

1

neřešil jsem

2

Označme si vektor skóre zpráv písmenem $s \in \mathbb{R}^{300}$, pro každou zprávu i její vektor četnosti slov písmenem $x_i \in \mathbb{R}^{250}$. Z dat vídíme, že máme celkem 300 zpráv a k nim jejich skóre. Dále vídíme že hledaný vektor a má 250 složek, tedy $a \in \mathbb{R}^{250}$. Nakonec označme e_i vektor šumu pro dané skóre. Ze zadání je zřejmé, že e_i pochází ze stejného pravděpodobnostního rozdělení. Rovnice tedy vypadají:

$$s_i = a^T \cdot x_i + b + e_i \quad (1)$$

Chceme nejdříve odhadnout a , takže se chceme zbavit neznámé b . To provedeme jednoduše například tak, že si pevně zvolíme dvojici s_i, x_i a od té odečteme ostatní rovnice. Zvolíme-li $i = 1$, tak máme nově rovnice:

$$\forall i \in \{2, \dots, 300\} : s_1 - s_i = a^T \cdot (x_1 - x_i) + e'_i$$

Kde $e'_i = e_1 - e_i$ je z vlastností součtu normálních rozdělení z rozdělení se střední hodnotou 0 a rozptylem 0.2. Soustavu si můžeme napsat následovně:

$$\begin{aligned} s_1 - s_2 &= (x_1 - x_2)^T \cdot a + e'_1 \\ &\vdots \\ s_1 - s_{300} &= (x_1 - x_{300})^T \cdot a + e'_{299} \\ \implies \\ \begin{pmatrix} s_1 - s_2 \\ \vdots \\ s_1 - s_{300} \end{pmatrix} &= \begin{pmatrix} (x_1 - x_2)^T \\ \vdots \\ (x_1 - x_{300})^T \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ \vdots \\ a_{250} \end{pmatrix} + \begin{pmatrix} e'_1 \\ \vdots \\ e'_{299} \end{pmatrix} \end{aligned}$$

Máme tedy klasickou rovnici pro problém nejmenších čtverců: $y = A \cdot x + e$. Stačí tedy spočítat pseudoinverzi matice A (matice kde jsou řádky rozdíly vektorů x) a vynásobit ho s vektorem s . Jelikož chyby e'_i jsou všechny ze stejného pravděpodobnostního rozdělení, tak víme, že toto řešení bude nejlepší možné. Nezáleží tedy na velikosti rozptylu daného rozdělení.

Dostaneme tedy odhad vektoru a . Dokážeme z každé původní rovnice (1) dostat odhad $b - e_i \forall i \in \{1, \dots, 300\}$. Nezbyvá tedy nic jiného než spočítat průměr těchto hodnot a dostaneme odhad b .

Protože všechny e_i pocházejí ze stejného normálního rozdělení se střední hodnotou 0, předpokládá se tedy, že se šum v průměru sám od sebe odečte, jelikož jeho střední hodnota je 0.

Výsledné odhady jsou:

$$\begin{aligned}
 b &= 0.65705 \\
 a &= (-0.469225, 0.0449372, -0.619684, -0.16555, 0.667982, -1.05396, \\
 &\quad 0.535969, 0.464577, -0.304336, -0.744404, -0.534875, -0.790616, \\
 &\quad \dots \\
 &\quad 0.625111, -10.9927, 3.74909, -1.61924, -1.8207, 0.052431, -0.870726, \\
 &\quad -0.342373, -2.66895, -0.791671, -0.158633, -0.455493)^T
 \end{aligned}$$

Email, který by měl projít spam filterem je například: *"Hi, my prey. This is my last warning. I write you since I embed a trojan on the web site with porn which you have visited. My trojan captured all your private data and switched on your camera which recorded the act of your solitary sex. Just after that the trojan saved your contact list. I will erase the compromising video records and information if you send me 1000 EURO in bitcoin. This is address for payment - 1taeJZbDZnhrk46nmDaG3GkQBdEQf9s49 (If you don't know what bitcoin / write to buy bitcoin in Google) I give you 30 hours after you open my message for making the payment. As soon as you read the message I'll see it right away. It is not necessary to tell me that you have sent money to me. This address is connected to you, my system will erased automatically after transfer confirmation. If you need 48h just Open the calculator on your desktop and press +++ If you don't pay, I'll send dirt to all your contacts. Let me remind you-I see what you're doing! You can visit the pulicc office but anybody can't help you. If you try to deceive me , I'll know it immediately! I don't live in your country. So anyone can not track my location even for 18 months. bye. Don't forget about the shame and to ignore, Your life can be ruined."*

Její skóre mi vyšlo ≈ 0.875946 .

3

3.1

Z definice násobení matic a vlastností sum:

$$\begin{aligned} \text{tr}(AA^T) &= \sum_{i=1}^m \left(\sum_{j=1}^n a_{i,j} \cdot a_{i,j} \right) = \sum_{j=1}^n \left(\sum_{i=1}^m a_{i,j} \cdot a_{i,j} \right) = \text{tr}(A^T A) \\ &\implies \\ \sum_{j=1}^n \left(\sum_{i=1}^m a_{i,j} \cdot a_{i,j} \right) &= \sum_{j=1}^n \sum_{i=1}^m a_{i,j}^2 = \|A\|_F^2 \end{aligned}$$

kde i je index řádku ve výsledné matici a v závorce je prvek na místě i, i ve výsledné matici.

