

Первое задание

Хохлов Алексей

25 февраля 2019 г.

Доделать: 1.3, 1.4, 4.2 (представить в матричном виде), 8 (возможно, доказывать необходимость)
Везде, где будет необходимо доказать выпуклость множества или функции, будет подразумеваться, что $0 \leq \theta_i \leq 1$, $\sum \theta_i = 1$. Для доказательства аффинности: $\sum \theta_i = 1$

1 Диаграмма Вороного

1.1

Условие $\|\mathbf{x} - \mathbf{x}_0\|_2 \leq \|\mathbf{x} - \mathbf{x}_i\|_2, i = 1, \dots, k$ означает, что

$$\sqrt{\sum_j (x_j - x_{0j})^2} \leq \sqrt{\sum_j (x_j - x_{ij})^2} \quad (1.1)$$

Возведя в квадрат и перенеся в левую сторону, получим

$$\sum_j 2(x_{ij} - x_{0j})(x_j - \frac{x_{0j} + x_{ij}}{2}) \leq 0 \quad (1.2)$$

$$\sum_j 2(x_{ij} - x_{0j})x_j \leq \sum_j (x_{ij} - x_{0j})(x_{0j} + x_{ij}) \quad (1.3)$$

Последнее неравенство можно представить в виде

$$\mathbf{Ax} \preceq \mathbf{b} \quad (1.4)$$

где у матрицы \mathbf{A} коэффициенты $a_{ij} = 2(x_{ij} - x_{0j})$, а у столбца $b_i = \sum_j (x_{ij} - x_{0j})(x_{0j} + x_{ij})$

Как видно, область Вороного является многоугольником.

1.2

Попробуем из известных \mathbf{A} и \mathbf{b} восстановить точки. Выразим из $(x_{ij} - x_{0j}) = a_{ij}/2$ точки x_{ij}

$$x_{ij} = \frac{a_{ij}}{2} + x_{0j} \quad (1.5)$$

и подставим в b_i

$$b_i = \sum_j \frac{a_{ij}}{2} (\frac{a_{ij}}{2} + 2x_{0j}) = \frac{1}{4} \sum_j a_{ij}^2 + \sum_j a_{ij}x_{0j} \quad (1.6)$$

$$\mathbf{Ax}_0 = \mathbf{b} - \frac{1}{4}\tilde{\mathbf{a}} \quad (1.7)$$

где столбец $\tilde{\mathbf{a}}$ таков, что их i элементы — сумма квадратов элементов i строки матрицы \mathbf{A}
Если матрица обратима, то

$$\mathbf{x}_0 = \mathbf{A}^{-1}(\mathbf{b} - \frac{1}{4}\tilde{\mathbf{a}}) \quad (1.8)$$

А из (1.5) следует, что матрица \mathbf{X} , столбцы которой — точки \mathbf{x}_i , такова:

$$\mathbf{X} = \frac{1}{2}\mathbf{A} + \mathbf{A}^{-1}(\mathbf{b} - \frac{1}{4}\tilde{\mathbf{a}}) \quad (1.9)$$

Если же матрица необратима, то в (1.7) ищется пространство решений \mathbf{x}_0 , и найденное решение подставляется в (1.5). Таким способом можно восстановить точки \mathbf{x}_i , $i = 0, \dots, k$

1.3

Заменяем \mathbf{x}_0 на \mathbf{x}_p . Разбиение будет выглядеть следующим образом

1.4

2 Множество решений квадратного неравенства

2.1

Пусть $\mathbf{x}_1, \mathbf{x}_2 \in C$ — т.е. являются решениями неравенства. Узнаем, является ли $\tilde{\mathbf{x}} = \theta\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2$ также решением. Пусть $f(\mathbf{x})$ — квадратичная функция.

$$f(\tilde{\mathbf{x}}) = (\theta\mathbf{x}_1^T + (1 - \theta)\mathbf{x}_2^T)\mathbf{A}(\theta\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2) + \mathbf{b}^T(\theta\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2) + c \quad (2.1)$$

Преобразуем квадратичное слагаемое

$$\begin{aligned} (\theta\mathbf{x}_1^T + (1 - \theta)\mathbf{x}_2^T)\mathbf{A}(\theta\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2) &= \theta^2\mathbf{x}_1^T\mathbf{A}\mathbf{x}_1 + (1 - \theta)^2\mathbf{x}_2^T\mathbf{A}\mathbf{x}_2 + \theta(1 - \theta)\mathbf{x}_1^T\mathbf{A}\mathbf{x}_2 + \theta(1 - \theta)\mathbf{x}_2^T\mathbf{A}\mathbf{x}_1 = \\ &= \theta\mathbf{x}_1^T\mathbf{A}\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2^T\mathbf{A}\mathbf{x}_2 - \theta(1 - \theta)\mathbf{x}_1^T\mathbf{A}\mathbf{x}_1 - \theta(1 - \theta)\mathbf{x}_2^T\mathbf{A}\mathbf{x}_2 + \theta(1 - \theta)\mathbf{x}_1^T\mathbf{A}\mathbf{x}_2 + \theta(1 - \theta)\mathbf{x}_2^T\mathbf{A}\mathbf{x}_1 = \\ &= \theta\mathbf{x}_1^T\mathbf{A}\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2^T\mathbf{A}\mathbf{x}_2 + \theta(1 - \theta) [\mathbf{x}_1^T\mathbf{A}(\mathbf{x}_2 - \mathbf{x}_1) + \mathbf{x}_2^T\mathbf{A}(\mathbf{x}_1 - \mathbf{x}_2)] = \\ &= \theta\mathbf{x}_1^T\mathbf{A}\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2^T\mathbf{A}\mathbf{x}_2 - \theta(1 - \theta)(\mathbf{x}_2^T - \mathbf{x}_1^T)\mathbf{A}(\mathbf{x}_2 - \mathbf{x}_1) \end{aligned} \quad (2.2)$$

