Mějme 32-stranou kostku, férovou. Jaká je její entropie?

$$H(X) = -\sum_{i=1}^{32} p(x) \cdot \log_2(p(x))$$

$$= -\sum_{i}^{32} \frac{1}{32} \cdot \log\left(2^{-5}\right) = -32 \cdot \frac{1}{32} \cdot -5 = \underline{\underline{5}}$$

$P(X,Y)$, $X = \{\text{wet}, \text{dry}\}$, $Y = \{\text{rain}, \text{sun}, \text{clouds}\}$

$P(\text{wet}, \text{sun}) = 0{,}1$

$P(\text{wet}, \text{rain}) = 0{,}3$

$P(\text{sun}) = 0{,}7$

$p(\text{dry}) = ?$

$\longrightarrow$ všechno musí sčítat do 1

$p(\text{sun}) = p(\text{wet}, \text{sun}) + p(\text{dry}, \text{sun}) \Rightarrow p(\text{dry}, \text{sun}) = 0{,}6$

$p(\text{wet}, \text{clouds}) = p(\text{wet}) - 0{,}4$

$p(\text{wet}, \text{clouds}) + p(\text{dry}, \text{rain}) + p(\text{dry}, \text{sun}) + p(\text{dry}, \text{clouds}) = 0{,}6$

$p(\text{wet}, \text{clouds}) + p(\text{dry}, \text{rain}) + 0{,}6 + p(\text{dry}, \text{clouds}) = 0{,}6$

$p(\text{wet}, \text{clouds}) + p(\text{dry}, \text{rain}) + p(\text{dry}, \text{clouds}) = 0$

$p(\text{dry}) = p(\text{dry}, \text{rain}) + p(\text{dry}, \text{clouds}) + p(\text{dry}, \text{sun}) = 0 + 0 + \underline{\underline{0{,}6}}$

Mám kostku, na které padne:

4x 2

4x 3

1x 6

1x 1

$$E(X) = ? \quad \frac{1}{10} \cdot 1 + \frac{4}{10} \cdot 2 + \frac{4}{10} \cdot 3 + \frac{1}{10} \cdot 6 = \frac{1}{10} + \frac{8}{10} + \frac{12}{10} + \frac{6}{10} = \frac{27}{10}$$

Mějme $p(A|B)$, $A = \{\text{dry}, \text{wet}\}$, $B = \{\text{rain}, \text{sun}\}$

všechny podmíněné budou $1/2$, $\downarrow$ tudíž všechny joint $1/4$

$p(\text{dry}/\text{rain}) = 1/2$, $p(\text{dry}/\text{sun}) = \frac{1}{2}$, $p(\text{rain}) = 1/2 \Rightarrow p(\text{sun}) = 1/2$

what is the cond. entropy $p(A|B)$?

$p(a/b) = 1/2$

$p(\text{dry} \mid x) = p(\text{dry}, x) / P(x)$

We need to calculate $p(A, B)$

$P(\text{dry}, x) = p(\text{dry}/x) \cdot p(x)$

$$-4 \cdot \frac{1}{4} \cdot -1 = 1$$

$p(\text{dry}, x) = 1/4$

- $\Omega = \{a,b,..,z\}$, prob. distribution (assumed/estimated from data):
  $p(a) = .25$, $p(b) = .5$, $p(\alpha) = 1/64$ for $\alpha \in \{c..r\}$, $= 0$ for the rest: s,t,u,v,w,x,y,z

Mějme „break" $\rightarrow$ p': $p(b/r/e/a/k) = \frac{1}{5}$

Compute cross-entropy:

$$H_{p'}(p) = -1/|T| \cdot \sum \log(p(x))$$

$\rightarrow$ přesto by to mohlo být i:

$$-\sum_{x \in \Omega} p'(x) \cdot \log(p(x))$$

$$= -\frac{1}{5} \cdot \left(\log\left(\frac{1}{2}\right) + \log\left(\frac{1}{64}\right) + \log\left(\frac{1}{64}\right) + \log\left(\frac{1}{4}\right) + \log\left(\frac{1}{64}\right)\right)$$

$$= -\frac{1}{5} \cdot (-1 + -6 - 6 - 2 - 6)$$

$$= -\frac{1}{5} \cdot -21 = \frac{21}{5}$$

What is the entropy of bigram distribution?

No two numbers are so similar numbers.

$p(x,y) = 1/8$

$p(x|y) = 1$, except $p(\text{are}|\text{numbers}) = p(\cdot|\text{numbers}) = 0{,}5$

$\rightarrow 2^{-1} \rightarrow \log(2^{-1}) = -1$

$$H(p) = -\sum_{x,y} p(x,y) \cdot \log(p(x|y))$$

$$= -\frac{1}{8} \cdot -2 = \frac{1}{4}$$

estimate bigram LM from this data ...

