Aufgaben zum Tut am 26.06.2006

Thomas Pajor

26. Juni 2006

Aufgabe 1.

Beweisen Sie die Kettenregel für die Kullback–Leibler–Divergenz:

$$D(p(x,y)||q(x,y)) = D(p(x)||q(x)) + D(p(y|x)||q(y|x))$$

Hinweis: Die bedingte Kullback–Leibler–Divergenz ist definiert als

$$D(p(y|x)||q(y|x)) = \sum_x p(x) \sum_y p(y|x) \log \frac{p(y|x)}{q(y|x)}$$

Lösung.

Wir lösen die Aufgabe durch Nachrechnen. Es gilt

$$\begin{aligned}
D(p(x,y)||q(x,y)) &= \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{q(x,y)} \\
&= \sum_x \sum_y p(x,y) \log \frac{p(x)p(y|x)}{q(x)q(y|x)} \\
&= \sum_x \sum_y p(x|y)p(y) \log \frac{p(x)}{q(x)} + \sum_x \sum_y p(x,y) \log \frac{p(y|x)}{q(y|x)} \\
&= \sum_x p(x) \log \frac{p(x)}{q(x)} + \sum_x \sum_y p(x)p(y|x) \log \frac{p(y|x)}{q(y|x)} \\
&= D(p(x)||q(x)) + D(p(y|x)||q(y|x))
\end{aligned}$$

Die linke Term in der vorletzte Zeile folgt mit dem Gesetz der totalen Wahrscheinlichkeiten[1], das besagt, dass für zwei stochastisch abhängige Zufallsvariablen X und Y gilt

$$\sum_y p(x|y)p(y) = p(x)$$

$\square$

Aufgabe 2.

Wir wollen die Logarithmen–Summen–Ungleichung beweisen.

(a) Zeigen Sie: $f(t) = t \log t$ ist ein streng konvexe Funktion für $t \geq 0$.

(b) Sei nun f eine konvexe Funktion und $p_i \geq 0$ mit $\sum_{i=1}^n p_i = 1$. Zeigen Sie, dass dann gilt:

$$\sum_{i=1}^n p_i f(t_i) \geq f(\sum_{i=1}^n p_i t_i)$$

(c) Beweisen Sie mit Hilfe von (b):

$$\sum_{i=1}^n a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^n a_i \right) \log \frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}$$

Diese Ungleichung wird auch *Logarithmen–Summen–Ungleichung* genannt, und könnte bei Aufgabe 5a auf dem aktuellen Übungsblatt sehr behilflich sein.

Lösung.

(a) Wir zeigen, dass die Funktion f (streng) konvex ist indem wir die zweite Ableitung untersuchen. Zur Vereinfachung sei $c := \frac{1}{\ln(2)}$, dann ist $\log t = c \cdot \ln t$. Es gilt also

$$f'(t) = c \cdot \ln t + c$$

und damit

$$f''(t) = c \cdot \frac{1}{t}$$

Offenbar ist $f''(t) > 0$ für alle $t \geq 0$, somit hat die Funktion f keine Wendepunkte und besteht überall aus einer Linkskrümmung. Damit folgt, dass die Funktion streng konvex ist. $\square$

[1] siehe WT

(b) Wir beweisen die Aussage durch Induktion über n.

IA: $n = 2$.

Wir haben zwei Werte p_1 und p_2 mit $p_1 + p_2 = 1$. Damit ist $p_2 = (1 - p_1)$. Die zu zeigende Aussage ist also

$$p_1 f(t_1) + (1 - p_1) f(t_2) \geq f(p_1 t_1 + (1 - p_1) t_2)$$

Das ist aber gerade die Definition einer konvexen Funktion, somit ist die Gleichung für $n = 2$ erfüllt.

IS: $n - 1 \rightsquigarrow n$

Definiere zunächst für die Werte p_i für $i = 1, \ldots, n - 1$ neue Werte

$$p'_i := \frac{p_i}{1 - p_n}$$

Also ist $p_i = (1 - p_n) p'_i$. Außerdem gilt

$$\sum_{i=1}^{n-1} p_i = (1 - p_n) \;\Rightarrow\; \sum_{i=1}^{n-1} \frac{p_i}{1 - p_n} = \frac{1 - p_n}{1 - p_n} \;\Rightarrow\; \sum_{i=1}^{n-1} p'_i = 1$$

was wir für die Anwendung der Induktionsvoraussetzung benötigen. Damit folgt nun:

$$\begin{aligned}
\sum_{i=1}^{n} p_i f(t_i) &= p_n f(t_n) + (1 - p_n) \sum_{i=1}^{n-1} p'_i f(t_i) \\
&\overset{\text{IV}}{\geq} p_n f(t_n) + (1 - p_n) f\left(\sum_{i=1}^{n-1} p'_i t_i \right) \\
&\overset{\text{IA}}{\geq} f\left(p_n t_n + (1 - p_n) \sum_{i=1}^{n-1} p'_i t_i \right) \\
&= f\left(p_n t_n + \sum_{i=1}^{n-1} p_i t_i \right) \\
&= f\left(\sum_{i=1}^{n} p_i t_i \right)
\end{aligned}$$

$\square$

(c) Mit der Vorarbeit aus Aufgabe (a) und (b) ist dies sehr einfach. Setze $f(t) = t \log t$. Aus Aufgabe (b) wissen wir, dass f konvex ist, und mit Aufgabe (b) folgt damit

$$\sum_i \alpha_i f(t_i) \geq f\left(\sum_i \alpha_i t_i \right)$$

Für $\alpha_i \geq 0$ und $\sum_i \alpha_i = 1$. Wir definieren die α_i und t_i durch die a_i und b_i aus der Aufgabenstellung wie folgt:

$$\alpha_i := \frac{b_i}{\sum_j b_j}$$

sowie

$$t_i := \frac{a_i}{b_i}$$

Damit ist $\sum_i \alpha_i = \frac{b_1 + \ldots + b_n}{\sum_j b_j} = \frac{\sum_i b_i}{\sum_j b_j} = 1$ erfüllt, und somit können wir Aufgabe (b) benutzen und erhalten

$$\sum_i \frac{b_i}{\sum_j b_j} \frac{a_i}{b_i} \log \frac{a_i}{b_i} \geq \sum_i \frac{b_i}{\sum_j b_j} \frac{a_i}{b_i} \log \sum_i \frac{b_i}{\sum_j b_j} \frac{a_i}{b_i}$$

Kürzen sowie Multiplikation beider Seiten mit $\sum_j b_j$ ergibt

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \sum_i a_i \log \sum_i \frac{a_i}{\sum_j b_j}$$

oder anders geschrieben

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \sum_i a_i \log \frac{\sum_i a_i}{\sum_j b_j}$$

was gerade die zu beweisende Aussage war. $\square$