Тогда квадратичная функция запишется в виде

$$\begin{aligned} f(\tilde{\mathbf{x}}) &= \theta\mathbf{x}_1^T\mathbf{A}\mathbf{x}_1 + \theta\mathbf{b}^T\mathbf{x}_1 + \theta c + (1 - \theta)\mathbf{x}_2^T\mathbf{A}\mathbf{x}_2 + (1 - \theta)\mathbf{b}^T\mathbf{x}_2 + (1 - \theta)c - \theta(1 - \theta)(\mathbf{x}_2^T - \mathbf{x}_1^T)\mathbf{A}(\mathbf{x}_2 - \mathbf{x}_1) = \\ &= \theta [\mathbf{x}_1^T\mathbf{A}\mathbf{x}_1 + \mathbf{b}^T\mathbf{x}_1 + c] + (1 - \theta) [\mathbf{x}_2^T\mathbf{A}\mathbf{x}_2 + \mathbf{b}^T\mathbf{x}_2 + c] - \theta(1 - \theta)(\mathbf{x}_2^T - \mathbf{x}_1^T)\mathbf{A}(\mathbf{x}_2 - \mathbf{x}_1) \end{aligned} \quad (2.3)$$

Первое и второе слагаемое неположительны, т.к. удовлетворяют условию квадратичного неравенства. Если $\mathbf{A} \succ 0$, то

$$-\theta(1 - \theta)(\mathbf{x}_2^T - \mathbf{x}_1^T)\mathbf{A}(\mathbf{x}_2 - \mathbf{x}_1) = -\theta(1 - \theta)\tilde{\mathbf{x}}^T\mathbf{A}\tilde{\mathbf{x}} \leq 0 \quad (2.4)$$

Итак, $\tilde{\mathbf{x}}$ также является решением, т.к.

$$f(\tilde{\mathbf{x}}) \leq 0 \quad (2.5)$$

Иными словами, множество решений квадратичного неравенства выпукло.

2.2

Рассмотрим третье слагаемое в выражении (2.3)

$$\begin{aligned} & -\theta(1-\theta)(\mathbf{x}_2^T - \mathbf{x}_1^T)\mathbf{A}(\mathbf{x}_2 - \mathbf{x}_1) = \\ & = -\theta(1-\theta)(\mathbf{x}_2^T - \mathbf{x}_1^T)(\mathbf{A} + \lambda\mathbf{g}\mathbf{g}^T)(\mathbf{x}_2 - \mathbf{x}_1) + \theta(1-\theta)(\mathbf{x}_2^T - \mathbf{x}_1^T)\lambda\mathbf{g}\mathbf{g}^T(\mathbf{x}_2 - \mathbf{x}_1) \end{aligned} \quad (2.6)$$

Первое слагаемое в (2.6) неположительно, что следует из $\mathbf{A} + \lambda\mathbf{g}\mathbf{g}^T \succeq 0$. Изучим теперь второе слагаемое (2.6).

Из определения гиперплоскости следует, что $\mathbf{g}^T\mathbf{x}_1 = \mathbf{g}^T\mathbf{x}_2$, $\mathbf{x}_1^T\mathbf{g} = \mathbf{x}_2^T\mathbf{g}$.

$$\theta(1-\theta)(\mathbf{x}_2^T - \mathbf{x}_1^T)\lambda\mathbf{g}\mathbf{g}^T(\mathbf{x}_2 - \mathbf{x}_1) = \theta(1-\theta)(\mathbf{x}_2^T\mathbf{g} - \mathbf{x}_1^T\mathbf{g})(\mathbf{g}^T\mathbf{x}_2 - \mathbf{g}^T\mathbf{x}_1) = 0 \quad (2.7)$$

Итак, для пересечения C и гиперплоскости при условии $\mathbf{A} + \lambda\mathbf{g}\mathbf{g}^T \succeq 0$ выполнено

$$\begin{aligned} f(\tilde{\mathbf{x}}) & \leq 0 \\ \mathbf{g}^T\tilde{\mathbf{x}} + b & = 0 \end{aligned} \quad (2.8)$$

Значит, множество выпукло.

3 Выпуклый и аффинный

Для удобства будем обозначать в каждом пункте исследуемые множества буквой C .

