

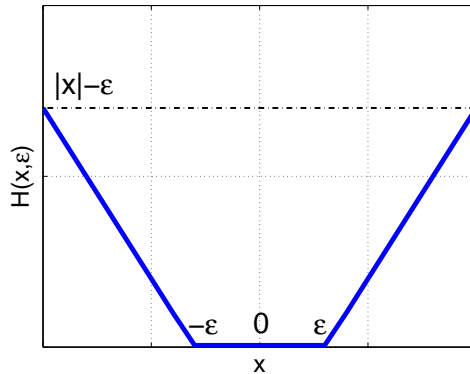
Robustní prokládání bodů přímkou pomocí Lineárního Programování

Karel Zimmermann

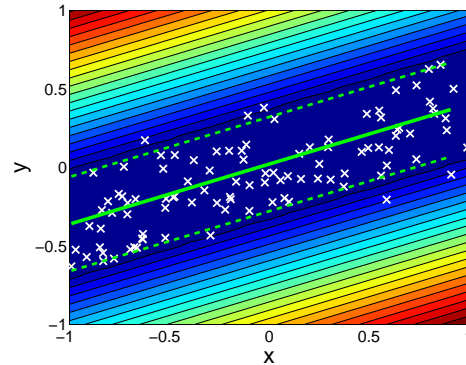
Úvod

Jistě si z minulých cvičení a přednášek pamatujete, že L_∞ a L_2 normy jsou velmi citlivé na chybná měření (tzv. "outliery"). V tomto cvičení se naučíme jak robustně prokládat body přímkou pomocí tzv. ϵ -necitlivé ztrátové funkce (ϵ -insensitive loss function).

Úkoly k vypracování



(a) ϵ -necitlivé ztrátové funkce.



(b) Prokládání bodů přímkou.

Obrázek 1: Prokládání bodů pomocí ϵ -necitlivé ztrátové funkce

1. Vaším prvním úkolem je matematicky zapsat ϵ -necitlivou ztrátovou funkci

$$H(x, \epsilon) : \mathcal{R}^2 \rightarrow \mathcal{R},$$

pro náš účel definovanou Obrázkem 1a, jako maximum třech lineárních funkcí. Správnost zápisu ověřte vykreslením funkce v MATLABu.

Všimněte si že tato ztrátová funkce nepenalizuje malé chyby vůbec a větší chyby jsou penalizovány pouze lineárně, což je velmi žádoucí vlastnost. Pomocí takto zavedené ztrátové funkce, definujeme ϵ -necitlivou normu m -rozměrného vektoru $\mathbf{x} \in \mathcal{R}^m$ jako:

$$\|\mathbf{x}\|_{H,\epsilon} = \sum_{i=1}^m H(x_i, \epsilon)$$

2. Stejně jako v předcházejících cvičení, naměřená data jsou m -tice dvojrozměrných bodů

$$T = \left[\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}, \dots, \begin{pmatrix} x_m \\ y_m \end{pmatrix} \right] = \begin{pmatrix} \mathbf{x}^\top \\ \mathbf{y}^\top \end{pmatrix},$$

jimiž chceme proložit přímkou $P(a, b) = \{(x, y) \mid y = a \cdot x + b, x, y \in \mathcal{R}\}$ zadanou parametry $a, b \in \mathcal{R}$. Tentokrát však budeme minimalizovat součet y -ových vzdáleností v ϵ -necitlivé normě, tj. hledáme

$$(a^*, b^*) = \arg \min_{a, b \in \mathcal{R}} \|a \cdot \mathbf{x} + b - \mathbf{y}\|_{H,\epsilon} = \arg \min_{a, b \in \mathcal{R}} \sum_{i=1}^m H(a \cdot x_i + b - y_i, \epsilon). \quad (1)$$

Napište lineární program odpovídající výše uvedenému problému a vyřešte pomocí funkce `linprog` v MATLABu. Výsledky pro data ze druhého cvičení (<http://cmp.felk.cvut.cz/cmp/courses/OPT/cviceni/02/data1.mat>) vizualizujte, např. jako obr. 1b. Zkuste jak se výslkedek mění pro různá ϵ a různé množiny bodů (body můžete samy jednoduše naklikat pomocí příkazu `ginput`).

Bonusové úlohy:

- Dokažte, že ϵ -necitlivá norma je konvexní funkce.
- Jaké body mají nulovou penalizaci?
- Jak souvisí ϵ -necitlivá norma s normou L_1 ?
- Proč je ϵ -necitlivá norma méně citlivá na outliery než norma L_2 ?