MI-SPOL-5

Matematika neurčitosti: vzdálenosti a další míry podobnosti, fuzzy množiny a operace s nimi, t-normy a t-konormy, entropie a její souvislost s neurčitostí.

Vzdálenost souvisí s podobností: vzdálenost roste, podobnost klesá

Vzdálenost založená na normě vektoru: $d(x, y) = ||y - x||$

- $||x|| \geq 0$
- $||x|| = 0 \Leftrightarrow x = 0$
- $||\alpha x|| = |\alpha| \cdot ||x||$
- $||x + y|| \leq ||x|| + ||y||$

Druhy:

- obecná (Minkovského): $||x||_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}$, kde $p \in [1, \infty]$
- eukleidovská: $||x||_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} = \sqrt{x^T x}$
- manhattanská: $||x||_1 = \sum_{i=1}^{n} |x_i|$
- čebyševská: $||x||_\infty = \max_{i=1,\dots,n} |x_i|$

Mahalanobisova vzdálenost: vzdálenost realizací xy náhodných vektorů X, Y splňujících $EX = EY = \mu, varX = varY = \Sigma$ pozitivně definitní
Jde o eukleidovskou vzdálenost $\sqrt{\Sigma^{-1}}(x - \mu)$ od $\sqrt{\Sigma^{-1}}(y - \mu)$.

$$||\sqrt{\Sigma^{-1}}(x - \mu) - \sqrt{\Sigma^{-1}}(y - \mu)|| = \sqrt{(x - y)^T \Sigma^{-1} (x - y)}$$

Další míry podobnosti:
Podobnost náhodných veličin podle korelačních koeficientů:

- Pearsonův: $corr(X, Y) = \frac{cov(X,Y)}{\sqrt{varXvarY}}$
- Spearmanův: $\rho_{X,Y} = 12E\left(H(X, Y) - F(X)G(Y)\right)$
- Kendalův: $\tau_{X,Y} = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0]$

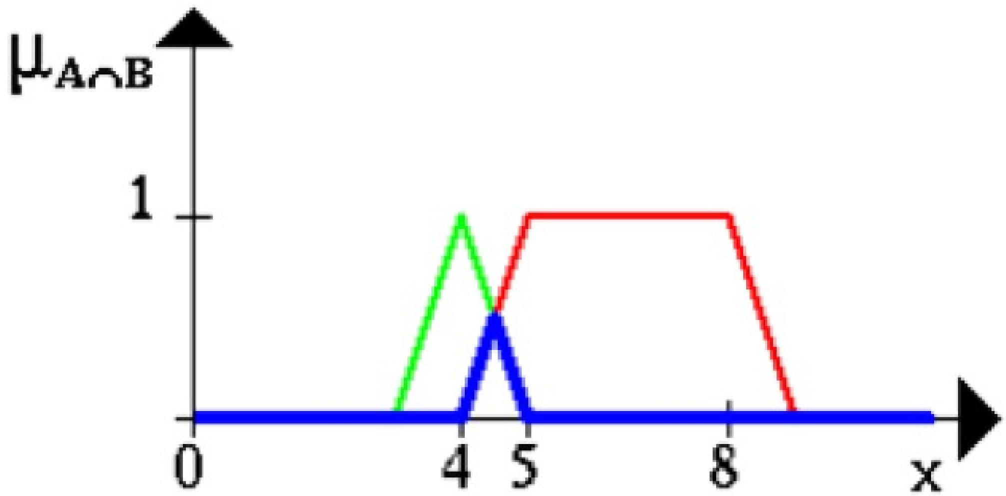
Podobnost binárních vektorů: Hammingova vzdálenost (počet lišících se bitů)

