Příklad: Statický systém v parlamentu

- podmnožina pěti vybraných poslanců $\mathbf{s} = \{v_1, \dots, v_5\}$
- někteří v_i hlasují podle v_j, $j \neq i$, ale nevíme kteří
- někteří hlasují nezávisle, nevíme kteří
- máme záznam 100 hlasování

	v_1	v_2	v_3	v_4	v_5	$p(\mathbf{s})$
s_0	0	0	0	0	0	0.02
	0	0	0	0	1	0.04
	0	0	0	1	0	0.02
	0	0	0	1	1	0.05
	0	0	1	0	0	0.03
	0	0	1	0	1	0.03
	0	0	1	1	0	0.09
	0	0	1	1	1	0.03
	0	1	0	0	0	0.03
	0	1	0	0	1	0.05
	0	1	0	1	0	0.04
	0	1	0	1	1	0.01
	0	1	1	0	0	0.03
	0	1	1	0	1	0.03
	0	1	1	1	0	0.04
	0	1	1	1	1	0.05

	v_1	v_2	v_3	v_4	v_5	$p(\mathbf{s})$
	1	0	0	0	0	0.01
	1	0	0	0	1	0.03
	1	0	0	1	0	0.03
	1	0	0	1	1	0.03
	1	0	1	0	0	0.01
	1	0	1	0	1	0.03
	1	0	1	1	1	0.01
	1	1	0	0	0	0.03
	1	1	0	0	1	0.03
	1	1	0	1	0	0.03
	1	1	0	1	1	0.04
	1	1	1	0	0	0.01
	1	1	1	0	1	0.04
	1	1	1	1	0	0.03
s_{30}	1	1	1	1	1	0.05

ostatní stavy se nevyskytly

pokračování

	v_1	v_2	v_3	$p(v_1, v_2, v_3)$
s_1	0	0	0	0.13
s_2	0	0	1	0.18
s_3	0	1	0	0.13
s_4	0	1	1	0.15
s_5	1	0	0	0.10
s_6	1	0	1	0.05
s_7	1	1	0	0.13
s_8	1	1	1	0.13

0.59

$\mathcal{A}$

s_1	s_2	s_3	s_4
		0.59	
s_5	s_6	s_7	s_8
		0.41	

rozklad proměnnou v_1

$\mathcal{B}$

s_1 s_2	s_3 s_4
s_5 s_6	s_7 s_8
0.46	0.54

rozklad proměnnou v_2

$\mathcal{A} \cdot \mathcal{B}$

s_1 s_2	s_3 s_4
0.31	0.28
s_5 s_6	s_7 s_8
0.15	0.26

$$H(\mathcal{A}) = 0.6769 \text{ nat} \qquad H(\mathcal{B}) = 0.6899 \text{ nat}$$

$$H(\mathcal{A} \cdot \mathcal{B}) = 1.3543 \text{ nat} \qquad H(\mathcal{E}) = 2.0342 \text{ nat}$$

Příklad na úvod k podmíněné entropii

Parlament, $\mathbf{S} = \{v_1, v_2, v_3\}$, v_1 indukuje rozklad $\mathcal{A}$

	v_1	v_2	v_3	$p(v_1, v_2, v_3)$	$p(v_2, v_3 \mid v_1 = 0)$
s_1	0	0	0	0.13	$0.13/0.59 = 0.2203$
s_2	0	0	1	0.18	0.3051
s_3	0	1	0	0.13	0.2203
s_4	0	1	1	0.15	0.2542
					$p(v_2, v_3 \mid v_1 = 1)$
s_5	1	0	0	0.10	0.2439
s_6	1	0	1	0.05	0.1220
s_7	1	1	0	0.13	0.3171
s_8	1	1	1	0.13	0.3171

$$H(v_1, v_2, v_3) = 2.0342 \text{ nat}$$

$$H(v_2, v_3) = 1.3827 \text{ nat}$$

$$H(v_2, v_3 \mid v_1 = 0) = 1.3769 \text{ nat}$$

$$H(v_2, v_3 \mid v_1 = 1) = 1.3291 \text{ nat}$$

$$H(v_2, v_3 \mid v_1) = 0.59 H(v_2, v_3 \mid v_1 = 0) + 0.41 H(v_2, v_3 \mid v_1 = 1) = 1.3573 \text{ nat}$$

