Opakování minule.

- X n.v., pak entropie X je $H(X) = -\sum_{x \in X} p(x) \lg_2 p(x)$
- lze též interpretovat $H(X)$ jako střední hodnotu n.v. $\lg(p(x)^{-1})$
- $H(x) \geq 0$ vždy
- joint entropy $H(X,Y)$ pro dvojici n.v. X, Y je $H(X,Y) = -\sum_{x \in X, y \in Y} p(X=x \,\&\, Y=y) \lg(p(X=x, Y=y))$

- podmíněná entropie

$$H(Y|X) = \sum_{x \in X} p(x) H(Y|X=x)$$

$$= -\sum_{x \in X} \Pr[X=x] \sum_{y \in Y} \Pr[Y=y|X=x] \lg(\Pr[Y=y|X=x])$$

$$= -\sum_{x \in X} \Pr[X=x, Y=y] \lg(\Pr[Y=y|X=x])$$

$$= -\mathbb{E}[\lg(p(y|x))]$$

- řetízkové pravidlo (joint rule)
  - $H(X,Y) = H(X) + H(Y|X)$

- $H(X,Y|Z) = H(X|Z) + H(Y|X,Z)$

- a obecně $H(X_1, \ldots, X_n) = \sum_{i=1}^{n} H(X_i | X_{i-1}, X_{i-2}, \ldots, X_1)$

- $H(X,Y) = H(X) + H(Y|X)$
  $= H(Y) + H(X|Y)$

- obecně neplatí $H(X|Y) = H(Y|X)$!!

Relativní entropie

- měří "vzdálenost" mezi dvěma distribucemi
- pokud p, q jsou pstí rozdělení, značíme jejich rel. entropii $D(p\|q)$
- vzdálenost ve smyslu, jak moc je neefektivní chovat se k p jakoby by byla q
  - př. délka prům. kód slova efekt. kódování p je $\sim H(p)$
  - ale kdybychom použili kód pro q, tak je to $H(p) + D(p\|q)$
- $D(p\|q) = \sum_{x \in X} p(x) \lg \frac{p(x)}{q(x)}$ $\quad \left(0 \cdot \lg \frac{0}{0} = 0,\ 0 \lg \frac{0}{q} = 0,\ p \lg \frac{p}{0} = \infty \right)$
- aka Kullback-Leibler divergence $\quad = \mathbb{E}\left[\lg \frac{p(X)}{q(X)} \right]$
- proč jde o vzdálenost v uvozovkách?
  - není symetrická
  - nesplňuje Δ nerovnost
- uvidíme $D(p\|q) \geq 0$, $D(p\|q) = 0 \iff p = q$

Def: X, Y n.v. se sdruženou pstní fcí $p(x,y)$ a marginál pst fce $p(x), p(y)$

Vzájemná informace $I(X:Y)$ je

$$I(X:Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y) \lg \frac{p(x,y)}{p(x)p(y)} = D(p(x,y) \| p(x)p(y)) =$$

$$= \mathbb{E}_{p(x,y)} \left[\lg \frac{p(x,y)}{p(x)p(y)} \right]$$

Př: uvažme $X = \{0,1\}$, $p(0) = 1-r$, $p(1) = r$, $q(0) = 1-s$, $q(s)$, p, q jsou pstní fce.

- $D(p\|q) = (1-r) \lg \frac{1-r}{1-s} + r \lg \frac{r}{s}$

  $D(q\|p) = (1-s) \lg \frac{1-s}{1-r} + s \lg \frac{s}{r}$

- $r = s \Rightarrow D(p\|q) = D(q\|p) = 0$
- $r = \frac{1}{2},\ s = \frac{1}{4} \Rightarrow D(p\|q) = 0.2075 \neq D(q\|p) = 0.1887$

Alternativní definice:

$$I(X:Y) = \sum_{x,y} p(x,y) \lg \frac{p(x,y)}{p(x)p(y)} = \sum_{x,y} p(x,y) \lg \frac{p(x|y)}{p(x)} =$$

$$= \underbrace{\sum_{x,y} p(x,y) \lg p(x|y)}_{-H(X|Y)} - \underbrace{\sum_{x,y} p(x,y) \lg p(x)}_{+H(X)}$$