3.1

Проверим на выпуклость:

$$\alpha = \theta\alpha + (1-\theta)\alpha \leq \mathbf{a}^T(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) = \theta\mathbf{a}^T\mathbf{x}_1 + (1-\theta)\mathbf{a}^T\mathbf{x}_2 \leq \theta\beta + (1-\theta)\beta = \beta \quad (3.1)$$

Множество выпукло. Проверим теперь на аффинность:

Пусть $\alpha \neq \beta$, $\mathbf{a}^T\mathbf{x}_1 = \acute{\alpha}$, $\mathbf{a}^T\mathbf{x}_2 = \acute{\beta}$, $\alpha \leq \acute{\alpha} < \acute{\beta} \leq \beta$

$$\mathbf{a}^T(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) = \theta\acute{\alpha} + (1-\theta)\acute{\beta} = \acute{\beta} + \theta(\acute{\alpha} - \acute{\beta}) \quad (3.2)$$

Если $\theta < 0$, то выражение (3.2) больше, чем β , или, говоря иначе, $(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) \notin C$. Множество не является аффинным.

Если же $\alpha = \beta$, то множество будет удовлетворять уравнению прямой $\mathbf{a}^T\mathbf{x} = \alpha = \beta$. Такое множество будет аффинным:

$$\mathbf{a}^T(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) = \theta\alpha + (1-\theta)\alpha = \alpha \quad (3.3)$$

3.2

Проверим на выпуклость:

$$\mathbf{a}_{1,2}^T(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) = \theta\mathbf{a}_{1,2}^T\mathbf{x}_1 + (1-\theta)\mathbf{a}_{1,2}^T\mathbf{x}_2 \leq \theta b_{1,2} + (1-\theta)b_{1,2} = b_{1,2} \quad (3.4)$$

Множество выпукло. Проверим теперь на аффинность:

Пусть, во-первых, множество непусто, и вместе с тем $\mathbf{a}_1^T \neq -\mathbf{a}_2^T$, $b_1 \neq -b_2$. $\alpha = \mathbf{a}_1^T\mathbf{x}_1$, $\beta = \mathbf{a}_1^T\mathbf{x}_2$, $\alpha < \beta \leq b_1$

$$\mathbf{a}_1^T(\theta\mathbf{x}_1 + (1-\theta)\mathbf{x}_2) = \theta\alpha + (1-\theta)\beta = \beta + \theta(\alpha - \beta) \quad (3.5)$$

Пусть $\theta = \frac{b_1 - \beta + 1}{\alpha - \beta}$. Тогда

$$\beta + \theta(\alpha - \beta) = b_1 - 1 \geq b_1 \quad (3.6)$$

Множество не аффинно.

В случае же $\mathbf{a}_1^T = -\mathbf{a}_2^T$, $b_1 = -b_2$ Множество будет задавать прямую, а такое множество, как показано в пункте 1, аффинное.

3.3

Проверим на выпуклость:

Воспользуемся (1.3), где $\mathbf{x}_i \in S$, а $\tilde{x}_j = \theta x_j^* + (1 - \theta)x_j^{**}$. Неравенство

$$\theta \sum_j 2(x_{ij} - x_{0j})x_j^* + (1 - \theta) \sum_j 2(x_{ij} - x_{0j})x_j^{**} \leq \sum_j (x_{ij} - x_{0j})(x_{0j} + x_{ij}) \quad (3.7)$$

для любого $\mathbf{x}_i \in S$. Множество выпукло. Проверим теперь на аффинность:

Пусть $\sum_j 2(x_{ij} - x_{0j})x_j^* = b_i^*$, $\sum_j 2(x_{ij} - x_{0j})x_j^{**} = b_i^{**}$, $b_i^* < b_i^{**} \leq b_i$, $\theta = \frac{b_i + 1 - b_i^*}{b_i^* - b_i^{**}}$

$$\theta \sum_j 2(x_{ij} - x_{0j})x_j^* + (1 - \theta) \sum_j 2(x_{ij} - x_{0j})x_j^{**} = b_i^{**} + \theta(b_i^* - b_i^{**}) = b_i + 1 > b_i \quad (3.8)$$

Множество не аффинно. В случае, если $S = \{\mathbf{x}_0\}$, то исследуемое множество — \mathbb{R}^n , а такое множество аффинно.

3.4

Условие $\|\mathbf{x} - \mathbf{a}\|_2 \leq \theta \|\mathbf{x} - \mathbf{b}\|_2$ равнозначено

$$\sum_{i=1}^n [(x_i - a_i)^2 - \theta^2(x_i - b_i)^2] = \sum_{i=1}^n [(1 - \theta^2)x_i^2 - 2(a_i - b_i)x_i + (a_i^2 - b_i^2)] \leq 0 \quad (3.9)$$

Искомое множество — множество решений квадратичного неравенства. Как было доказано в задаче 2, множество решений выпукло, если $\mathbf{A} \succ 0$. В нашем случае квадратичная форма такова: $\mathbf{A} = \text{diag}((1 - \theta^2), \dots, (1 - \theta^2))$. Поскольку по условию $0 \leq \theta \leq 1$, то при $\theta < 1$ будет $(1 - \theta^2) > 0$. Значит, все угловые миноры положительны, матрица положительно определена, и, следовательно, множество выпукло. При $\theta = 1$ множество будет $\|\mathbf{x} - \mathbf{a}\|_2 \leq \|\mathbf{x} - \mathbf{b}\|_2$, которое, как было доказано, выпукло.

