

Vzdálenostní metriky používané pro měření vzdáleností mezi obrázky

Pavel Jakš

6. prosince 2022

Úvod

Pod pojmem metrika na prostoru X si každý matematik představí zobrazení $\rho : X \times X \rightarrow [0, +\infty)$ splňující

1. $\rho(x, y) = 0 \iff x = y \quad \forall x, y \in X$,
2. $\rho(x, y) = \rho(y, x) \quad \forall x, y \in X$,
3. $\rho(x, z) \leq \rho(x, y) + \rho(y, z) \quad \forall x, y, z \in X$.

Taková metrika může být na lineárním prostoru V nad číselným tělesem (pro naše účely zůstaňme nad \mathbb{R}) snadno zadána pomocí normy, která je buď indukována skalárním součinem v případě pre-Hilbertových prostorů, nebo dána vlastnostmi, že se jedná o zobrazení $\|\cdot\| : V \rightarrow [0, +\infty)$ a splňuje:

1. $\|x\| = 0 \iff x = 0 \quad \forall x \in V$,
2. $\|\alpha x\| = |\alpha| \cdot \|x\| \quad \forall \alpha \in \mathbb{R}, \forall x \in V$,
3. $\|x + y\| \leq \|x\| + \|y\| \quad \forall x, y \in V$.

S metrikami, které jsou tzv. indukované normami se setkáme.

1 Metriky indukované l_p normami

Vzhledem k tomu, že obrázky, které jsou středem naší pozornosti, lze reprezentovat jako tenzory standardně o rozměrech $C \times W \times H$, kde C značí počet kanálů (nejčastěji kanály po řadě pro červenou, zelenou a modrou barvu), W označuje šířku a H výšku, tak lze na tyto tenzory vpustit L^p normy. Pro $p \in [1, +\infty)$ je L^p norma z $f \in L_p(X, \mu)$ definována vztahem:

$$\|f\|_p = \left(\int_X |f|^p d\mu \right)^{\frac{1}{p}}.$$

Pro naše obrázky lze za X vzít $\{1, \dots, C\} \times \{1, \dots, W\} \times \{1, \dots, H\}$ a za μ *počítací míru*. Potom naše L^p norma přejde v l_p normu, která má pro naše obrázky, tedy tenzory $x \in \mathbb{R}^{C \times W \times H}$, tvar:

$$\|x\|_p = \left(\sum_{i=1}^C \sum_{j=1}^W \sum_{k=1}^H |x_{i,j,k}|^p \right)^{\frac{1}{p}}. \quad (1)$$

Trochu mimo stojí l_∞ norma, která má tvar pro tenzor $x \in \mathbb{R}^{C \times W \times H}$:

$$\|x\|_\infty = \max_{i \in \{1, \dots, C\}} \max_{j \in \{1, \dots, W\}} \max_{k \in \{1, \dots, H\}} |x_{i,j,k}|. \quad (2)$$

A úplně mimo stojí L_0 norma, která svou povahou *není* norma ve smyslu výše uvedené definice, ale pro účely porovnávání obrázků se používá rozdíl obrázků v této pseudo-normě, proto ji zde zmiňuji:

$$\|x\|_0 = |\{x_{i,j,k} \neq 0\}|. \quad (3)$$

2 MSE a RMSE

Vzdálenosti, které mají blízko k metrikám indukovaným l_2 normou, jsou *MSE* (z anglického *Mean Squared Error*) a *RMSE* (z anglického *Root Mean Squared Error*). Pro tenzory $x, \tilde{x} \in \mathbb{R}^{C \times W \times H}$ mají definici:

$$\text{MSE}(x, \tilde{x}) = \frac{1}{CWH} \sum_{i=1}^C \sum_{j=1}^W \sum_{k=1}^H |x_{i,j,k} - \tilde{x}_{i,j,k}|^2 \quad (4)$$

$$\text{RMSE}(x, \tilde{x}) = \left(\frac{1}{CWH} \sum_{i=1}^C \sum_{j=1}^W \sum_{k=1}^H |x_{i,j,k} - \tilde{x}_{i,j,k}|^2 \right)^{\frac{1}{2}} \quad (5)$$

3 Wassersteinova vzdálenost

Bud' (M, d) metrický prostor, který je zároveň *Radonův*. Zvolme $p \in [1, +\infty)$. Potom máme *Wassersteinovu p -vzdálenost* mezi dvěma pravděpodobnostními mírami μ a ν na M , které mají konečné p -té momenty, jako:

$$W_p(\mu, \nu) = \left(\inf_{\gamma \in \Gamma(\mu, \nu)} \mathbb{E}_{(x,y) \sim \gamma} d(x, y)^p \right)^{\frac{1}{p}}, \quad (6)$$

kde $\Gamma(\mu, \nu)$ je množina všech sdružených pravděpodobnostních měr na $M \times M$, které mají po řadě μ a ν za marginální pravděpodobnostní míry [1].