Fuzzy množiny

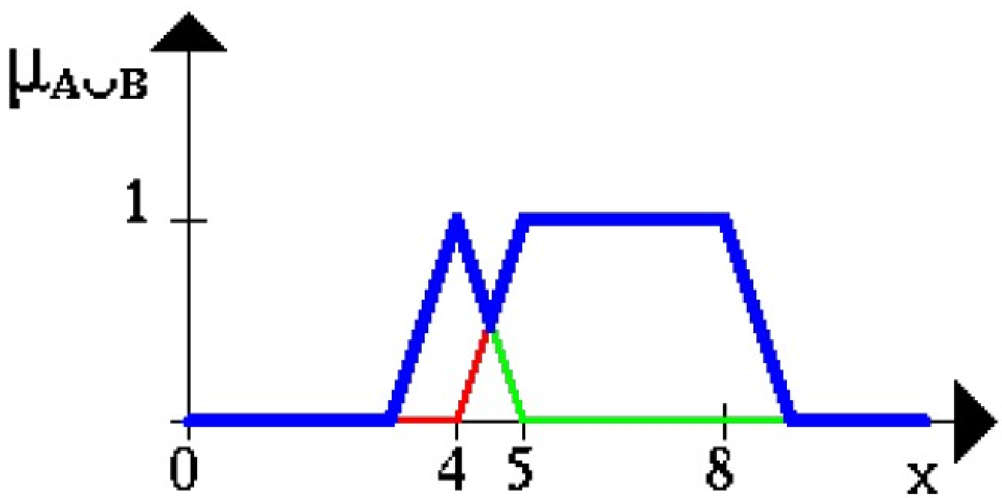
Fuzzy množina: $A = (U, \mu_A)$, kde A je množina a $\mu_A : U \to [0, 1]$ je funkce příslušnosti prvků univerza U k množině A
Stupeň příslušnosti x k A: $\mu_A(x)$
$\mu_A(x) > 0$: nosič A
$\mu_A(x) = 1$: jádro A

Operace s fuzzy množinami

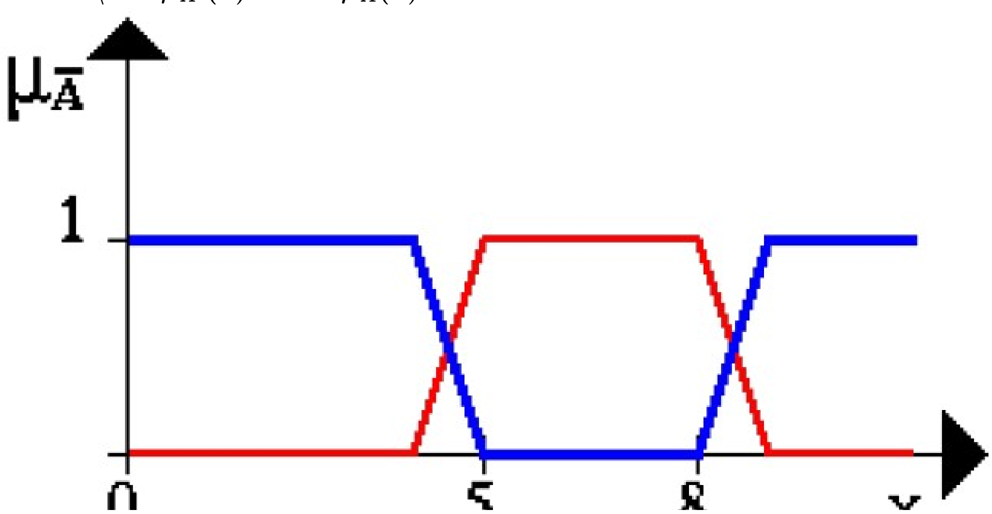
- $A \cap B : \mu_{A \cap B}(x) = \mu_A(x) \top \mu_B(x)$, kde $\top : [0,1]^2 \to [0,1]$ je **t-norma**



- $A \cup B : \mu_{A \cup B}(x) = \mu_A(x) \perp \mu_B(x)$, kde $\perp : [0,1]^2 \to [0,1]$ je **t-konorma**



- $A^c = U \setminus A : \mu_{A^c}(x) = 1 - \mu_A(x)$



T-normy

Vlastnosti:

- $u \top 0 = 0 \top v = 0$
- $u \top 1 = u$
- $1 \top v = v$
- Rostoucí v u, v
- $u > 0, v < 1 \Rightarrow 0 \leq u \top v \leq \min(u, v)$
- komutativní
- asociativní

Příklady:

- minimová (Gödelova): $u \top v = \min(u, v)$
- součinová: $u \top v = uv$
- Lukaszewiczova: $u \top v = \max(0, u + v - 1)$
- Hamacherova rodina: $H_p(u, v) = \frac{uv}{p+(1-p)(u+v-uv)}, p \geq 0$

T-konormy

Vlastnosti:

- $u \perp 1 = 1 \perp v = 1$
- $u \perp 0 = u$
- $0 \perp v = v$
- Rostoucí v u, v
- $u > 0, v < 1 \Rightarrow \max(u, v) \leq u \perp v \leq 1$
- komutativní
- asociativní

Příklady:

- maximová: $1 - \max(u, v) = \min(1 - u, 1 - v)$

Duální pár:

