Первое задание

Хохлов Алексей

25 февраля 2019 г.

Доделать: 1.3,1.4, 8 (возможно, доказывать необходимость)

Везде, где будет необходимо доказать выпуклость множества или функции, будет подразумеваться, что $0 \leqslant \theta_i \leqslant 1$, $\sum \theta_i = 1$. Для доказательства афинности:$\sum \theta_i = 1$

1 Диаграмма Вороного

1.1

Условие $\|\mathbf{x} - \mathbf{x_0}\|_2 \leqslant \|\mathbf{x} - \mathbf{x_i}\|_2, i = 1, ..., k$ означает, что

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

Возведя в квадрат и перенеся в левую сторону, получим

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

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

Последнее неравенство можно представить в виде

$$\mathbf{Ax} \preceq \mathbf{b} \tag{1.4}$$

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

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}

$$x_{ij} = \frac{a_{ij}}{2} + x_{0j} \tag{1.5}$$

и подставим в b_i

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

$$\mathbf{Ax_0} = \mathbf{b} - \frac{1}{4}\tilde{\mathbf{a}} \tag{1.7}$$

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

$$\mathbf{x_0} = \mathbf{A}^{-1}(\mathbf{b} - \frac{1}{4}\tilde{\mathbf{a}}) \tag{1.8}$$

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

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

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 $\mathbf{x}_i$, $i = 0, ..., k$

1.3

Заменим $\mathbf{x}_0$ 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.

$$f(\tilde{\mathbf{x}}) = (\theta\mathbf{x_1^T} + (1-\theta)\mathbf{x_2^T})\mathbf{A}(\theta\mathbf{x_1} + (1-\theta)\mathbf{x_2}) + \mathbf{b^T}(\theta\mathbf{x_1} + (1-\theta)\mathbf{x_2}) + c \tag{2.1}$$

Преобразуем квадратичное слагаемое

$$\begin{aligned}
&(\theta\mathbf{x_1^T} + (1-\theta)\mathbf{x_2^T})\mathbf{A}(\theta\mathbf{x_1} + (1-\theta)\mathbf{x_2}) = \theta^2\mathbf{x_1^T}\mathbf{A}\mathbf{x_1} + (1-\theta)^2\mathbf{x_2^T}\mathbf{A}\mathbf{x_2} + \theta(1-\theta)\mathbf{x_1^T}\mathbf{A}\mathbf{x_2} + \theta(1-\theta)\mathbf{x_2^T}\mathbf{A}\mathbf{x_1} = \\
&= \theta\mathbf{x_1^T}\mathbf{A}\mathbf{x_1} + (1-\theta)\mathbf{x_2^T}\mathbf{A}\mathbf{x_2} - \theta(1-\theta)\mathbf{x_1^T}\mathbf{A}\mathbf{x_1} - \theta(1-\theta)\mathbf{x_2^T}\mathbf{A}\mathbf{x_2} + \theta(1-\theta)\mathbf{x_1^T}\mathbf{A}\mathbf{x_2} + \theta(1-\theta)\mathbf{x_2^T}\mathbf{A}\mathbf{x_1} = \\
&= \theta\mathbf{x_1^T}\mathbf{A}\mathbf{x_1} + (1-\theta)\mathbf{x_2^T}\mathbf{A}\mathbf{x_2} + \theta(1-\theta)\left[\mathbf{x_1^T}\mathbf{A}(\mathbf{x_2} - \mathbf{x_1}) + \mathbf{x_2^T}\mathbf{A}(\mathbf{x_1} - \mathbf{x_2})\right] = \\
&= \theta\mathbf{x_1^T}\mathbf{A}\mathbf{x_1} + (1-\theta)\mathbf{x_2^T}\mathbf{A}\mathbf{x_2} - \theta(1-\theta)(\mathbf{x_2^T} - \mathbf{x_1^T})\mathbf{A}(\mathbf{x_2} - \mathbf{x_1})
\end{aligned} \tag{2.2}$$

