpretation **ETH**

- Theorem:** Es gilt $I(X; Y|Z) \geq 0$
- Die Aussage, dass Zusatzinformation die Entropie nicht vergrössern kann, gilt jedoch nicht für $I(X; Y|Z) > I(X; Y)$, da $R(X; Y; Z)$ negativ werden kann
- Z bewirkt in diesem Falle eine stärkere Reduktion von $H(X|Y)$, als von $H(X)$
- Hierbei gilt:

$$H(X|Y) - H(X|YZ) > H(X) - H(X|Z)$$



Mit Hilfe von Entropien kann intuitiv gerechnet werden. Wir betrachten zwei Beispiele

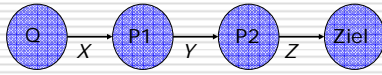
Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

24

Beispiel 1: Markov-Kette

ETH



- Wir betrachten folgende Markov-Kette
- P1 und P2 sind Prozessoren, die Berechnungen durchführen
- X wird durch P1 in Y übergeführt usw.
- P1 und P2 können beliebige deterministische oder probabilistische Operationen durchführen
- Einschränkung: Kein versteckter Pfad von X nach Z, also Markov-Eigenschaft

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

25

Beispiel 1: Markov-Kette

ETH

- Dies ist, wie bekannt, gleichbedeutend mit

$$p_{Z|XY}(x, y, z) = p_{Z|Y}(y, z)$$

- Mittels Entropie wird daraus

$$H(Z | XY) = H(Z | Y)$$

- oder auch

$$I(Z; X | Y) = 0$$

- Es gilt die Symmetrie-Eigenschaft

$$I(Z; X | Y) = I(X; Z | Y)$$

- Und damit die **Umkehrbarkeit** der Markov-Kette, also $X \rightarrow Y \rightarrow Z$ sowie $Z \rightarrow Y \rightarrow X$

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

26

Beispiel 1: Markov-Kette

ETH

- Lemma (Informationsverarbeitung):**

Falls $X \rightarrow Y \rightarrow Z$, dann gelten

$$I(X; Z) \leq I(Y; Z), \text{ sowie } I(X; Z) \leq I(X; Y)$$

- Beweis: Es gilt

$$I(Y; Z) = R(X; Y; Z) + I(Y; Z | X)$$

- Und

$$I(X; Z) = R(X; Y; Z) + I(X; Z | Y)$$

- Also gilt (zweiter Teil geht ebenso)

$$I(Y; Z) - I(X; Z) = I(Y; Z | X) - I(X; Z | Y) = I(Y; Z | X) \geq 0$$



Es gibt keine Operation auf Y, die die Information erhöhen kann, welche Y über X enthält

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

27

Perfekte Verschlüsselung

ETH

- Ein Klartext M soll mit einem geheimen Schlüssel K zu einem Chiffre C verschlüsselt werden
- Empfänger kann mit Schlüssel den Code entschlüsseln

- Definition:** Ein Verschlüsselungsverfahren heisst perfekt sicher, wenn

$$I(M; C) = 0 \Leftrightarrow P(M | C) = P(M)$$

- Bedeutung: Auch mit unbegrenzten Rechenressourcen ist der Code nicht zu knacken!
- Stärkste (theoretische) Sicherheit

Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

28

Perfekte Verschlüsselung

ETH

- Theorem:** In einem perfekt sicheren Verschlüsselungssystem gilt:

$$H(K) \geq H(M)$$

- Aufgrund der Entschlüsselbarkeit von M mit C und K gilt:

$$H(M | CK) = 0$$

- Perfekte Sicherheit bedeutet $b = -a$ im Bild

- Wir erinnern uns, dass $b < 0$ möglich ist

- Umgekehrt ist auch

$$I(C; K) \geq 0 \Rightarrow c \geq -b = a$$

- $H(K) \geq H(M)$ folgt aus Vergleich der Regionen

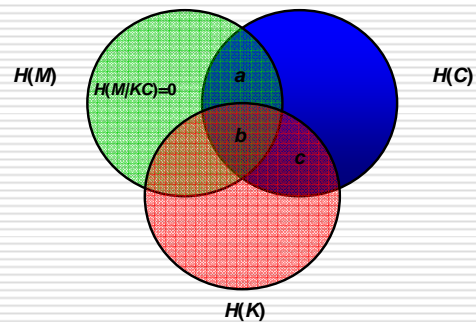
Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

29

Bild dazu

ETH



Entropie II

Informationstheorie
Copyright M. Gross, ETH Zürich 2005, 2006

30