Interpretace

- Pokud nevíme nic o stavu systému, nejistota o jeho stavu $\mathbf{S}$ je 2.0342 nat.
- Pokud známe hodnotu proměnné v_1, nejistota o $\mathbf{S}$ klesne na 1.3573, což je méně než $H(v_2, v_3) = 1.3827$, protože v_1 obsahuje o (v_2, v_3) nějakou informaci.
- Informace, kterou obsahuje v_1 o (v_2, v_3) je $H(v_2, v_3) - H(v_2, v_2 \mid v_1) = 1.3827 - 1.3573 = 0.0254$ nat.

Podmíněná entropie

Příklad: Uvažujme sekvenci pokusů, v nichž nastal jev L (padla lichá při házení kostkou), v takové sekvenci je nejistota o elementárním rozkladu rovna $H(\mathcal{E} \mid L)$.

Uvažujme sekvenci, v níž nastal doplněk jevu S (sudá), v ní je nejistota o $\mathcal{E}$ rovna $H(\mathcal{E} \mid S)$.

Vážený součet $H(\mathcal{E} \mid L) \cdot P(L) + H(\mathcal{E} \mid S) \cdot P(S)$ je podmíněná entropie $\mathcal{E}$, pozorujeme-li rozklad $\{L, S\}$.

Def. Nechť rozklady $\mathcal{A}$ a $\mathcal{B}$ jsou $\mathcal{A} = \{A_1, A_2, \ldots, A_{N_A}\}$, $\mathcal{B} = \{B_1, B_2, \ldots, B_{N_B}\}$. Potom podmíněná entropie rozkladu $\mathcal{A}$ za předpokladu, že nastal jev B_j je

$$H(\mathcal{A} \mid B_j) = -\sum_{i=1}^{N_A} P(A_i \mid B_j) \ln P(A_i \mid B_j)$$

Podmíněná entropie $H(\mathcal{A} \mid \mathcal{B})$ je pak střední hodnota přes $\mathcal{B}$

$$H(\mathcal{A} \mid \mathcal{B}) = \sum_{j=1}^{N_B} P(B_j)\, H(\mathcal{A} \mid B_j) = -\sum_{j=1}^{N_B} P(B_j) \left(\sum_{i=1}^{N_A} P(A_i \mid B_j) \ln P(A_i \mid B_j) \right) =$$

$P(B_j)\, P(A_i \mid B_j)$

$$= -\sum_{i} \sum_{j} P(A_i, B_j) \ln P(A_i \mid B_j)$$

Pozn: Střední nejistota o $\mathcal{A}$ je-li pozorováno $\mathcal{B}$ je $H(\mathcal{A} \mid \mathcal{B})$.

Vlastnosti I

Pro každé rozklady $\mathcal{A}$ a $\mathcal{B}$ platí:

V5. Jestliže $\mathcal{B} \preceq \mathcal{A}$ potom $H(\mathcal{A} \mid \mathcal{B}) = 0$. Intuice: víme, které jevy v $\mathcal{A}$ nastaly.