Проверим теперь на аффинность:

Если $\theta = 1$, то, как показал пункт 3, множество не будет аффинным. Поскольку по условиям задачи $\theta \in [0; 1]$, то достаточно найти только одно такое θ , что рушило бы аффинность множества. Однако если быть до конца честным, можно исследовать аффинность множества решений квадратичного неравенства, т.е. при $\theta \neq 1$. Пусть $\mathbf{x} = \alpha \mathbf{x}_1 + (1 - \alpha)\mathbf{x}_2$, $\sum_{i=1}^n [(1 - \theta^2)x_{1i}^2 - 2(a_i - b_i)x_{1i} + (a_i^2 - b_i^2)] = y_1$,

$$\sum_{i=1}^n [(1 - \theta^2)x_{2i}^2 - 2(a_i - b_i)x_{2i} + (a_i^2 - b_i^2)] = y_2, \quad y_2 \neq y_1$$

Легко видеть, что при

$$\alpha > \frac{1}{2} \left| \frac{y_1 - y_2}{(1 - \theta^2)\|\mathbf{x}_1 - \mathbf{x}_2\|^2} - 1 \right| \sqrt{\left(\frac{y_1 - y_2}{(1 - \theta^2)\|\mathbf{x}_1 - \mathbf{x}_2\|^2} - 1 \right)^2 - \frac{4y_2}{(1 - \theta^2)\|\mathbf{x}_1 - \mathbf{x}_2\|^2}} \quad (3.10)$$

$$\sum_{i=1}^n [(1 - \theta^2)\tilde{x}_i^2 - 2(a_i - b_i)\tilde{x}_i + (a_i^2 - b_i^2)] > 0 \quad (3.11)$$

Множество не аффинное.

3.5

Проверим на выпуклость:

$$\mathbf{F}_0 + \sum_{i=1}^n (\theta x_{1i} + (1 - \theta)x_{2i})\mathbf{F}_i = \theta(\mathbf{F}_0 + \sum_{i=1}^n x_{1i}\mathbf{F}_i) + (1 - \theta)(\mathbf{F}_0 + \sum_{i=1}^n x_{2i}\mathbf{F}_i) \succeq 0 \quad (3.12)$$

Множество выпукло. Проверим теперь на аффинность:

Пусть $\mathbf{a}^T \left[\mathbf{F}_0 + \sum_{i=1}^n x_{1i}\mathbf{F}_i \right] \mathbf{a} = \alpha$, $\mathbf{a}^T \left[\mathbf{F}_0 + \sum_{i=1}^n x_{2i}\mathbf{F}_i \right] \mathbf{a} = \beta$, $\beta > \alpha$

$$\mathbf{a}^T \left[\mathbf{F}_0 + \sum_{i=1}^n (\theta x_{1i} + (1 - \theta)x_{2i})\mathbf{F}_i \right] \mathbf{a} = \beta + \theta(\alpha - \beta) \quad (3.13)$$

При $\theta = \frac{-\beta - 1}{\alpha - \beta}$

$$\mathbf{a}^T \left[\mathbf{F}_0 + \sum_{i=1}^n (\theta x_{1i} + (1 - \theta)x_{2i})\mathbf{F}_i \right] \mathbf{a} = -1 < 0 \quad (3.14)$$

Множество не аффинное.

4 Градиенты и гессианы

4.1

4.1.1 а

Как известно, след матрицы — сумма собственных чисел.

$$f(\mathbf{X}) = \text{Tr}(\mathbf{X}) = \sum_i x_{ii} \quad (4.1)$$

Скалярное представление градиента:

$$\frac{\partial f(\mathbf{X})}{\partial x_{pq}} = \delta_{pq} \quad (4.2)$$

Векторное представление градиента:

$$\nabla f(\mathbf{X}) = \mathbf{I} \quad (4.3)$$

4.1.2 б

Из уравнения на собственные числа и теоремы Виета

$$\det(\mathbf{X} - \lambda \mathbf{I}) = (-1)^n \lambda^n + (-1)^{n-1} \lambda^{n-1} \sum_{i=1}^n x_{ii} + \dots + \det(\mathbf{X} - 0\mathbf{I}) = 0 \quad (4.4)$$

следует, что

$$\det(\mathbf{X}) = \prod_{i=1}^n \lambda_i(\mathbf{X}) = f(\mathbf{X}) \quad (4.5)$$

Скалярное представление функции:

$$f(\mathbf{X}) = \sum_j (-1)^{i+j} x_{ij} M_{ij} \quad (4.6)$$

Скалярное представление градиента:

$$\frac{\partial f(\mathbf{X})}{\partial x_{pq}} = (-1)^{p+q} M_{pq} \quad (4.7)$$

Векторное представление градиента:

$$\nabla f(\mathbf{X}) = (\text{adj}(\mathbf{X}))^T = (\mathbf{X}^{-1} \cdot \det(\mathbf{X}))^T = \det(\mathbf{X}) \cdot \mathbf{X}^{-T} \quad (4.8)$$

4.2

Матричное представление функции:

$$J(\mathbf{U}, \mathbf{V}) = \|\mathbf{UV} - \mathbf{Y}\|_F^2 + \frac{\lambda}{2} (\|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2) \quad (4.9)$$

