Jméno:

Konvexní optimalizace

Sada 9 domácích úkolů

Termín odevzdání: 5. prosince 2017 ve 12:21

Všechna svá řešení zdůvodněte.

Problém	Bodů max	Bodů
1	2	
2	2	
3	3	
4	3	
Σ	10	

Problém 1. Zformulujte duální problém k problémům:

a) Nejmenší čtverce přímočaré

$$\text{minimalizujte} \quad \|A\mathbf{x} - \mathbf{b}\|_2$$

b) Nejmenší čtverce upravené (pozor na to, že $\|\mathbf{y}\|_2 \neq \mathbf{y}^T\mathbf{y}$; duál vyjde jinak než pro kvadratickou účelovou funkci):

$$\begin{array}{rl} \text{minimalizujte} & \|\mathbf{y}\|_2 \\ \text{za podmínek} & \mathbf{y} = A\mathbf{x} - \mathbf{b} \end{array}$$

Problém 2. Buď A matice $2 \times n$ jejíž oba řádky jsou „náhodné“ v tom smyslu, že pro všechny volby indexů $0 < i < j < k \leq n$ a všechny volby $u_i, u_j, u_k \in \{-1, 1\}$ má matice

$$\begin{pmatrix} a_{1i} & a_{1j} & a_{1k} \\ a_{1i} & a_{2j} & a_{2k} \\ u_i & u_j & u_k \end{pmatrix}$$

lineárně nezavislé řádky. Dokažte, že pak optimální řešení $\mathbf{x}^\star$ problému

$$\begin{aligned} \text{minimalizujte} \quad & \|\mathbf{x}\|_1 \\ \text{za podmínek} \quad & A\mathbf{x} = \mathbf{b} \end{aligned}$$

má nejvýše dvě nenulové složky $x_i^\star \neq 0$.

Rada: Přepište si problém do ekvivalentní podoby

$$\begin{aligned} \text{minimalizujte} \quad & \mathbf{1}^T\mathbf{x}^+ + \mathbf{1}^T\mathbf{x}^- \\ \text{za podmínek} \quad & A\mathbf{x}^+ - A\mathbf{x}^- = \mathbf{b} \\ & \mathbf{x}^+, \mathbf{x}^- \succeq 0 \end{aligned}$$

a podívejte se na duální problém a komplementaritu.

Problém 3. Zvolme si $M > 0$ parametr. Dokažte, že robustní nejmenší čtverce, tj. regrese vůči Huberově penalizační funkci s parametrem M

$$\begin{array}{rl} \text{minimalizujte} & \displaystyle\sum_{i=1}^{m} \phi(u_i) \\ \text{za podmínek} & \mathbf{u} = A\mathbf{x} - \mathbf{b} \end{array}$$

(kde $\phi(u) = u^2$ pro $|u| \leq M$ a $\phi(u) = 2M|u| - M^2$ pro $|u| > M$) je ekvivalentní kvadratickému programu

$$\begin{array}{rl} \text{minimalizujte} & \mathbf{u}^T\mathbf{u} + 2M\displaystyle\sum_{i=1}^{m} v_i \\ \text{za podmínek} & A\mathbf{x} - \mathbf{b} \succeq -\mathbf{u} - \mathbf{v} \\ & A\mathbf{x} - \mathbf{b} \preceq \mathbf{u} + \mathbf{v} \\ & \mathbf{u}, \mathbf{v} \succeq 0. \end{array}$$

Problém 4. Goblini chtějí konkurovat trpaslíkům v těžbě uhlí. Je opět rozumné předpokládat, že množství uhlí, které se za den v gobliním dole vytěží, je zhruba součet produktivity všech přítomných goblinů a že produktivita daného goblina je konstantní.

Na rozdíl od trpaslíků mají goblini v elektronické evidenci těžeb záznamy o docházce a těžbě za posledních 100 dní. Také na rozdíl od trpaslíků jsou ale záznamy ze zhruba 10 % dní (docházky i celkového vytěženého uhlí) zfalšované. Goblin, co to udělal, se snažil vylepšit si uměle výkonnost a falešné záznamy by se měly chovat jako odlehlé hodnoty.

1. Odhadněte produktivitu jednotlivých goblinů pomocí minimalizace vůči Hubnerově penalizační funkci s parametrem $M = 1$ (při implementaci si můzete pomoci řešením předchozí úlohy). Svůj program používající CVXOPT/CVXPY mi pošlete na mail kazda@karlin.mff.cuni.cz a tamtéž mi pošlete i vektor odhadnutých produktivit.

2. Vymyslete, jak na základě odhadu produktivity identifikovat odlehlé hodnoty, tj. cca 10 dní se zfalšovanými záznamy. Tato druhá část nemá jednoznačné řešení. Nemusíte ale vyrábět složité statistické testy – navrhněte postup, který lze zdůvodnit a dává příčetné výsledky. Čísla dnů s podezřelými záznamy mi napište do řešení.

3. Odstraňte podezřelé záznamy ze vstupních dat, znovu spočtěte produktivity a okomentujte, jak moc se změnil odhad produktivit po eliminaci odlehlých hodnot. (Svůj program mi už nemusíte posílat.)

Gobliní záznamy najdete v souboru goblini.csv na webu cvičení.

Při rešení úloh je možné se poradit s dalšími lidmi (nejlépe dalšími studenty a studentkami Konvexní optimalizace), ale svá řešení (včetně programů!) *pište samostatně* a před termínem odevzdání úloh sepsaná řešení (a programy) nikomu *neukazujte*.