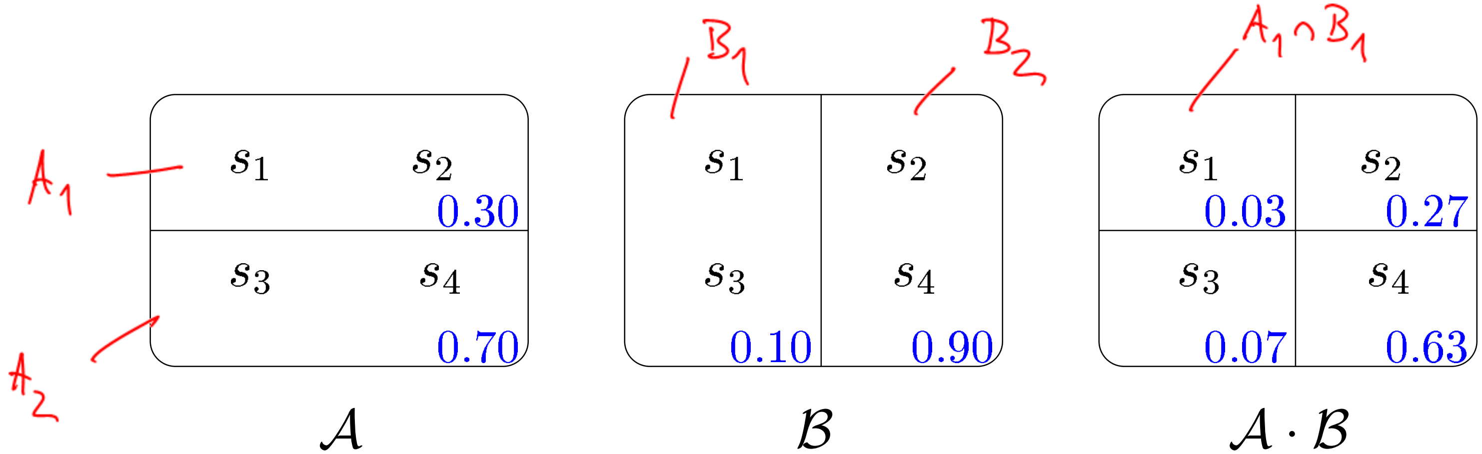
D:

1. $\mathcal{B}$ je zjemnění $\Rightarrow \forall j\, \exists!\, i : B_j \subseteq A_i$, takže $A_i \cap B_j = \begin{cases} B_j & \text{když } B_j \subseteq A_i \\ \emptyset & \text{jinak} \end{cases}$
2. $P(A_i \mid B_j) = \dfrac{P(A_i, B_j)}{P(B_j)} = \begin{cases} 1 & \text{když } B_j \subseteq A_i \\ 0 & \text{jinak} \end{cases}$
3. všechny členy v součtu $H(\mathcal{A} \mid B_j) = \sum_i P(A_i \mid B_j) \log P(A_i \mid B_j)$ jsou nulové.

$p = 0 \quad p \log p = 0$

$p = 1 \quad p \log p = 0$

Def. Rozklady $\mathcal{A}$ a $\mathcal{B}$ nezávislé $\Leftrightarrow \forall A_i \in \mathcal{A},\ B_j \in \mathcal{B}\colon P(A_i,\ B_j) = P(A_i) \cdot P(B_j)$.



V6. Jsou-li $\mathcal{A}$ a $\mathcal{B}$ nezávislé, potom $\quad H(\mathcal{A} \mid \mathcal{B}) = H(\mathcal{A}), \qquad H(\mathcal{B} \mid \mathcal{A}) = H(\mathcal{B})$

Intuice: Jsou-li jevy nezávislé, potom pozorováním $\mathcal{B}$ nezískáme žádnou informaci o $\mathcal{A}$. $\circledast 1$

V7. $H(\mathcal{A} \cdot \mathcal{B}) \le H(\mathcal{A}) + H(\mathcal{B})$ Intuice: entropie složeného jevu není větší než entropie dílčích jevů. $\circledast 2$

V8. Jsou-li $\mathcal{A}$ a $\mathcal{B}$ nezávislé, potom $\quad H(\mathcal{A} \cdot \mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B})$ $\circledast 1$

Vlastnosti III

V9. $H(\mathcal{A}\cdot\mathcal{B}) = H(\mathcal{B}) + H(\mathcal{A}\mid\mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B}\mid\mathcal{A})$

Intuice: Pozorujeme-li $\mathcal{B}$, získáme o $\mathcal{A}\cdot\mathcal{B}$ informaci $H(\mathcal{B})$, takže zbytková nejistota je $H(\mathcal{A}\mid\mathcal{B}) = H(\mathcal{A}\cdot\mathcal{B}) - H(\mathcal{B})$.