Тогда квадратичная функция запишется в виде

$$\begin{aligned}
&f(\tilde{\mathbf{x}}) = \theta\mathbf{x_1^T}\mathbf{A}\mathbf{x_1} + \theta\mathbf{b^T}\mathbf{x_1} + \theta c + (1-\theta)\mathbf{x_2^T}\mathbf{A}\mathbf{x_2} + (1-\theta)\mathbf{b^T}\mathbf{x_2} + (1-\theta)c - \theta(1-\theta)(\mathbf{x_2^T} - \mathbf{x_1^T})\mathbf{A}(\mathbf{x_2} - \mathbf{x_1}) = \\
&= \theta\left[\mathbf{x_1^T}\mathbf{A}\mathbf{x_1} + \mathbf{b^T}\mathbf{x_1} + c\right] + (1-\theta)\left[\mathbf{x_2^T}\mathbf{A}\mathbf{x_2} + \mathbf{b^T}\mathbf{x_2} + c\right] - \theta(1-\theta)(\mathbf{x_2^T} - \mathbf{x_1^T})\mathbf{A}(\mathbf{x_2} - \mathbf{x_1})
\end{aligned} \tag{2.3}$$

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

$$-\theta(1-\theta)(\mathbf{x_2^T} - \mathbf{x_1^T})\mathbf{A}(\mathbf{x_2} - \mathbf{x_1}) = -\theta(1-\theta)\check{\mathbf{x}}^\mathbf{T}\mathbf{A}\check{\mathbf{x}} \leqslant 0 \tag{2.4}$$

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

$$f(\tilde{\mathbf{x}}) \leqslant 0 \tag{2.5}$$

Иными словами, множество решений квадратичного неравенства выпукло.

2.2

Рассмотрим третье слагаемое в выражении (2.3)

$$
\begin{aligned}
&-\theta(1-\theta)(\mathbf{x_2^T}-\mathbf{x_1^T})\mathbf{A}(\mathbf{x_2}-\mathbf{x_1}) = \\
&= -\theta(1-\theta)(\mathbf{x_2^T}-\mathbf{x_1^T})(\mathbf{A}+\lambda\mathbf{g}\mathbf{g}^{\mathrm{T}})(\mathbf{x_2}-\mathbf{x_1}) + \theta(1-\theta)(\mathbf{x_2^T}-\mathbf{x_1^T})\lambda\mathbf{g}\mathbf{g}^{\mathrm{T}}(\mathbf{x_2}-\mathbf{x_1})
\end{aligned}
\tag{2.6}
$$

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

Из определения гиперплоскости следует, что $\mathbf{g}^{\mathrm{T}}\mathbf{x_1} = \mathbf{g}^{\mathrm{T}}\mathbf{x_2}$, $\mathbf{x_1}^{\mathrm{T}}\mathbf{g} = \mathbf{x_2}^{\mathrm{T}}\mathbf{g}$.

$$
\theta(1-\theta)(\mathbf{x_2^T}-\mathbf{x_1^T})\lambda\mathbf{g}\mathbf{g}^{\mathrm{T}}(\mathbf{x_2}-\mathbf{x_1}) = \theta(1-\theta)(\mathbf{x_2}^{\mathrm{T}}\mathbf{g}-\mathbf{x_1}^{\mathrm{T}}\mathbf{g})(\mathbf{g}^{\mathrm{T}}\mathbf{x_2}-\mathbf{g}^{\mathrm{T}}\mathbf{x_1}) = 0 \tag{2.7}
$$

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

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

Значит, множество выпукло.

3 Выпуклый и аффинный

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

3.1

Проверим на выпуклость:

$$
\alpha = \theta\alpha + (1-\theta)\alpha \leqslant \mathbf{a^T}(\theta\mathbf{x_1}+(1-\theta)\mathbf{x_2}) = \theta\mathbf{a^T}\mathbf{x_1} + (1-\theta)\mathbf{a^T}\mathbf{x_2} \leqslant \theta\beta + (1-\theta)\beta = \beta \tag{3.1}
$$

Множество выпукло. Проверим теперь на аффинность:

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

$$
\mathbf{a^T}(\theta\mathbf{x_1}+(1-\theta)\mathbf{x_2}) = \theta\acute{\alpha} + (1-\theta)\acute{\beta} = \acute{\beta} + \theta(\acute{\alpha}-\acute{\beta}) \tag{3.2}
$$