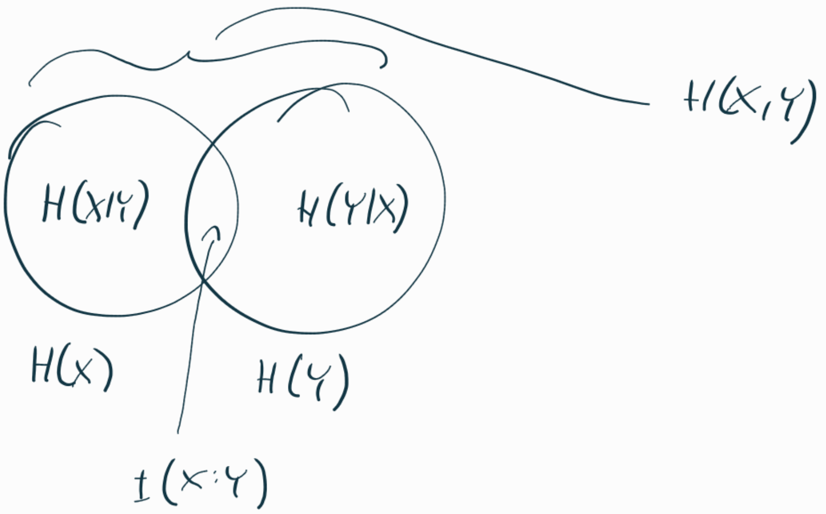
$$= H(X) - H(X|Y)$$

≈ o kolik se sníží neurčitost X, když znám Y

$$I(X:Y) \underset{\text{symetricky}}{=} H(Y) - H(Y|X)$$

→ symetrie informace $\Downarrow$

- $H(X,Y) = H(X) + H(Y|X) \Rightarrow I(X:Y) = H(X) + H(Y) - H(X,Y)$
- $I(X:X) = H(X) - H(X|X) = H(X)$



Def.: Podmíněná vzájemná informace n.v. X a Y za předpokladu Z je

$$I(X:Y|Z) = H(X|Z) - H(X|Y,Z)$$

Věta: $$I(\overbrace{X_1, X_2, \dots, X_n}^{\text{náhodný vektor}} : Y) = \sum_{i=1}^{n} I(X_i : Y | X_{i-1}, X_{i-2}, \dots, X_1)$$

Dk: $$I(X_1, \dots, X_n : Y) = H(X_1, \dots, X_n) - H(X_1, \dots, X_n | Y) =$$

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

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

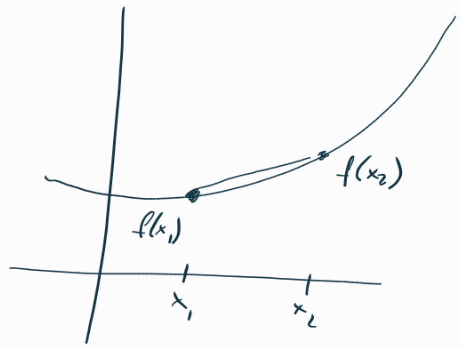
$$= \sum_{i=1}^{n} I(X_i : Y, X_{i-1}, \dots, X_1)$$

· lze definovat podm. KL-divergenci a řetízkové pravidlo pro KL, ale nebudeme dělat

Jensenova nerovnost

· fce $f : \mathbb{R} \to \mathbb{R}$ je konvexní na intervalu (a,b) pokud $\forall x_1, x_2 \in (a,b)$ a $0 \le \lambda \le 1$

platí:
$$f(\lambda x_1 + (1-\lambda) x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$$



Věta: (Jensenova nerovnost) f: konvexní fce na (a,b),

n.v. $X \subseteq (a,b)$. Pak
$$\mathbb{E}[f(X)] \ge f(\mathbb{E}[X])$$

Dk: pro $|X|$ konečné

· indukcí podle $|X|$ → $X = \{x_1, x_2\}$

· pokud $|X| = 2$, kde p je pst prvního jevu a $1-p$ toho druhého, pak

J.n. je
$$p f(x_1) + (1-p) f(x_2) \ge f(p x_1 + (1-p) x_2)$$

· to je konvexita sama o sobě ... ✓

· IK:

· ozn $p_i' = \dfrac{p_i}{1-p_n}$

$$\sum_{i=1}^{n} p_i f(x_i) = p_n f(x_n) + \sum_{i=1}^{n-1} p_i f(x_i) = p_n f(x_n) + (1-p_n) \sum_{i=1}^{n-1} p_i' f(x_i) \overset{IP}{\ge}$$
$$\overset{IP}{\ge} p_n f(x_n) + (1-p_n) f\left(\sum_{i=1}^{n-1} p_i' x_i\right) \underset{\text{konvexita}}{\ge} f\left(p_n x_n + (1-p_n) \sum_{i=1}^{n-1} p_i' x_i\right) =$$
$$= f\left(\sum_{i=1}^{n} p_i x_i\right) \quad \square$$

Důsledky

Věta: Nechť $p(x)$, $q(x)$ jsou pstní rozdělení na X. Pak $D(p \| q) \ge 0$

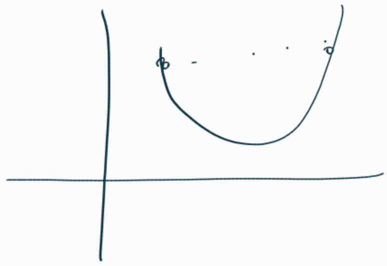
Dk:

