

Sada 1 (deadline 1.4. 23:59)

1.1 Snímání obrázků

Přišla vám data z přístroje, který má 67 dvojic protilehlých vertikálních detektorů, 67 dvojic horizontálních protilehlých detektorů a pojme obrázek o rozměrech 338x338. Tyto detektory jsou rozmístěné rovnoměrně po obvodu (tj. tvoří mřížku 67x67). Měření u dvojice protilehlých detektorů probíhá tak, že sečte hodnoty (v rozsahu 0-255) pixelů nacházející se na přímce (šířka 1px) mezi těmito detektory. V záznamu jsou vždy prvně data z horizontálních detektorů a následují data z vertikálních detektorů.

Dělali jsme dvě měření – na jednom obrázku byla jednička a na druhém dvojka. Dokážete tyto obrázky zrekonstruovat? Které měření snímalo obrázek s dvojkou? Vyzkoušejte stručně popsat nevýhody tohoto rozdělení detektorů a navrhněte, jakým způsobem byste umístili další (řádově jednotky dvojic).

- dataA: <http://adolf.streda.matfyz.cz/2018/1a/du1/dataA.csv>
- dataB: <http://adolf.streda.matfyz.cz/2018/1a/du1/dataB.csv>

Hint: Rekonstruuje obrázek v rozlišení 67x67 pixelů.

1.2 SPAM

Rozhodli jste se šířit spam, bohužel vaše velkolepé plány zmařil spamfilter. Naštěstí spamfilter v hlavičce každé zprávy poznamená její skóre. Víte, že tento spamfilter používá jednoduchý model, který převede zprávu na vektor frekvencí slov za pomoci slovníku. Spamfilter je dále určený nenulovým vektorem $a \in \mathbb{R}^n$ a konstantou $b \in \mathbb{R}$. Zprávě reprezentované vektorem $x \in \mathbb{R}^n$ spamfilter přiřadí skóre $a^T x + b$ a vyhodnotí ji jako spam, pokud je skóre zprávy kladné.

Aby tyto parametry nešly snadno vyčíst, je skóre zašuměno šumem s normálním rozdělením o střední hodnotě 0 a variancí 0,1. Zrekonstruuje z již vektorizovaných e-mailů *emails.csv* a jejich skóre *score.csv* odhad parametrů spamfilteru a dokažte, že hodnota variance nemá vliv na spočtený odhad parametrů. Za pomoci parametrů a slovníku *dict.csv* vytvořte zprávu na způsob dopisů Nigerijského prince (<http://www.hoax.cz/scam419/>), která projde spamfilterem jako validní zpráva.

- emails: <http://adolf.streda.matfyz.cz/2018/1a/du1/emails.csv>
- score: <http://adolf.streda.matfyz.cz/2018/1a/du1/score.csv>
- dict: <http://adolf.streda.matfyz.cz/2018/1a/du1/dict.csv>

Pozn.: Upravena formulace skóre a popis spamfilteru.

1.3 Normy

Definice. Pro matici $A = (a_{ij}) \in \mathbb{R}^{m \times n}$ definujeme Frobeniovu normu $\|A\|_F$ předpisem $\|A\|_F^2 = \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2$.

Pro čtvercovou matici $U = (u_{ij}) \in \mathbb{R}^{n \times n}$ definujeme stopu $\text{tr}(U)$ matice U předpisem $\text{tr}(U) = \sum_{i=1}^n u_{ii}$.

Dokažte:

1. $\|A\|_F^2 = \text{tr}(A^T A) = \text{tr}(A A^T)$.
2. Jestliže $B \in \mathbb{R}^{m \times n}$, pak $\text{tr}(AB) = \text{tr}(BA)$.
3. Dokažte, že mají-li obě matice $U \in \mathbb{R}^{k \times m}$, $V \in \mathbb{R}^{l \times n}$ ortonormální posloupnosti sloupcových vektorů, pak $\|UAV^T\|_F = \|A\|_F$

1.4 Semidefinitnost

Definice. Pro semipozitivně definitní matice $U, V \in \mathbb{R}^{n \times n}$ definujeme $U \preceq V$ právě když je $V - U$ pozitivně semidefinitní matice.

1. Dokažte, že $\preceq$ je částečné uspořádání na prostoru semidefinitních matic.
2. Dokažte, že je-li $U \preceq V$, pak platí $\text{tr}(U) \leq \text{tr}(V)$.
3. Dokažte BLUE property pro klasické nejmenší čtverce (tj. bez váhy): je-li posloupnost sloupcových vektorů matice $A \in \mathbb{R}^{m \times n}$ lineárně nezávislá, pak pro matici $L_0 = (A^T A)^{-1} A^T$ a pro každý levý inverz L matice A platí $\|L_0\|_F^2 \leq \|L\|_F^2$.

Hint: použijte BLUE property obecné metody nejmenších čtverců dokázanou na přednášce.