Homework 3

Lecturer:Xiangyu Chang *Scribe: Xiangyu Chang*

Edited by: Xiangyu Chang

HW 1 *(1) Let* $A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \end{bmatrix}$*, then compute the column space* $C(A)$ *and null space* $N(A)$*.*

(2) Prove that for any $A \in \mathbb{R}^{m\times n}$*, it has*

$$C(A) \perp N(A^\top). \tag{1}$$

HW 2 *(1) Prove* ℓ_0 *norm is not a vector norm.*

(2) Prove $\sigma_1 = \sup_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2$*, where* σ_1 *is the biggest singular value of* A*.*

(3) $\|AB\|_F \leq \|A\|_2 \|B\|_F$ *for any* $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times p}$*.*

HW 3 *Let*

$$f(\mathbf{x}) = \sum_{i=1}^{m} \log(1 + \exp(\langle \mathbf{a}_i, \mathbf{x}\rangle)) - \langle \mathbf{b}, A\mathbf{x}\rangle$$

be the objective function of logistic regression, then

(1) please prove that $f(\mathbf{x})$ *is* β*-smooth;*

(2) write down the iterative formulation of Gradient Descent algorithm for Logistic Regression.

HW 4 *(1) Let* $C \subseteq \mathbb{R}^n$ *be a convex set, with* $\mathbf{x}_1, \ldots, \mathbf{x}_k \in C$*, and let* $\theta_1, \theta_2, \ldots, \theta_k \in \mathbb{R}$ *satisfy* $\theta_i \geq 0, \sum_i \theta_i = 1$*. Show that* $\theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2 + \cdots + \theta_k\mathbf{x}_k \in C$*.*

(2) Prove that the ellipsoid $E(\mathbf{x}_c) = \{\mathbf{x}|\ (\mathbf{x} - \mathbf{x}_c)^\top A(\mathbf{x} - \mathbf{x}_c) \leq 1, A \in \mathcal{S}^n_{++}\}$ *is a convex set.*

(3) Suppose that f *is a convex function, and denote the set contains all the points which can achieve the global minimum of* f *is* G*. Please prove that* G *is convex.*

HW 5 *(1) Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be distinct points in* $\mathbb{R}^n$*. Show that the set of all points that are closer to* $\mathbf{a}$ *than* $\mathbf{b}$ *is a halfspace.*

(2) What is the distance between two parallel hyperplances $\{\mathbf{x}|\mathbf{a}^\top\mathbf{x} = b_1\}$ *and* $\{\mathbf{x}|\mathbf{a}^\top\mathbf{x} = b_2\}$*?*

HW 6 *Prove the following functions are convex.*

(1) Negative Entropy: $f(x) = x\log(x), x > 0$*.*

(2) Quadratic-over-linear function: $f(x, y) = \frac{x^2}{y}$ *with* $dom(f) = \{(x, y) \in \mathbb{R}^2 | y > 0\}$*.*

(3) $f(\mathbf{x}) = \|A\mathbf{x} - b\|$*.*

(4) Supporting function: $S_C(\mathbf{x}) = \sup_{\mathbf{b}\in C} \mathbf{b}^\top\mathbf{x}$*.*

(5) The distance of a point $\mathbf{x}$ *to a set* $S \subset \mathbb{R}^n$ *defined as* $d(\mathbf{x}, S) = \inf_{\mathbf{b} \in S} \|\mathbf{x} - \mathbf{b}\|$.

HW 7 *Let* $A = U\Sigma V = \begin{bmatrix} -0.91 & 0.37 & -0.18 \\ 0.19 & 0.78 & 0.59 \\ 0.36 & 0.50 & -0.78 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.25 & -0.73 & -0.62 \\ -0.74 & -0.56 & 0.35 \\ -0.61 & 0.37 & -0.69 \end{bmatrix}$ *and* $\mathbf{b} = (-0.29, -2.09, -0.98)^\top$, *then consider a LS problem:*

$$\min_{\mathbf{x}} \frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|^2. \tag{2}$$

- *Implement gradient descent algorithm with the backtracking line search for solving the LS problem.*
- *Implement gradient descent for* β*-smooth function for solving the LS problem.*
- *Compare the convergence speed for different* A.

$$A_1 = \begin{bmatrix} -0.91 & 0.37 & -0.18 \\ 0.19 & 0.78 & 0.59 \\ 0.36 & 0.50 & -0.78 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 10^{-2} \end{bmatrix} \begin{bmatrix} 0.25 & -0.73 & -0.62 \\ -0.74 & -0.56 & 0.35 \\ -0.61 & 0.37 & -0.69 \end{bmatrix}, \tag{3}$$

and

$$A_2 = \begin{bmatrix} -0.91 & 0.37 & -0.18 \\ 0.19 & 0.78 & 0.59 \\ 0.36 & 0.50 & -0.78 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 10^{-6} \end{bmatrix} \begin{bmatrix} 0.25 & -0.73 & -0.62 \\ -0.74 & -0.56 & 0.35 \\ -0.61 & 0.37 & -0.69 \end{bmatrix}. \tag{4}$$

HW 8 *(Additional Question) Let* $A \in \mathbb{R}^{m\times n}$ *with its singular value decomposition as* $A = \sum_{i=1}^r \sigma_i u_i v_i^\top, \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$. *Assume that* $A_k = \sum_{i=1}^k \sigma_i u_i v_i^\top$, *where* $k \leq r$. *Then prove that*

$$A_k = \arg \min_{B \in \mathbb{R}^{m\times n}, rank(B) \leq k} \|A - B\|_F^2. \tag{5}$$

HW 9 *(Additional Question) Logistic Regression for classifying digital numbers 0 and 1. Please see the readme file.*

References