- $1 - u \top v = (1 - u) \perp (1 - v)$
- $1 - u \perp v = (1 - u) \top (1 - v)$
  (de Morganovy zákony)

Kopule: $C : [0, 1]^2 \rightarrow [0, 1]$ s vlastnostmi:

- $C(x, 0) = C(0, y) = 0$
- $C(x, 1) = x$
- $C(1, y) = y$
- $C(x_2, y_2) - C(x_2, y_1) - C(x_1, y_2) + C(x_1, y_1) \geq 0$

Kopule je někdy t-norma, vždy při $C(x, y) = f(f^{-1}(x) + f^{-1}(y))$, kde f je klesající konvexní

Sklarova věta: H je rozložení $(X, Y) \Leftrightarrow$ existuje kopule C splňující $H(X, Y) = C(F(X), G(Y))$ s pravděpodobností 1

Fuzzy zobrazení U do $\hat{U}$ pro $A = (U, \mu_A)$:
Zobrazení $u \rightarrow F_A(u) = (\hat{U}, \mu_{F_A(u)})$
Mamdaniho metoda konstrukce $F_A(u)$**:** na základě pravidel
$\text{if}(U, \mu_{A_1}) \text{ then } (\hat{U}, \mu_{B_1})..., \text{if}(U, \mu_{A_p}) \text{ then } (\hat{U}, \mu_{B_p})$
$F_A(u) = \bigcup_{i=1}^{p}(B_i \cap \mu_{A_i}(u))$ při Gödelových $\top, \perp$
$\mu_{F_A(u)}(\hat{u}) = \max(\min(\mu_{B_1}(\hat{\mu}), \mu_{A_1}(\hat{\mu})), ..., \min(\mu_{B_p}(\hat{\mu}), \mu_{A_p}(\hat{\mu})))$

Defuzzifikace: zobrazení fuzzy množiny do jejího univerza: $(U, \mu_A) \rightarrow U$

- Maximum μ_A
- Metoda těžiště: $u_{CG} = \frac{\int_S u\mu_A(u)du}{\int_S \mu_A(u)du}$

Fuzzy integrály: kombinace hodnot integrované funkce T-normami a T-konormami

- **Sugenův integrál:**
  $\int_U^{(S)} f dm = \max_i \min(f(x_{(i)}, m(A_i))),$
  Alternativní zápis: $\perp_i(f(x_{(i)} \top m(A_i))$
  kde $A_i = \{x_{(1)}, ..., x_{(k)}\}$, m je fuzzy míra
  ($m(U) = 1, A \subset B \Rightarrow m(A) < m(B)$)
- **Choquetův integrál:**
  $\int_U^{(C)} f dm = \sum_{i=1}^{k} f(x_{(i)})(m(A_i) - m(A_{i-1}))$

Entropie

Entropie diskrétní náhodné veličiny X**:** střední hodnota délky zápisu $P(x) = P(X = x)$ abedecou délky b

$$\mathcal{E}(X) = \sum_{x \in \mathcal{X}} P(x) \log_{\frac{1}{b}} P(x) = -\sum_{x \in \mathcal{X}} P(x) \log_b P(x)$$

Entropie spojité náhodné veličiny X**:**

$$\mathcal{E}(X) = -\int f(x) \log_b f(x)$$

, kde f je hustota X

Platí $0 \leq \mathcal{E} \leq \log_b n$

Minimální entropie: $\mathcal{E} = 0 \Leftrightarrow P(x_i) = 1$ pro nějaké x_i.
Žádná nejistota

Maximální entropie: $\mathcal{E} = \log_b n \Leftrightarrow P(x_1) = ... = P(x_n) = 1/n.$
Největší nejistota

Relativní entropie kódování při rozdělení $(Q(x))_{x \in \mathcal{X}}$ vůči $(P(x))_{x \in \mathcal{X}}$

$$D(P||Q) = -\sum_{x \in \mathcal{X}} P(x) \log_b Q(x) - \mathcal{E}(X) = \sum P(x) \log_b \frac{P(x)}{Q(x)}$$

Vzájemná informace X a Y:

$$I(X,Y) = D(P_{(X,Y)}||P_X P_Y) = \sum P(X=x, Y=y) \log_b \frac{P(X=x, Y=y)}{P(x)P(y)} = \sum_{x \in \mathcal{X}} \sum_{yiny} P(x,y) \log_b \frac{P(x,y)}{P(x)P(y)}$$