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 $\mathbf{a}^{\mathrm{T}}\mathbf{x} = \alpha = \beta$. Такое множество будет аффинным:

$$
\mathbf{a^T}(\theta\mathbf{x_1}+(1-\theta)\mathbf{x_2}) = \theta\alpha + (1-\theta)\alpha = \alpha \tag{3.3}
$$

3.2

Проверим на выпуклость:

$$
\mathbf{a_{1,2}^T}(\theta\mathbf{x_1}+(1-\theta)\mathbf{x_2}) = \theta\mathbf{a_{1,2}^T}\mathbf{x_1} + (1-\theta)\mathbf{a_{1,2}^T}\mathbf{x_2} \leqslant \theta b_{1,2} + (1-\theta)b_{1,2} = b_{1,2} \tag{3.4}
$$

Множество выпукло. Проверим теперь на аффинность:

Пусть, во-первых, множество непусто, и вместе с тем $\mathbf{a_1^T} \neq -\mathbf{a_2^T}$, $b_1 \neq -b_2$.$\alpha = \mathbf{a_1^T}\mathbf{x_1}$, $\beta = \mathbf{a_1^T}\mathbf{x_2}$, $\alpha < \beta \leqslant b_1$

$$
\mathbf{a_1^T}(\theta\mathbf{x_1}+(1-\theta)\mathbf{x_2}) = \theta\alpha + (1-\theta)\beta = \beta + \theta(\alpha-\beta) \tag{3.5}
$$

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

$$\beta + \theta(\alpha - \beta) = b_1 - 1 \geqslant b_1 \tag{3.6}$$

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

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

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

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

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

Множество не аффинно. В случае, если $S = \{\mathbf{x_0}\}$, то исследуемое множество — $\mathbb{R}^n$, а такое множество аффинно.

3.4

Условие $\|\mathbf{x} - \mathbf{a}\|_2 \leqslant \theta\|\mathbf{x} - \mathbf{b}\|_2$ равнозначно

$$\sum_{i=1}^{n}\left[(x_i - a_i)^2 - \theta^2(x_i - b_i)^2\right] = \sum_{i=1}^{n}\left[(1-\theta^2)x_i^2 - 2(a_i - b_i)x_i + (a_i^2 - b_i^2)\right] \leqslant 0 \tag{3.9}$$

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: $\mathbf{A} = \mathrm{diag}((1-\theta^2), ..., (1-\theta^2))$. Поскольку по условию $0 \leqslant \theta \leqslant 1$, 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 $\|\mathbf{x} - \mathbf{a}\|_2 \leqslant \|\mathbf{x} - \mathbf{b}\|_2$, 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 множество. Однако если быть до конца честным, можно исследовать аффинность множества решений квадратичного неравенства, т.е. при $\theta \neq 1$. Пусть $\mathbf{x} = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2$, $\sum\limits_{i=1}^{n}[(1-\theta^2)x_{1i}^2 - 2(a_i - b_i)x_{1i} + (a_i^2 - b_i^2)] = y_1$, $\sum\limits_{i=1}^{n}[(1-\theta^2)x_{2i}^2 - 2(a_i - b_i)x_{2i} + (a_i^2 - b_i^2)] = y_2$, $y_2 > y_1$

Легко видеть, что при

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

выполняется

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

Множество не афинное.

3.5

Проверим на выпуклость:

$$\mathbf{F_0} + \sum_{i=1}^{n} (\theta x_{1i} + (1-\theta) x_{2i}) \mathbf{F_i} = \theta(\mathbf{F_0} + \sum_{i=1}^{n} x_{1i}\mathbf{F_i}) + (1-\theta)(\mathbf{F_0} + \sum_{i=1}^{n} x_{2i}\mathbf{F_i}) \succeq 0 \tag{3.12}$$

Множество выпукло. Проверим теперь на аффинность:

Пусть $\mathbf{a^T} \left[\mathbf{F_0} + \sum\limits_{i=1}^{n} x_{1i}\mathbf{F_i}\right] \mathbf{a} = \alpha$, $\mathbf{a^T} \left[\mathbf{F_0} + \sum\limits_{i=1}^{n} x_{2i}\mathbf{F_i}\right] \mathbf{a} = \beta$, $\beta > \alpha$

$$\mathbf{a^T} \left[\mathbf{F_0} + \sum_{i=1}^{n} (\theta x_{1i} + (1-\theta) x_{2i}) \mathbf{F_i}\right] \mathbf{a} = \beta + \theta(\alpha - \beta) \tag{3.13}$$

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

$$\mathbf{a^T} \left[\mathbf{F_0} + \sum_{i=1}^{n} (\theta x_{1i} + (1-\theta) x_{2i}) \mathbf{F_i}\right] \mathbf{a} = -1 < 0 \tag{3.14}$$

Множество не аффинное.