D:

1. dle definice se snažíme $H(\mathcal{A}\mid\mathcal{B})$ přepsat na $H(\mathcal{A}\cdot\mathcal{B}) - H(\mathcal{B})$
2. $P(B_j)H(\mathcal{A}\mid B_j) = -\sum_i \underbrace{P(B_j)P(A_i\mid B_j)}_{P(A_i,B_j)} \log \underbrace{P(A_i\mid B_j)}_{\frac{P(A_i,B_j)}{P(B_j)}} =$

$$\underbrace{-\sum_j \sum_i P(A_i,B_j)\log P(A_i,B_j)}_{H(\mathcal{A}\cdot\mathcal{B})} + \underbrace{\sum_j \sum_i P(A_i,B_j)\log P(B_j)}_{\sum_i P(A_i,B_j) = P(B_j)} \longrightarrow \sum_j P(B_j)\log P(B_j) = -H(\mathcal{B})$$

Vlastnosti IV

V10. $H(\mathcal{A}) - H(\mathcal{A} \mid \mathcal{B}) = H(\mathcal{B}) - H(\mathcal{B} \mid \mathcal{A})$ (plyne z V9)

V11. $H(\mathcal{B}) \leq H(\mathcal{A} \cdot \mathcal{B}) \leq H(\mathcal{A}) + H(\mathcal{B})$ (plyne z V9+V7)

V12. Pro každé $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ platí: (podmíníme V9 rozkladem $\mathcal{C}$)

$$H(\mathcal{A} \cdot \mathcal{B} \mid \mathcal{C}) = H(\mathcal{B} \mid \mathcal{C}) + H(\mathcal{A} \mid \mathcal{B} \cdot \mathcal{C}) = H(\mathcal{A} \mid \mathcal{C}) + H(\mathcal{B} \mid \mathcal{A} \cdot \mathcal{C})$$

V13. $0 \leq H(\mathcal{A} \mid \mathcal{B}) \leq H(\mathcal{A})$ (plyne z V9+V11)

Intuice:

1. Pozorováním $\mathcal{B}$ se nejistota o $\mathcal{A}$ nemůže zvětšit.
2. Entropie podmnožiny rozkladu je menší.

V14. Jestliže $\mathcal{B} \preceq \mathcal{C}$ potom $H(\mathcal{A} \mid \mathcal{B}) \leq H(\mathcal{A} \mid \mathcal{C})$. (plyne z V7 a vlastnosti 5 rozkladu MEJ.)

Intuice: Jemnějším rozkladem se dozvíme více o $\mathcal{A}$. $\circledast 2$

Vzájemná informace

Pozorování rozkladu $\mathcal{B}$ redukuje nejistotu o $\mathcal{A}$ z $H(\mathcal{A})$ na $H(\mathcal{A} \mid \mathcal{B})$ a získá tedy o $\mathcal{A}$ informaci $I(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}) - H(\mathcal{A} \mid \mathcal{B})$. srv. V9, kde jsme získávali inf. o $\mathcal{A} \cdot \mathcal{B}$ a příklad na str. 13

Def.

$$I(\mathcal{A}, \mathcal{B}) = H(\mathcal{A}) + H(\mathcal{B}) - H(\mathcal{A} \cdot \mathcal{B})$$ plyne z V9