Compute its CE against the following data

$\rightarrow$ actually $H: \frac{1}{|T|} \cdot \sum \log(p(x_i|y_i))$ can be used.

$\rightarrow$ $p(x,y)$ goes from the data

$p(x|y)$ goes from the trained model

No two numbers. $\rightarrow p(x,y) = 1/4$

$\left.\begin{array}{l} \text{no / \#} \\ \text{two / no} \\ \text{numbers / two} \end{array}\right\} = 1 \rightarrow \log = 0$

only: $p(\cdot|\text{numbers}) = \frac{1}{2} \rightarrow \log = -1$

$$CE: -\sum_{x,y} p(x,y) \cdot \log(p(x|y)) = -\frac{1}{4} \cdot -1 = \frac{1}{4}$$

$$p'(w|h) = \frac{(c(w,h)+1)}{(c(h)+|V|)}$$ $\rightarrow$ Add 1 smoothing

$$p = \frac{1}{16}, \text{ uniform} \rightarrow E(p) = \sum_x \frac{1}{16} \cdot \log(2^{-4}) = -\sum_x -\frac{1}{4} = -16 \cdot -\frac{1}{4} = 4$$

Perplexity: $2^{E(p)} = 2^4 = 16$

$P_j(X,Y)$, $X = \{dry, wet\}$, $Y = \{rain, sun, clouds\}$

$p(wet, sun) = 0,1$

$p(wet, rain) = 0,3$

$p(sun) = 0,7$

$p(dry) = ?$

$p(wet) = p(wet, sun) + p(wet, rain) + p(wet, clouds)$

$p(sun) = 0,7$

$1 = 0,4 + p(wet, clouds) + p(dry, clouds) + p(dry, rain) + p(dry, sun)$

$$0,6 = p(wet, clouds) + \underbrace{p(dry, clouds) + p(dry, rain) + p(dry, sun)}_{= p(dry)}$$

$p(dry) = 0,6 - p(wet, clouds)$

$p(wet) = 0,4 + p(wet, clouds)$

$p(dry, sun) + p(wet, sun) = p(sun) = 0,7$

$p(dry, sun) + 0,1 = 0,7$

$p(dry, sun) = 0,6$

$0,6 = p(wet, clouds) + p(dry, clouds) + p(dry, rain) + 0,6)$

$0 = p(wet, clouds) + p(dry, clouds) + p(dry, rain)$

Since $p(.,..) \geq 0$, then all must be 0.

	= konferenci	≠ konferenci
= tiskové	37	13
≠ tiskové	20	222 341

Compute PMI(tiskové, konferenci)

$p(t, k) = \frac{37}{222\,441}$ $\qquad \log(p(t,k) / p(t) \cdot p(k)) = 11,49$ $\rightarrow$ they are pretty strong collocations

$p(t) = \frac{50}{222\,441}$

$p(k) = \frac{57}{222\,441}$

	B	CS	EN
B	0	1/2	1/2
CS	0	1	0
EN	0	0	1

	#	a	to	ne	set
B	1	0	0	0	0
CS	0	1/3	1/3	1/3	0
EN	0	1/2	1/3	0	1/3

\# to <u>ne</u>

$$1 \cdot \frac{1}{2} \cdot \frac{1}{3} \cdot 1 \cdot \frac{1}{3} = \frac{1}{6} \cdot \frac{1}{3} = \frac{1}{18}$$

$$1 \cdot \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$$

\# to ne

$$\begin{array}{cccc} \# & to & & ne \\ & CS & - & CS \\ B \Big< & & & \\ & EN & - & CS \\ & to & & x \end{array} \quad = 1 \cdot \frac{1}{2} \cdot \frac{1}{3} \cdot 1 \cdot \frac{1}{3} = \frac{1}{18}$$

	B	CS	EN
B	0	1/2	1/2
CS	0	1	0
EN	0	0	1

	#	to	je	zde	Pinguin	go
B	1	0	0	0	0	0
CS	0	1/3	1/3	1/3	0	0
EN	0	1/3	0	0	1/3	1/3

b a a a

$$p(a|b) = p(b,a) / p(b) = \frac{1}{8} / \frac{1}{8} = 1$$

$$p(a|a) = p(a,a) / p(a) = \frac{1}{4} / \frac{1}{2} = \frac{1}{2}$$

$$-\frac{1}{3} \cdot \left(\log\left(\frac{1}{2}\right) + \log\left(\frac{1}{2}\right) \right) = 2/3$$