4 Градиенты и гессианы

4.1

4.1.1 a

Как известно, след матрицы — сумма собственных чисел.

$$f(\mathbf{X}) = \mathrm{Tr}(\mathbf{X}) = \sum_i x_{ii} \tag{4.1}$$

Скалярное представление градиента:

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

Векторное представление градиента:

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

4.1.2 b

Из уравнения на собственные числа и теоремы Виета

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

следует, что

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

Скалярное представление функции:

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

Скалярное представление градиента:

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

Векторное представление градиента:

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

4.2

Матричное представление функции:

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

Скалярное представление функции:

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

Скалярное представление градиентов:

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

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

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

Матричное представление градиентов:

$$\nabla_u J(\mathbf{U}, \mathbf{V}) = 2(\mathbf{UV} - \mathbf{Y})\mathbf{V}^{\mathrm{T}} + \lambda\mathbf{U} \tag{4.13}$$

$$\nabla_v J(\mathbf{U}, \mathbf{V}) = 2\mathbf{U}^{\mathrm{T}}(\mathbf{UV} - \mathbf{Y}) + \lambda\mathbf{V} \tag{4.14}$$

4.3

Матричное представление функции:

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

Скалярное представление:

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

Скалярное представление градиента:

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

Матричное представление градиента:

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

Скалярное представление гессиана:

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

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

Матричное представление гессиана:

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

5 Кратчайший путь в графе

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

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

где

$$P_{pq}(\mathbf{c}) = \begin{cases} 1,\ (p,q) \subseteq (i,j) \\ 0,\ (p,q) \nsubseteq (i,j) \end{cases} \tag{5.2}$$

$$\tag{5.3}$$

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 $\tilde{\mathbf{c}} = \theta \mathbf{c^a} + (1-\theta)\mathbf{c^b}$

$$p_{ij}(\tilde{\mathbf{c}}) = \sum P_{pq}(\tilde{\mathbf{c}})(\theta c^a_{pq} + (1-\theta)c^b_{pq}) = \theta \sum P_{pq}(\tilde{\mathbf{c}}) c^a_{pq} + (1-\theta) \sum P_{pq}(\tilde{\mathbf{c}}) c^b_{pq} \tag{5.4}$$

Проверим на выпуклость или вогнутость:

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

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

$$\begin{aligned} \sum P_{pq}(\tilde{\mathbf{c}}) c^a_{pq} \geqslant \sum P_{pq}(\mathbf{c^a}) c^a_{pq} \\ \sum P_{pq}(\tilde{\mathbf{c}}) c^b_{pq} \geqslant \sum P_{pq}(\mathbf{c^b}) c^b_{pq} \end{aligned} \tag{5.6}$$

Следовательно, функция кратчайшего пути является вогнутой:

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

Эта вогнутость не является строгой. Рассмотрим граф

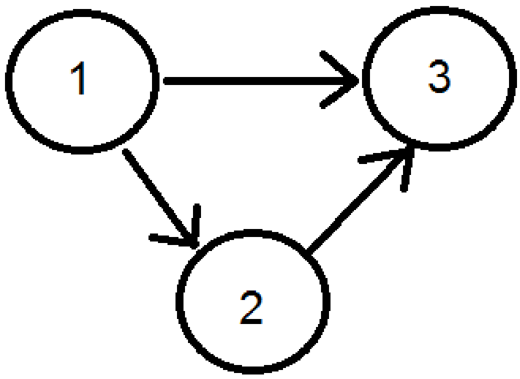

Рис. 1: Граф.