Vlastnosti

$$I(\mathcal{A}, \mathcal{B}) = I(\mathcal{B}, \mathcal{A})$$ symetrie

$$I(\mathcal{A}, \mathcal{B}) \geq 0$$ nonnegativita, plyne z V7

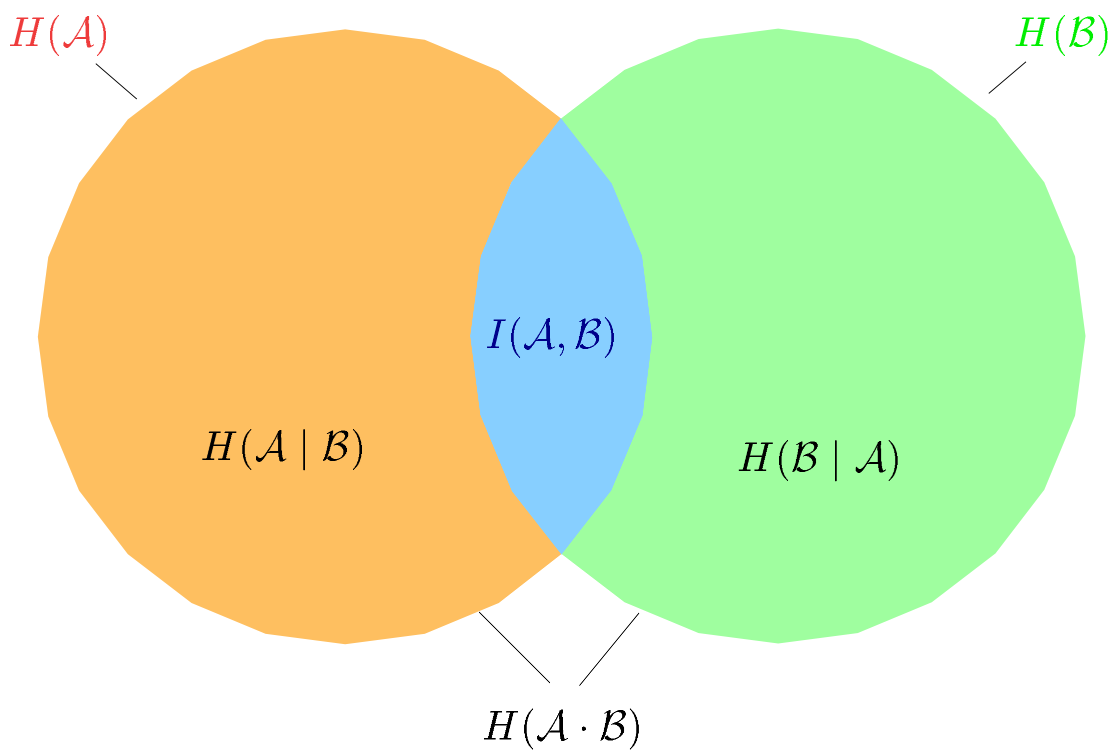
$I(\mathcal{A}, \mathcal{B})$

1. Informace o $\mathcal{A}$ obsažená v $\mathcal{B}$
2. Informace o $\mathcal{B}$ obsažená v $\mathcal{A}$

Pro více rozkladů:

$$I(\mathcal{A}_1, \mathcal{A}_2, \cdots, \mathcal{A}_k) = \sum_{i=1}^{k} H(\mathcal{A}_i) - H(\mathcal{A}_1 \cdot \mathcal{A}_2 \cdots \mathcal{A}_k)$$

Mnemotechnická pomůcka: Vztah mezi podmíněnými druhy entropie



Čteme takto: platí aditivita plochy, příčemž levý kruh představuje $H(\mathcal{A})$, pravý kruh $H(\mathcal{B})$ a jejich sjednocení $H(\mathcal{A} \cdot \mathcal{B})$. Potom

$$H(\mathcal{A}) + H(\mathcal{B} \mid \mathcal{A}) = H(\mathcal{A} \cdot \mathcal{B})$$

$$H(\mathcal{A} \cdot \mathcal{B}) - I(\mathcal{A} \cdot \mathcal{B}) = H(\mathcal{A} \mid \mathcal{B}) + H(\mathcal{B} \mid \mathcal{A})$$

$$\ldots$$

Entropie náhodné proměnné

Nechť X je diskrétní náhodná proměnná nabývající hodnot $x_i \in R(X)$ s pravděpodobnostmi

$$P(X = x_i) \stackrel{\text{def}}{=} p_i$$

1-D :

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Sjednocení událostí $\{X = x_i\}$ tvoří rozklad $\mathcal{E}_X$ (tj. pokrytí oboru hodnot $R(X)$).

Def. 1 Entropie $H(X)$ diskrétní náhodné proměnné X je rovna

$$\mu \approx \frac{1}{N}\sum_{i=1}^{N} x_i$$

odhad střední hodnoty

$$H(X) = H(\mathcal{E}_X) = -\sum_i p_i \ln p_i$$

$$S \approx \frac{1}{N-1}\sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T$$

odhad kovarianční matice

Def. 2 Diferenciální entropie $H(X)$ spojité náhodné proměnné X je

$$H(X) \stackrel{\text{def}}{=} -\int_{-\infty}^{\infty} f(x) \ln f(x)\, dx$$

n-D :

$$f(x) = \frac{1}{\sqrt{\det(2\pi S)}} e^{-\frac{1}{2}(x-\mu)^T S^{-1}(x-\mu)}$$

Pozn 1: události $\{X = x_i\}$ spojité X netvoří rozklad, jsou nespočetné

Pozn 2: pro spojitou X je $H(X) \in (-\infty, \infty)$

Pozn 3: rovnoměrné rozdělení na intervalu $\langle 0, a\rangle$: $H(x) = \ln a$, normální rozdělení: $H(x) = \ln \sigma\sqrt{2\pi e}$