Скалярное представление функции:

$$J(\mathbf{U}, \mathbf{V}) = \sum_{i=1}^n \sum_{j=1}^n \left(\sum_{r=1}^n u_{ir} v_{rj} - y_{ij} \right)^2 + \frac{\lambda}{2} \sum_{i=1}^n \sum_{j=1}^n (u_{ij}^2 + v_{ij}^2) \quad (4.10)$$

Скалярное представление градиентов:

$$\begin{aligned} \frac{\partial J(\mathbf{U}, \mathbf{V})}{\partial u_{pq}} &= \sum_{j=1}^n 2v_{pj} \left(\sum_{r=1}^n u_{pr} v_{rj} - y_{pj} \right) + \lambda u_{pq} = \sum_{j=1}^n \sum_{r=1}^n 2v_{pj} u_{pr} v_{rj} - \sum_{j=1}^n 2v_{pj} y_{pj} + \lambda u_{pq} = \\ &= \sum_{j=1}^n \sum_{r=1}^n 2v_{pj} v_{jr}^T u_{rp}^T - \sum_{j=1}^n 2v_{pj} y_{pj}^T + \lambda u_{pq} = 2(\mathbf{V}\mathbf{V}^T \mathbf{U}^T)_{pp} - 2(\mathbf{V}\mathbf{Y}^T)_{pp} + \lambda(\mathbf{U})_{pq} = \\ &= 2(\mathbf{V}(\mathbf{UV} - \mathbf{Y})^T)_{pp} + \lambda(\mathbf{U})_{pq} \end{aligned} \quad (4.11)$$

где $(\mathbf{A})_{pq}$ означает pq компоненту \mathbf{V} . Аналогично и для градиента по \mathbf{V}

$$\begin{aligned} \frac{\partial J(\mathbf{U}, \mathbf{V})}{\partial v_{pq}} &= \sum_{i=1}^n 2u_{iq} \left(\sum_{r=1}^n u_{ir} v_{rq} - y_{iq} \right) + \lambda v_{pq} = 2 \sum_{i=1}^n \sum_{r=1}^n u_{qi}^T u_{ir} v_{rq} - 2 \sum_{i=1}^n u_{qi}^T y_{iq} + \lambda v_{pq} = \\ &= 2(\mathbf{U}^T(\mathbf{UV} - \mathbf{Y}))_{qq} + \lambda(\mathbf{V})_{pq} \end{aligned} \quad (4.12)$$

Матричное представление градиентов:

$$\nabla_u J(\mathbf{U}, \mathbf{V}) = \dots + \lambda \mathbf{U} \quad (4.13)$$

$$\nabla_v J(\mathbf{U}, \mathbf{V}) = \dots + \lambda \mathbf{V} \quad (4.14)$$

4.3

Матричное представление функции:

$$f(\mathbf{w}) = \sum_{i=1}^m \log(1 + e^{-y_i \mathbf{w}^T \mathbf{x}_i}) \quad (4.15)$$

Скалярное представление:

$$f(\mathbf{w}) = \sum_{i=1}^m \log(1 + e^{-y_i \sum_j w_j x_{ij}}) \quad (4.16)$$

Скалярное представление градиента:

$$\frac{\partial f(\mathbf{w})}{\partial w_k} = \sum_{i=1}^m \frac{1}{1 + e^{-y_i \mathbf{w}^T \mathbf{x}_i}} \cdot e^{-y_i \mathbf{w}^T \mathbf{x}_i} \cdot (-y_i x_{ik}) \quad (4.17)$$

Матричное представление градиента:

$$\nabla f(\mathbf{w}) = - \sum_{i=1}^m \frac{e^{-y_i \mathbf{w}^T \mathbf{x}_i}}{1 + e^{-y_i \mathbf{w}^T \mathbf{x}_i}} y_i \mathbf{x}_i \quad (4.18)$$

Скалярное представление гессиана:

$$\frac{\partial^2 f(\mathbf{w})}{\partial w_k \partial w_p} = \sum_{i=1}^m (-y_i x_{ik}) \frac{e^{-y_i \mathbf{w}^T \mathbf{x}_i} \cdot (-y_i x_{ip}) \cdot (1 + e^{-y_i \mathbf{w}^T \mathbf{x}_i}) - e^{-y_i \mathbf{w}^T \mathbf{x}_i} \cdot e^{-y_i \mathbf{w}^T \mathbf{x}_i} \cdot (-y_i x_{ip})}{(1 + e^{-y_i \mathbf{w}^T \mathbf{x}_i})^2} \quad (4.19)$$

$$\frac{\partial^2 f(\mathbf{w})}{\partial w_k \partial w_p} = \sum_{i=1}^m y_i^2 x_{ik} x_{ip} \frac{e^{-y_i \mathbf{w}^T \mathbf{x}_i}}{(1 + e^{-y_i \mathbf{w}^T \mathbf{x}_i})^2} \quad (4.20)$$

Матричное представление гессиана:

$$\mathbf{H} = \sum_{i=1}^m y_i^2 \frac{e^{-y_i \mathbf{w}^T \mathbf{x}_i}}{(1 + e^{-y_i \mathbf{w}^T \mathbf{x}_i})^2} \mathbf{x}_i \otimes \mathbf{x}_i^T \quad (4.21)$$