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:

$$\begin{aligned}&\tilde{t}^2 - \tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}} = \theta^2 t_1^2 + (1-\theta)^2 t_2^2 + 2\theta(1-\theta)t_1t_2 - \theta^2\mathbf{x}_1^T - (1-\theta)^2\mathbf{x}_2^T - \theta(1-\theta)\mathbf{x_1}^{\mathrm{T}}\mathbf{x_2} - \theta(1-\theta)\mathbf{x_2}^{\mathrm{T}}\mathbf{x_1} = \\ &= \theta^2(t_1^2 - \mathbf{x_1}^{\mathrm{T}}\mathbf{x_1}) + (1-\theta)^2(t_2^2 - \mathbf{x_2}^{\mathrm{T}}\mathbf{x_2}) - 2\theta(1-\theta)(t_1t_2 - \mathbf{x_1}^{\mathrm{T}}\mathbf{x_2})\end{aligned} \tag{6.1}$$

Проверим, верно ли неравенство

$$\tilde{t}^2 - \tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}} \geqslant \theta(t_1^2 - \mathbf{x_1}^{\mathrm{T}}\mathbf{x_1}) + (1-\theta)(t_2^2 - \mathbf{x_2}^{\mathrm{T}}\mathbf{x_2}) \tag{6.2}$$

Вычтем правую часть из левого и получим

$$\begin{aligned}&\tilde{t}^2 - \tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}} - \theta(t_1^2 - \mathbf{x_1}^{\mathrm{T}}\mathbf{x_1}) + (1-\theta)(t_2^2 - \mathbf{x_2}^{\mathrm{T}}\mathbf{x_2}) = \\ &= \theta(1-\theta)\left[(t_1^2 - \mathbf{x_1}^{\mathrm{T}}\mathbf{x_1}) + (t_2^2 - \mathbf{x_2}^{\mathrm{T}}\mathbf{x_2}) - 2(t_1t_2 - \mathbf{x_1}^{\mathrm{T}}\mathbf{x_2})\right] = \\ &= \theta(1-\theta)\left[(t_1 - t_2)^2 - (\mathbf{x_2} - \mathbf{x_1})^{\mathrm{T}}(\mathbf{x_2} - \mathbf{x_1})\right]\end{aligned} \tag{6.3}$$

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

Теперь проверим неравенство

$$-\log(\tilde{t}^2 - \tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}) \leqslant -\theta\log(t_1^2 - \mathbf{x_1}^{\mathrm{T}}\mathbf{x_1}) - (1-\theta)\log(t_2^2 - \mathbf{x_2}^{\mathrm{T}}\mathbf{x_2}) \tag{6.4}$$

Избавляясь от логарифмов, получим

$$(\tilde{t}^2 - \tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}) \geqslant (t_1^2 - \mathbf{x_1}^{\mathrm{T}}\mathbf{x_1})^{\theta}(t_2^2 - \mathbf{x_2}^{\mathrm{T}}\mathbf{x_2})^{1-\theta} \tag{6.5}$$

Воспользуемся неравенством Юнга

$$a^{\theta}b^{1-\theta} \leqslant \theta a + (1-\theta)b \tag{6.6}$$

Из неравенства Юнга и (6.2) следует

$$(\tilde{t}^2 - \tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}) \geqslant \theta(t_1^2 - \mathbf{x_1}^{\mathrm{T}}\mathbf{x_1}) + (1-\theta)(t_2^2 - \mathbf{x_2}^{\mathrm{T}}\mathbf{x_2}) \geqslant (t_1^2 - \mathbf{x_1}^{\mathrm{T}}\mathbf{x_1})^{\theta}(t_2^2 - \mathbf{x_2}^{\mathrm{T}}\mathbf{x_2})^{1-\theta} \tag{6.7}$$

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

$$f(\lambda_1\mathbf{x_1} + ... + \lambda_n\mathbf{x_n}) \geqslant \lambda_1 f(\mathbf{x_1}) + ... + \lambda_n f(\mathbf{x_n}) \tag{7.1}$$

Докажем неравенство, равносильное данному

$$f(\mathbf{x_1}) \leqslant \frac{1}{\lambda_1} f(\lambda_1\mathbf{x_1} + ... + \lambda_n\mathbf{x_n}) + \sum_{i=2}^{n}(-\frac{\lambda_i}{\lambda_1}) f(\mathbf{x_i}) \tag{7.2}$$