· ozn $A = \operatorname{supp}(p(x)) = \{x \in X : p(x) > 0\}$

· z J.n. plyne $\mathbb{E}[f(X)] \ge f(\mathbb{E}[X]) \Rightarrow -\mathbb{E}[f(X)] \le -f(\mathbb{E}[X])$

· $D(p\|q) = \sum_{x\in A} p(x) \lg \frac{p(x)}{q(x)} = -\sum_{x\in A} p(x) \lg \frac{q(x)}{p(x)} \le -\lg \sum_{x\in A} \cancel{p(x)} \frac{q(x)}{\cancel{p(x)}} = -\lg \sum_{x\in A} q(x) \le$

$$\le -\lg \sum_{x\in X} q(x) = -\lg 1 = 0$$



· $D(p\|q)=0 \iff p=q$

Dusl: $I(X:Y)\ge 0$

Dk: $I(X:Y) = D(p(x,y)\|p(x)p(y)) \ge 0$ □

Věta: $H(X)\le \lg(|X|)$

Dk:

· nechť $u(x)=\frac{1}{|X|}$, (p je pstní fce X)

· $D(p\|u) = \sum_{x\in X} p(x)\lg\frac{p(x)}{u(x)} = \sum_{x\in X} p(x)\lg p(x) - \sum p(x) \lg u(x)$

$$= +\sum_{x\in X} p(x) \lg|X| - H(X)$$

$$= \lg|X| - H(X)$$

· $0\le D(p\|u) = \lg|X| - H(X)$ s rovností, pokud $p=u$

Věta: $H(X|Y)\le H(X)$

· "více informace neuškodí"

Dk:

$0\le I(X:Y) = H(X)-H(X|Y)$

pouze v průměru, může se stát $H(X|Y=y) \ge H(X)$ pro nějaké y

Př:

Y \ X	1	2
1	0	$\frac{3}{4}$
2	$\frac{1}{8}$	$\frac{1}{8}$

· $H(X) = H(\frac{7}{8}) = 0.544$

· $H(X|Y=1) = 0$, $H(X|Y=2) = \lg 2 = 1$

· $H(X|Y) = \frac{3}{4} H(X|Y=1) + \frac{1}{4} H(X|Y=2) = 0.25$

Data processing ineq

Alenka   Bob

$$X \xrightarrow[\text{šum}]{} Y \xrightarrow{g} Z = g(Y)$$

$$I(X:Z) > I(X:Y) \; ? \quad \text{NE!}$$

· "chytré zpracování dat nám nezlepší info hodnotu"

· X, Y, Z mají Markovskou vl. pokud $\Pr[Z=z \mid Y=y \,\&\, X=x] = \Pr[Z=z \mid Y=y]$ $\quad \forall x,y,z$

$$\rightsquigarrow p(x,y,z) = p(x)\, p(y,z \mid x) = p(x)\, p(y \mid x)\, p(z \mid x,y) \underset{\text{Markov}}{=} p(x)\, p(y \mid x)\, p(z \mid y)$$

· důsledky:

- X a Z jsou podm. Y nezávislé $\quad p(x,z \mid y) = p(x \mid y)\, p(z \mid y)$
- $X \to Y \to Z$ je M.ř. $\iff$ $Z \to Y \to X$ je M.ř.

$$= \frac{p(x,y,z)}{p(y)} = \frac{p(x,y)\, \overset{p(z \mid x,y) \,\text{"}}{p(z \mid y)}}{p(y)} =$$

<u>Věta:</u> $I(X:Y) \geq I(X:Z)$

<u>Dk:</u>

$$I(X:Y,Z) = I(X:Z) + I(X:Y \mid Z)$$

$$= I(X:Y) + I(X:Z \mid Y)$$

· Z a X jsou nezávislé podm. Y $\Rightarrow$ $I(X:Z \mid Y) = 0$

· $I(X:Z) + I(X:Y \mid Z) = I(X:Y)$

· $I(X:Y \mid Z) \geq 0 \Rightarrow I(X:Y) \geq I(X:Z)$

<u>Důsl:</u> $I(X:Y) \geq I(X:g(Y))$ pro lib. fci g

<u>Důsl:</u> $X \to Y \to Z$, pak $I(X:Y) \geq I(X:Y \mid Z)$

· takže pokud vysledujeme Z, pak se sníží závislost X a Y

· $X, Y \sim \mathrm{Ber}(\frac{1}{2})$, $Z = X+Y$

· $I(X:Y) = 0$, $\quad I(X:Y \mid Z) = H(X \mid Z) - \underbrace{H(X \mid Y,Z)}_{0} = \frac{1}{2}$