5 Кратчайший путь в графе

Пусть \mathbf{c} — вектор-столбец весов $(c_{pk}, c_{pq}, c_{qm}, \dots)^T$ ориентированного взвешенного графа. Функция кратчайшего пути будет выглядеть следующим образом:

$$p_{ij}(\mathbf{c}) = \sum P_{pq}(\mathbf{c}) c_{pq} \quad (5.1)$$

где

$$P_{pq}(\mathbf{c}) = \begin{cases} 1, & (p, q) \subseteq (i, j) \\ 0, & (p, q) \not\subseteq (i, j) \end{cases} \quad (5.2)$$

$$(5.3)$$

или, иными словами, функция $P_{pq}(\mathbf{c})$ принимает значение 1, если ребро лежит в траектории с минимальным путём, и 0, если не лежит. Возьмем два произвольных столбца весов \mathbf{c}^a и \mathbf{c}^b , а также столбец $\tilde{\mathbf{c}} = \theta \mathbf{c}^a + (1 - \theta) \mathbf{c}^b$

$$p_{ij}(\tilde{\mathbf{c}}) = \sum P_{pq}(\tilde{\mathbf{c}}) (\theta c_{pq}^a + (1 - \theta) c_{pq}^b) = \theta \sum P_{pq}(\tilde{\mathbf{c}}) c_{pq}^a + (1 - \theta) \sum P_{pq}(\tilde{\mathbf{c}}) c_{pq}^b \quad (5.4)$$

Проверим на выпуклость или вогнутость:

$$\begin{aligned} & p_{ij}(\tilde{\mathbf{c}}) - \theta \sum P_{pq}(\mathbf{c}^a) c_{pq}^a - (1 - \theta) \sum P_{pq}(\mathbf{c}^b) c_{pq}^b \\ &= \theta \left[\sum P_{pq}(\tilde{\mathbf{c}}) c_{pq}^a - \sum P_{pq}(\mathbf{c}^a) c_{pq}^a \right] + (1 - \theta) \left[\sum P_{pq}(\tilde{\mathbf{c}}) c_{pq}^b - \sum P_{pq}(\mathbf{c}^b) c_{pq}^b \right] \end{aligned} \quad (5.5)$$

Поскольку $\sum P_{pq}(\mathbf{c}^a) c_{pq}^a$ и $\sum P_{pq}(\mathbf{c}^b) c_{pq}^b$ — кратчайшие пути для весов \mathbf{c}^a и \mathbf{c}^b , то для любого иного вектора весов $\tilde{\mathbf{c}}$ выполнены неравенства

$$\begin{aligned} \sum P_{pq}(\tilde{\mathbf{c}}) c_{pq}^a &\geq \sum P_{pq}(\mathbf{c}^a) c_{pq}^a \\ \sum P_{pq}(\tilde{\mathbf{c}}) c_{pq}^b &\geq \sum P_{pq}(\mathbf{c}^b) c_{pq}^b \end{aligned} \quad (5.6)$$

Следовательно, функция кратчайшего пути является вогнутой:

$$p_{ij}(\theta \mathbf{c}^a + (1 - \theta) \mathbf{c}^b) \geq \theta p_{ij}(\mathbf{c}^a) + (1 - \theta) p_{ij}(\mathbf{c}^b) \quad (5.7)$$

Эта вогнутость не является строгой. Рассмотрим граф

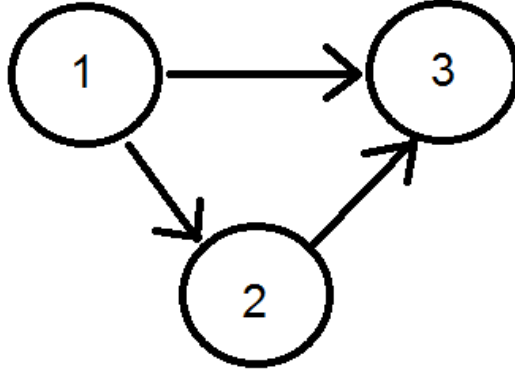


Рис. 1: Граф.

Пусть $c_{13}^a = 2, c_{13}^b = 6, c_{12}^a = 1, c_{12}^b = 1, c_{23}^a = 2, c_{23}^b = 2$. Тогда $p_{13}(\mathbf{c}^a) = 2, p_{13}(\mathbf{c}^b) = 4$. Пусть $\theta = 0.5$. Для $\tilde{\mathbf{c}}$ веса будут $c_{13} = 4, c_{12} = 1.5, c_{23} = 1.5$, а минимальный путь $p_{13}(\tilde{\mathbf{c}}) = 3$. Отсюда $p_{13}(\tilde{\mathbf{c}}) = 0.5p_{13}(\mathbf{c}^a) + 0.5p_{13}(\mathbf{c}^b)$. Как видим, строгого неравенства не наблюдается, и функция кратчайшего пути оказывается нестрого вогнутой.