3.2

Obdobně jako 3.1

$$B \in \mathbb{R}^{n \times m}, \text{tr}(AB) = \sum_{i=1}^m \left(\sum_{j=1}^n a_{i,j} \cdot b_{i,j} \right) = \sum_{i=1}^m \left(\sum_{j=1}^n b_{i,j} \cdot a_{i,j} \right) = \sum_{j=1}^n \left(\sum_{i=1}^m b_{i,j} \cdot a_{i,j} \right) = \text{tr}(BA)$$

3.3

Dokážeme rovnosti 2. mocnin, což je ekvivalentní, jelikož norma je vždy kladná.

$$\begin{aligned} \|UAV^T\|^2 &\stackrel{3.1}{=} \text{tr}(UAV^TVA^TU^T) \stackrel{\text{V má ortonormální posl. sloupce}}{=} \\ \text{tr}(UAA^TU^T) &\stackrel{3.1}{=} \text{tr}(A^TU^TUA) \stackrel{\text{U má ortonormální posl. sloupce}}{=} \text{tr}(A^TA) = \\ &= \|A\|_F^2 \end{aligned}$$

4

4.1

reflexivita:

$$U \preceq U \iff \forall x \in \mathbb{R}^n : x^T(U - U)x \geq 0 \implies x^T 0 x = 0 \forall x \in \mathbb{R}^n$$

tranzitivita:

$$\begin{aligned}
 U, V, W \in \mathbb{R}^{n \times n} : U \preceq V \wedge V \preceq W &\iff \forall x \in \mathbb{R}^n : x^T(V - U)x \geq 0 \wedge x^T(W - V)x \geq 0 \\
 x^T(W - U)x &= x^T(W - V + V - U)x \stackrel{\text{vlastnosti mat. nás.}}{=} \\
 x^T(W - V)x + x^T(V - U)x &\stackrel{1. \text{ řádek}}{\geq} 0 \quad \forall x \in \mathbb{R}^n \implies U \preceq W
 \end{aligned}$$

slabá antisymetrie:

$$U, V \in \mathbb{R}^{n \times n} : U \preceq V \wedge V \preceq U \iff V - U \text{ poz. semidef.} \wedge U - V \text{ poz. semidef.}$$

Víme, že pro pozitivně semidefinitní matice platí, že jsou ortogonálně diagonalizovatelné a jejich vlastní čísla jsou nezáporná. Označme $A := U - V$. Necht $\lambda \in \mathbb{R}$ je vlastní číslo A a $v \in \mathbb{R}^n$ jemu příslušný vlastní vektor. Platí:

$$Av = \lambda v \iff -Av = -\lambda v.$$

Víme že $\lambda \geq 0$ díky tomu, že je to vlastní číslo A , což je poz. semidef. matice. Ale $-A = V - U$, což je také semidefinitní matice, má evidentně $-\lambda$ jako své vlastní číslo. Takže $0 \geq \lambda \geq 0 \implies \lambda = 0$. Tedy libovolné vlastní číslo matice A je 0. Z diagonalizace víme, že $A = BDB^T$, kde D je diagonální (B ortogonální) a na diagonále jsou všechna vlastní čísla A . Tudíž $D = 0 \implies A = 0 \iff U = V$.

4.2

$$\begin{aligned}
 C &= \text{diag}(c_1, \dots, c_n), \quad D = \text{diag}(d_1, \dots, d_n), \quad D - C = \text{diag}(d_1 - c_1, \dots, d_n - c_n) \\
 C \preceq D &\iff \forall (x_1, \dots, x_n)^T \in \mathbb{R}^n : x^T(D - C)x \geq 0 \iff \\
 \forall (x_1, \dots, x_n)^T \in \mathbb{R}^n : &x_1^2 \cdot (d_1 - c_1) + \dots + x_n^2 \cdot (d_n - c_n) \geq 0
 \end{aligned}$$

x je libovolný vektor, tedy můžeme vzít $x = e_i \quad \forall i \in \{1, \dots, n\}$ a vidíme, že musí platit $\forall i \in \{1, \dots, n\} : d_i - c_i \geq 0 \iff d_i \geq c_i$. Naopak pokud je tato podmínka splněna tak rovnost bude vždy nezáporná, jelikož jsou tam druhé mocniny násobené nezáporným číslem. Tedy $C \preceq D \iff \forall i \in \{1, \dots, n\} : d_i \geq c_i$

4.3

Matice U, V jsou symetrické, tedy nám udávají symetrické bilineární formy. Tvzení 11.34 nám říká, protože je U pozitivně definitní, že existuje

matice Q ortogonální tž.:

$$c_i, d_i \in \mathbb{R}_0^+, \quad C = \text{diag}(c_1, \dots, c_n), \quad D = \text{diag}(d_1, \dots, d_n) : U = Q C Q^T \wedge V = Q D Q^T$$

$$U \preceq V \iff \forall x \in \mathbb{R}^n : x^T (V - U) x \geq 0 \iff x^T V x \geq x^T U x \iff \\ x^T (Q D Q^T) x \geq x^T (Q C Q^T) x$$

$$Q \text{ je bijekce tedy označíme-li } y := Q^T x \implies$$

$$\forall y \in \mathbb{R}^n : y^T D y \geq y^T C y \text{ tedy speciálně pro } y = e_i \ \forall i \in \{1, \dots, n\} \implies$$

$$\forall i \in \{1, \dots, n\} : d_i \geq c_i$$

Z 1.3 víme, že $\|U\|_F = \|Q C Q^T\|_F = \|C\|_F$ a $\|V\|_F = \|Q D Q^T\|_F = \|D\|_F$.
 Díky čemu plyne $\|D\|_F \geq \|C\|_F \implies \|V\|_F \geq \|U\|_F$.