⊛2 entropie multidimenzionálního normálního rozdělení

Sdružená a podmíněná entropie diskrétní náhodné proměnné

Sdružená entropie

$$H(X,Y) = H(\mathcal{A}_X \cdot \mathcal{A}_Y) = -\sum_i \sum_j p_{ij} \ln p_{ij}$$

Podmíněná entropie

$$H(X \mid Y) = H(\mathcal{A}_X \mid \mathcal{A}_Y)$$

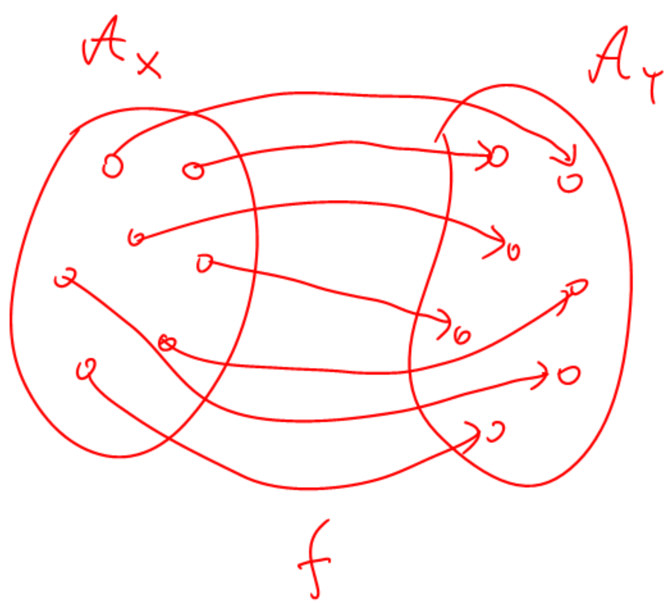
Věta o bijekci

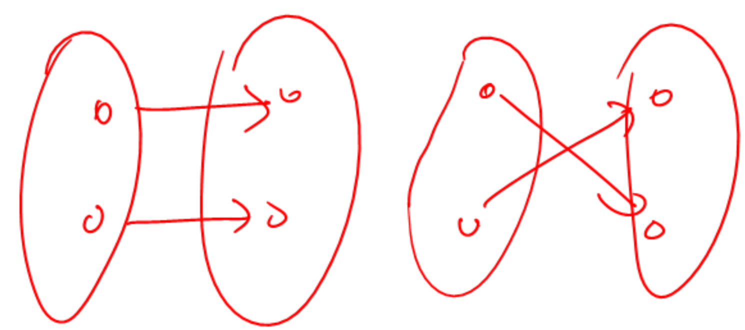
Jestliže $y = f(x)$ je zobrazení prosté a na (bijekce), potom

V15. $H(Y \mid X) = H(X \mid Y) = 0$

D: Rozklady $\mathcal{A}_X$ a $\mathcal{A}_Y$ jsou ekvivalentní, $\mathcal{A}_X = \mathcal{A}_Y$. Takže $\mathcal{A}_X \preceq \mathcal{A}_Y$ a platí V5.

V16. $I(X,Y) = H(Y)$, $\quad I(X,Y) = H(X)$. $\quad$ Také $H(X,Y) = H(X) = H(Y)$. (z V9)





Aplikace 1: Který poslanec má největší vliv na výsledek hlasování?