6 Логарифмический барьер для конуса второго порядка

Итак, как обычно, пусть $\tilde{\mathbf{x}} = \theta\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2, \tilde{t} = \theta t_1 + (1 - \theta)t_2$

Раскроем аргумент логарифма:

$$\begin{aligned} \tilde{t}^2 - \tilde{\mathbf{x}}^T \tilde{\mathbf{x}} &= \theta^2 t_1^2 + (1 - \theta)^2 t_2^2 + 2\theta(1 - \theta)t_1 t_2 - \theta^2 \mathbf{x}_1^T \mathbf{x}_1 - (1 - \theta)^2 \mathbf{x}_2^T \mathbf{x}_2 - \theta(1 - \theta)\mathbf{x}_1^T \mathbf{x}_2 - \theta(1 - \theta)\mathbf{x}_2^T \mathbf{x}_1 = \\ &= \theta^2(t_1^2 - \mathbf{x}_1^T \mathbf{x}_1) + (1 - \theta)^2(t_2^2 - \mathbf{x}_2^T \mathbf{x}_2) - 2\theta(1 - \theta)(t_1 t_2 - \mathbf{x}_1^T \mathbf{x}_2) \end{aligned} \quad (6.1)$$

Проверим, верно ли неравенство

$$\tilde{t}^2 - \tilde{\mathbf{x}}^T \tilde{\mathbf{x}} \geq \theta(t_1^2 - \mathbf{x}_1^T \mathbf{x}_1) + (1 - \theta)(t_2^2 - \mathbf{x}_2^T \mathbf{x}_2) \quad (6.2)$$

Вычтем правую часть из левого и получим

$$\begin{aligned} \tilde{t}^2 - \tilde{\mathbf{x}}^T \tilde{\mathbf{x}} - \theta(t_1^2 - \mathbf{x}_1^T \mathbf{x}_1) + (1 - \theta)(t_2^2 - \mathbf{x}_2^T \mathbf{x}_2) &= \\ &= \theta(1 - \theta) [(t_1^2 - \mathbf{x}_1^T \mathbf{x}_1) + (t_2^2 - \mathbf{x}_2^T \mathbf{x}_2) - 2(t_1 t_2 - \mathbf{x}_1^T \mathbf{x}_2)] = \\ &= \theta(1 - \theta) [(t_1 - t_2)^2 - (\mathbf{x}_2 - \mathbf{x}_1)^T (\mathbf{x}_2 - \mathbf{x}_1)] \end{aligned} \quad (6.3)$$

Множество E таково, что $\|x\|_2 < t$. Следовательно, выражение (6.3) больше нуля, а значит, неравенство (6.2) верно.

Теперь проверим неравенство

$$-\log(\tilde{t}^2 - \tilde{\mathbf{x}}^T \tilde{\mathbf{x}}) \leq -\theta \log(t_1^2 - \mathbf{x}_1^T \mathbf{x}_1) - (1 - \theta) \log(t_2^2 - \mathbf{x}_2^T \mathbf{x}_2) \quad (6.4)$$

Избавляясь от логарифмов, получим

$$(\tilde{t}^2 - \tilde{\mathbf{x}}^T \tilde{\mathbf{x}}) \geq (t_1^2 - \mathbf{x}_1^T \mathbf{x}_1)^\theta (t_2^2 - \mathbf{x}_2^T \mathbf{x}_2)^{1-\theta} \quad (6.5)$$

Воспользуемся неравенством Юнга

$$a^\theta b^{1-\theta} \leq \theta a + (1 - \theta)b \quad (6.6)$$

Из неравенства Юнга и (6.2) следует

$$(\tilde{t}^2 - \tilde{\mathbf{x}}^T \tilde{\mathbf{x}}) \geq \theta(t_1^2 - \mathbf{x}_1^T \mathbf{x}_1) + (1 - \theta)(t_2^2 - \mathbf{x}_2^T \mathbf{x}_2) \geq (t_1^2 - \mathbf{x}_1^T \mathbf{x}_1)^\theta (t_2^2 - \mathbf{x}_2^T \mathbf{x}_2)^{1-\theta} \quad (6.7)$$

Отсюда следует, что неравенство (6.5) верно, и вместе с ним верно (6.4). И, следовательно, логарифмический барьер для конуса второго порядка является выпуклым на множестве E .

7 Обратное неравенство Йенсена

Необходимо доказать, что

$$f(\lambda_1 \mathbf{x}_1 + \dots + \lambda_n \mathbf{x}_n) \geq \lambda_1 f(\mathbf{x}_1) + \dots + \lambda_n f(\mathbf{x}_n) \quad (7.1)$$

Докажем неравенство, равносильное данному

$$f(\mathbf{x}_1) \leq \frac{1}{\lambda_1} f(\lambda_1 \mathbf{x}_1 + \dots + \lambda_n \mathbf{x}_n) + \sum_{i=2}^n \left(-\frac{\lambda_i}{\lambda_1}\right) f(\mathbf{x}_i) \quad (7.2)$$

Сделаем преобразование переменных. Пусть $\sum_{i=1}^n \lambda_i \mathbf{x}_i = \tilde{\mathbf{x}}_1$, $\mathbf{x}_i = \tilde{\mathbf{x}}_i$, $i = 2, \dots, n$. Неравенство преобразуется в

$$f\left(\frac{1}{\lambda_1} \tilde{\mathbf{x}}_1 + \sum_{i=2}^n \left(-\frac{\lambda_i}{\lambda_1}\right) \tilde{\mathbf{x}}_i\right) \leq \frac{1}{\lambda_1} f(\tilde{\mathbf{x}}_1) + \sum_{i=2}^n \left(-\frac{\lambda_i}{\lambda_1}\right) f(\tilde{\mathbf{x}}_i) \quad (7.3)$$

Поскольку $\lambda_1 > 0$, $\lambda_i \leq 0$, $i = 2, \dots, n$, то $\lambda_1 = 1 - \sum_{i=2}^n \lambda_i \geq 1$, и, соответственно, $0 < \frac{1}{\lambda_1} \leq 1$.