Jak to souvisí s obrázky? Přeš dopravní problém. Pod pravděpodobnostní distribucí μ či ν na X si lze představit rozložení jakési hmoty o celkové hmotnosti 1. Sdružená rozdělení $\gamma \in \Gamma(\mu, \nu)$ potom odpovídají transportnímu plánu, kde

$\gamma(x, y) dx dy$ vyjadřuje, kolik hmoty se přesune z x do y . Tomu lze přiřadit nějakou cenu C , totiž kolik stojí přesun jednotkové hmoty z x do y : $c(x, y)$. V případě *Wassersteinovy vzdálenosti* za cenu dosadíme $c(x, y) = d(x, y)^p$, tedy p -tou mocninu vzdálenosti mezi x a y . Potom cena celkového dopravního problému s transportním plánem γ bude:

$$c_\gamma = \int c(x, y) \gamma(x, y) dx dy \quad (7)$$

$$= \int c(x, y) d\gamma(x, y) \quad (8)$$

a optimální cena bude:

$$c = \inf_{\gamma \in \Gamma(\mu, \nu)} c_\gamma. \quad (9)$$

Po dosazení:

$$c = \inf_{\gamma \in \Gamma(\mu, \nu)} \int c(x, y) d\gamma(x, y) \quad (10)$$

$$= \inf_{\gamma \in \Gamma(\mu, \nu)} \int c(x, y) \gamma(x, y) dx dy \quad (11)$$

$$= \inf_{\gamma \in \Gamma(\mu, \nu)} \mathbb{E}_{(x, y) \sim \gamma} c(x, y) \quad (12)$$

$$= \inf_{\gamma \in \Gamma(\mu, \nu)} \mathbb{E}_{(x, y) \sim \gamma} d(x, y)^p \quad (13)$$

$$= W_p(\mu, \nu)^p \quad (14)$$

Dostáváme tedy interpretaci, že p -tá mocnina *Wassersteinovy vzdálenosti* odpovídá ceně dopravního problému.

Pro obrázky má tato konstrukce následující uplatnění: Obrázky je třeba chápat jako diskrétní pravděpodobnostní rozdělení, proto je třeba je normalizovat, aby součet prvků tenzoru obrázku byl roven 1. Pak střední hodnota v definici *Wassersteinovy vzdálenosti* přejde ve váženou sumu cen, tedy p -tých mocnin vzdáleností mezi jednotlivými pixely.

Jak je to barevnými obrázky?

4 PSNR

Vzdálenost označená zkratkou *PSNR* z anglického *Peak Signal-to-Noise Ratio* vyjadřuje vztah mezi obrázkem $x \in \mathbb{R}^{C \times W \times H}$ a jeho pokažením $\tilde{x} \in \mathbb{R}^{C \times W \times H}$ za přidání šumu. Definice je následující:

$$\text{PSNR}(x, \tilde{x}) = 10 \cdot \log_{10} \left(\frac{l^2}{\text{MSE}(x, \tilde{x})} \right), \quad (15)$$

$$= 20 \cdot \log_{10} \left(\frac{l}{\text{RMSE}(x, \tilde{x})} \right), \quad (16)$$

kde l je dynamický rozsah obrázků, tedy rozdíl mezi maximální možnou hodnotou pixelů a minimální možnou hodnotou pixelů. Jedná se tedy o transformaci metriky *MSE*.

5 SSIM

Zkratka *SSIM* pochází z anglického *structural similarity index measure*. Tato metrika se při výpočtu indexu dvou obrázků x a \tilde{x} dívá na podokna, ze kterých vybere jisté statistiky a z nich vytvoří index pro daná podokna obrázků. Potom se jako celkový index bere průměr přes tato okna. Uveďme vzorce pro výpočet indexu SSIM pro případ, že máme jediné okno, které splývá s obrázkem, které pro jednoduchost zvolme jednokanálové, tedy černobílé. Označme $N = W \times H$ počet pixelů v obrázku a indexujme prvky matice obrázku jediným číslem. Potom definujeme pro obrázky x a \tilde{x} následující:

$$\begin{aligned}\mu_x &= \frac{1}{N} \sum_{i=1}^N x_i, \\ \mu_{\tilde{x}} &= \frac{1}{N} \sum_{i=1}^N \tilde{x}_i, \\ \sigma_x^2 &= \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2, \\ \sigma_{\tilde{x}}^2 &= \frac{1}{N-1} \sum_{i=1}^N (\tilde{x}_i - \mu_{\tilde{x}})^2, \\ \sigma_{x\tilde{x}} &= \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(\tilde{x}_i - \mu_{\tilde{x}}).\end{aligned}$$

Potom:

$$\text{SSIM}(x, \tilde{x}) = \frac{(2\mu_x\mu_{\tilde{x}} + C_1)(2\sigma_{x\tilde{x}} + C_2)}{(\mu_x^2 + \mu_{\tilde{x}}^2 + C_1)(\sigma_x^2 + \sigma_{\tilde{x}}^2 + C_2)}, \quad (17)$$

kde C_1, C_2 jsou konstanty pro stabilitu dělení volené kvadraticky úměrně dynamickému rozsahu.

Jak volíme celkový SSIM pro barevné obrázky? Jako průměr přes kanály.

Reference

- [1] L. Vaserstein, *Markov processes over denumerable products of spaces, describing large systems of automata*. Problemy Peredači Informacii 5, 1969.