Systém: $\mathbf{S} = \{v_1, v_2, v_3, v_4, v_5\}$

Pozorování:

$$H_1 = H(v_2, v_3, v_4, v_5 \mid v_1) \text{ tím menší, čím větší je vliv } v_1$$

$$= H(v_1, v_2, v_3, v_4, v_5) - H(v_1) \qquad \text{(z V9)}$$

Výsledky:

v_i	H_i
v_1	2.6423
v_2	2.6292
v_3	2.6262
v_4	2.6310
v_5	2.6310

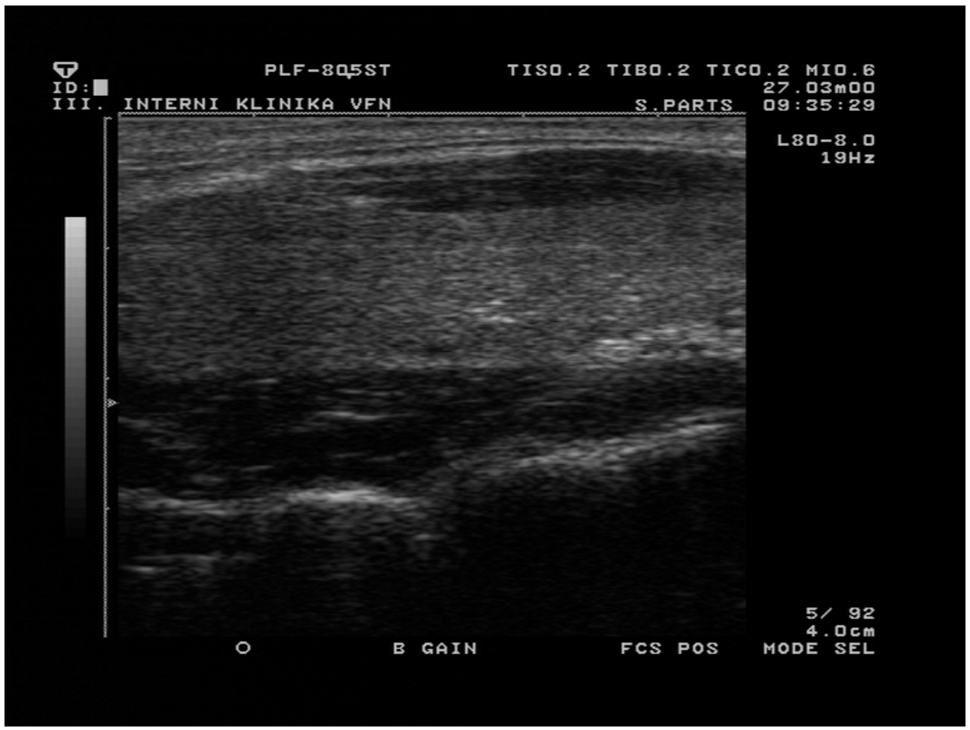
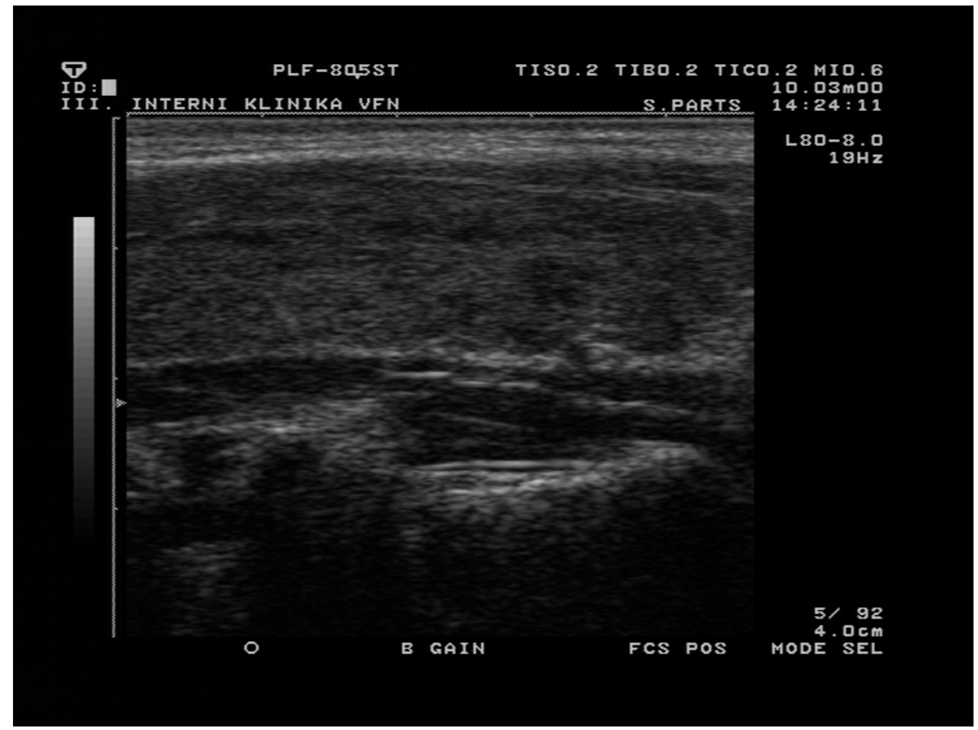
– nemá být červené