В то же время $\lambda_1 = 1 - \sum_{i=2}^n \lambda_i = 1 + \sum_{i=2}^n (-\lambda_i) = 1 + \sum_{i=2}^n |\lambda_i| > |\lambda_i|$, и, следовательно, $0 \leq \left(-\frac{\lambda_i}{\lambda_1}\right) < 1$.

Переобозначим $\theta_1 = \frac{1}{\lambda_1}$, $\theta_i = \left(-\frac{\lambda_i}{\lambda_1}\right)$, $i = 2, \dots, n$.

Как видно, $0 < \theta_1 \leq 1$, $0 \leq \theta_i < 1$, $\sum_i \theta_i = \frac{1 - \sum_{i=2}^n \lambda_i}{\lambda_1} = 1$. Неравенство переписывается в виде

$$f\left(\sum_{i=1}^n \theta_i \tilde{\mathbf{x}}_i\right) \leq \sum_{i=1}^n \theta_i f(\tilde{\mathbf{x}}_i) \quad (7.4)$$

Поскольку функция выпукла, то неравенство верно (неравенство Йенсена).

8 Выпуклая композиция

Пусть функция $\mathbf{y} = f(\mathbf{x})$ определена на множестве \mathbf{X} , а $h(\mathbf{y})$ на множестве $\mathbf{Y} = f(\mathbf{X})$. Их суперпозиция $g(\mathbf{x}) = h(f(\mathbf{x}))$. Также будем придерживаться следующих обозначений:

$$\begin{aligned} \tilde{\mathbf{x}} &= \theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2 \\ \tilde{\mathbf{y}} &= f(\theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2) \\ \tilde{\mathbf{y}} &= \theta f(\mathbf{x}_1) + (1 - \theta) f(\mathbf{x}_2) = \theta y_1 + (1 - \theta) y_2. \end{aligned} \quad (8.1)$$

Рассмотрим следующие случаи:

8.1 $f(\mathbf{x})$ — выпуклая, $h(y)$ — выпуклая и нестрого возрастающая

Т.к. $f(\mathbf{x})$ — выпуклая, то $\tilde{y} \leq \tilde{\tilde{y}}$. Поскольку $h(y)$ — нестрого возрастающая, то $h(\tilde{y}) \leq h(\tilde{\tilde{y}})$. А из выпуклости $h(y)$ следует, что $h(\tilde{\tilde{y}}) \leq \theta h(y_1) + (1 - \theta)h(y_2)$

$$h(f(\tilde{\mathbf{x}})) = h(\tilde{y}) \leq h(\tilde{\tilde{y}}) \leq \theta h(y_1) + (1 - \theta)h(y_2) = \theta h(f(\mathbf{x}_1)) + (1 - \theta)h(f(\mathbf{x}_2)) \quad (8.2)$$

Выпуклость f , выпуклость и неубывающая монотонность h одновременно — достаточное условие для выпуклости суперпозиции.

8.2 $f(\mathbf{x})$ — вогнутая, $h(y)$ — выпуклая и нестрого убывающая

Из вогнутости $f(\mathbf{x})$ следует $\tilde{y} \geq \tilde{\tilde{y}}$. Из невозрастания $h(y)$ следует $h(\tilde{y}) \leq h(\tilde{\tilde{y}})$. А из выпуклости $h(y)$ следует, что $h(\tilde{\tilde{y}}) \leq \theta h(y_1) + (1 - \theta)h(y_2)$

$$h(f(\tilde{\mathbf{x}})) = h(\tilde{y}) \leq h(\tilde{\tilde{y}}) \leq \theta h(y_1) + (1 - \theta)h(y_2) = \theta h(f(\mathbf{x}_1)) + (1 - \theta)h(f(\mathbf{x}_2)) \quad (8.3)$$

Вогнутость f , выпуклость и невозрастающая монотонность h одновременно — достаточное условие для выпуклости суперпозиции.

8.3 $f(\mathbf{x})$ и $h(y)$ — дважды дифференцируемы

$$\frac{\partial g}{\partial x_i} = h' \frac{\partial f}{\partial x_i} \quad (8.4)$$

$$\frac{\partial^2 g}{\partial x_i \partial x_j} = h'' \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} + h' \frac{\partial^2 f}{\partial x_i \partial x_j} \quad (8.5)$$

$$\mathbf{H}_g = h'' \nabla f \otimes (\nabla f)^T + h' \mathbf{H}_f \quad (8.6)$$