Сделаем преобразование переменных. Пусть $\sum\limits_{i=1}^{n}\lambda_i\mathbf{x_i} = \tilde{\mathbf{x_1}}$, $\mathbf{x_i} = \tilde{\mathbf{x_i}}, i = 2, ..., n$. Неравенство преобразуется в

$$f(\frac{1}{\lambda_1}\tilde{\mathbf{x_1}} + \sum_{i=2}^{n}(-\frac{\lambda_i}{\lambda_1})\tilde{\mathbf{x_i}}) \leqslant \frac{1}{\lambda_1} f(\tilde{\mathbf{x_1}}) + \sum_{i=2}^{n}(-\frac{\lambda_i}{\lambda_1}) f(\tilde{\mathbf{x_i}}) \tag{7.3}$$

Поскольку $\lambda_1 > 0$, $\lambda_i \leqslant 0, i = 2, ..., n$, то $\lambda_1 = 1 - \sum\limits_{i=2}^{n}\lambda_i \geqslant 1$, и, соответственно, $0 < \dfrac{1}{\lambda_1} \leqslant 1$.

В то же время $\lambda_1 = 1 - \sum\limits_{i=2}^{n}\lambda_i = 1 + \sum\limits_{i=2}^{n}(-\lambda_i) = 1 + \sum\limits_{i=2}^{n}|\lambda_i| > |\lambda_i|$, и, следовательно, $0 \leqslant (-\dfrac{\lambda_i}{\lambda_1}) < 1$.

Переобозначим $\theta_1 = \dfrac{1}{\lambda_1}$, $\theta_i = (-\dfrac{\lambda_i}{\lambda_1}), i = 2, ..., n$.

Как видно, $0 < \theta_1 \leqslant 1$, $0 \leqslant \theta_i < 1$, $\sum\limits_i \theta_i = \dfrac{1 - \sum\limits_{i=2}^{n}\lambda_i}{\lambda_1} = 1$. Неравенство перепишется в виде

$$f(\sum_{i=1}^{n}\theta_i\tilde{\mathbf{x_i}}) \leqslant \sum_{i=1}^{n}\theta_i f(\tilde{\mathbf{x_i}}) \tag{7.4}$$

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:

$$\begin{aligned}
&\tilde{\mathbf{x}} = \theta\mathbf{x_1} + (1-\theta)\mathbf{x_1} \\
&\tilde{y} = f(\theta\mathbf{x_1} + (1-\theta)\mathbf{x_2}) \\
&\tilde{\tilde{y}} = \theta f(\mathbf{x_1}) + (1-\theta) f(\mathbf{x_2}) = \theta y_1 + (1-\theta) y_2.
\end{aligned} \tag{8.1}$$

Рассмотрим следующие случаи:

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

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

$$h(f(\tilde{\mathbf{x}})) = h(\tilde{y}) \leqslant h(\tilde{\tilde{y}}) \leqslant \theta h(y_1) + (1-\theta)h(y_2) = \theta h(f(\mathbf{x_1})) + (1-\theta)h(f(\mathbf{x_2})) \tag{8.2}$$

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 $\tilde{y} \geqslant \tilde{\tilde{y}}$. Из невозрастания $h(y)$ следует $h(\tilde{y}) \leqslant h(\tilde{\tilde{y}})$. А из выпуклости $h(y)$ следует, что $h(\tilde{\tilde{y}}) \leqslant \theta h(y_1) + (1-\theta)h(y_2)$

$$h(f(\tilde{\mathbf{x}})) = h(\tilde{y}) \leqslant h(\tilde{\tilde{y}}) \leqslant \theta h(y_1) + (1-\theta)h(y_2) = \theta h(f(\mathbf{x_1})) + (1-\theta)h(f(\mathbf{x_2})) \tag{8.3}$$

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

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

$$\frac{\partial g}{\partial x_i} = h' \frac{\partial f}{\partial x_i} \tag{8.4}$$

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

$$\mathbf{H_g} = h'' \nabla f \otimes (\nabla f)^{\mathrm{T}} + h' \mathbf{H_f} \tag{8.6}$$