v_3 má největší vliv

Pozn: $H(v_1, v_2, v_3, v_4, v_5) = 3.3191$ [Nat]

Aplikace 2: Hledání informativních příznaků pro rozpoznávání

Sonogram štítné žlázy v podélném řezu

zdravá	lymfocitická thyroitida
	

Zajímá nás, kolik se lze z dat dozvědět o třídě c a „kde“ ta informace je.

Příznaky $\mathbf{x}_i$, $i = 1, 2, \ldots n$.

Informace o třídě v příznaku $\mathbf{x}_i$ je podmíněná entropie $H(c \mid \mathbf{x}_i)$.

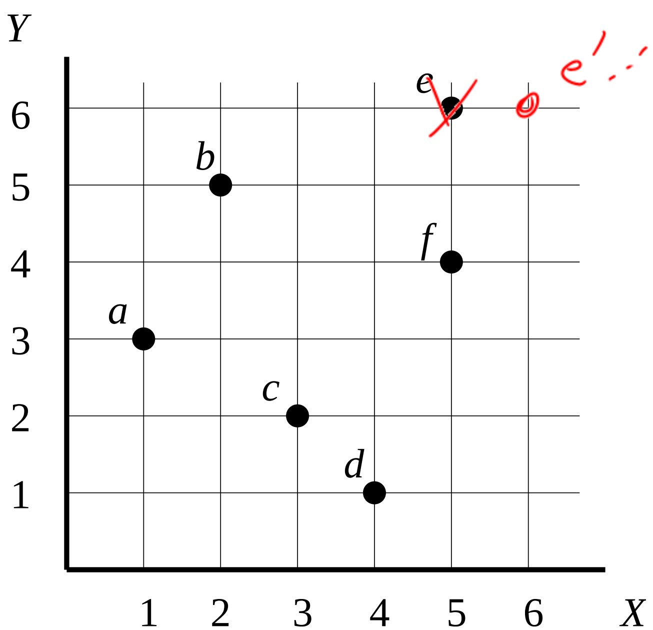
Aplikace 3: Normovaná střední vzájemná informace

$$i(X,Y) = \frac{I(X,Y)}{H(X,Y)}, \qquad 0 \le i(X,Y) \le 1$$

$i(X,Y) = 0$ pokud X, Y jsou nezávislé

$i(X,Y) = 1$ pokud mezi X, Y existuje bijekce

potom $i(X,Y) = 1$



$$X = \{1, 2, 3, 4, 5\}, \quad Y = \{3, 5, 2, 1, 6, 4\}$$

$$X \cdot Y \supseteqq \{(1,3), (2,5), (3,2), (4,1), (5,6), (5,4)\}$$

$$H(X) = 1.5607 \text{ nat}, \quad H(Y) = 1.7918 \text{ nat}$$

$$H(X,Y) = 1.7918 \text{ nat}, \quad I(X,Y) = 1.5607 \text{ nat}$$

normovaná stř. vzájemná inf. $i(X,Y) = 0.8710$

normalizovaný korelační koeficient $\rho(X,Y) = 0.1964$

Spearmanův rankový koeficient $r(X,Y) = 0.2609$

Dva aspekty relace mezi proměnnými

1. síla asociace $i(X,Y)$: jak moc jsou závislé?
2. statistická významnost této asociace : postačují data k takovému závěru?