Решение домашнего задания № 1

Роман Тарасов, группа 876

1. Задача первая

Пусть путешественник может находиться в некоторой области пространства $X \in \mathbb{R}^n$, и известные приблизительные расстояния до m объектов равны $x_i \in \mathbb{R}, i \in \overline{1, m}$, а координаты самих объектов равны $\mathbf{y}_j \in X, j \in \overline{1, m}$. Нам нужно найти такую точку $\mathbf{x} \in X$, чтобы расстояние от $\mathbf{x}$ до объектов как можно меньше отличались от соответствующих приблизительных расстояний. Введём вектор $\mathbf{u} \in \mathbb{R}^m$, такой что $u_i = ||\mathbf{x} - \mathbf{y}_1||_2 - x_1$ есть разность расстояние от точки $\mathbf{x}$ до i-того объекта минус приблизительноое расстояния. Оптимизационная задача выглядит так:

$$\min_{\mathbf{x} \in X} ||\mathbf{u}||$$

Причём норму можно взять какую может быть нужно в данном случае. Например, если взять первую норму, то задача будет выглядеть так:

$$\min_{\mathbf{x} \in X} \sum_{i=1}^{m} \Big| ||\mathbf{x} - \mathbf{y}_1||_2 - x_1 \Big|$$

2. Задача вторая

Сделаем несколько предположений насчёт входных данных. Во-первых, будем считать, что если $p_{ij} < 0$, то студент i не хочет жить со студентом j. То есть, i лучше бы выбрал пустое место в комнате, чем студента j на это место. Если $p_{ij} = 0$, то человеку i без разницы, будет ли там пустое место или студент j. В случае когда $p_{ij} > 0$, студент i хорошо относится к j и выбрал бы лучше жить с ним, чем вместо j свободное место в комнате.

Перед тем, как решать задачу, надо проверить её на корректность. То есть что мы вообще можем поселить всех студентов в комнаты: $n \leqslant 3m$.

Пусть $\mathbf{S}$ – матрица размера $m \times n$. $\mathbf{S}_{ij} = 1$, если j-тый студент поселился в i-тую комнату, и $\mathbf{S}_{ij} = 0$, если j-тый студент поселился не в i-тую комнату. Назовём $\mathbf{S}$ матрицей поселения. Заметим, что в силу условия задачи в одном столбце должна быть ровно одна единица (каждый студент поселён ровно в одну комнату) и в каждой строке не больше 3 единиц (в комнате может быть не больше 3 человек).

Рассмотрим матрицу $\mathbf{R} = \mathbf{S}^\top \mathbf{S}$. $\mathbf{R}_{ij} = \sum\limits_{\alpha=1}^{m} \mathbf{S}_{\alpha i} \mathbf{S}_{\alpha j}$. В этой сумме α пробегает по всем номерам комнат, и если студенты i, j живут в одной комнате k, то $\mathbf{R}_{ij} = \sum\limits_{\alpha=1}^{m} \mathbf{S}_{\alpha i} \mathbf{S}_{\alpha j} = \mathbf{S}_{ki} \mathbf{S}_{kj} = 1 \cdot 1 = 1$. Если i, j живут в разных комнатах, то $\mathbf{R}_{ij} = 0$.

Обозначим матрицы $\mathbf{P} = ||p_{ij}|| \in \mathbb{R}^{n \times n}$, $\mathbf{B} = ||b_{ik}|| \in \mathbb{R}^{n \times m}$.

Рассмотрим некоторую матрицу поселения $\mathbf{S}$ и следующие суммы:

$$\mathrm{trace}(\mathbf{RP}) = \sum_{1 \leqslant i,j \leqslant n} \mathbf{R}_{ij}\mathbf{P}_{ij}$$

$$\mathrm{trace}(\mathbf{SB}) = \sum_{k=1}^{m}\sum_{i=1}^{n} \mathbf{S}_{ki}\mathbf{B}_{ik}$$

Заметим, что первая сумма это есть сумма всех предпочтений p_{ij} таких, что студенты i, j живут в одной комнате. А вторая сумма это есть сумма всех предпочтений b_{ik} таких, что студент i живёт в комнате k. Нам нужно эти суммы максимизировать для того, чтобы удовлетворённость поселением была как можно больше. То есть, максимизировать сумму $\mathrm{trace}(\mathbf{RP})\mathrm{trace}(\mathbf{SB}) = \mathrm{trace}(\mathbf{RP} + \mathbf{SB})$.

Итак, задача выглядит следующим образом:

$$\max_{\mathbf{S}} \mathrm{trace}(\mathbf{RP} + \mathbf{SB})$$

$$s.t. : \mathbf{S} \in \{0,1\}^{m \times n}$$

$$\sum_{i=1}^{n} \mathbf{S}_{ki} \leqslant 3,\ \sum_{k=1}^{m} \mathbf{S}_{ki} = 1$$

Если же известны не все p_{ij}, то можно им присвоить какие-то значение. Можно для простоты присвоить 0, если предположить, что студентам нет большой разницы, жить с незнакомым студентом или же вместо него пустое место. Можно присвоить неизвестному p_{ij} значение среднего арифметического всех известных предпочтений студента p_{ik}. В этом случае мы предполагаем, что если студент ладит с большинством из известных ему людей, то и с незнакомым студентом он поладит с большей вероятностью. В обоих этих случаях остальная часть постановки задачи не изменится.

3. Задача третья

Обозначим $\mathbf{S}$ – матрица размера $n \times m$, в которой $\mathbf{S}_{ij}$ – количество единиц товара, которые надо перевезти из i-того склада в j-тый магазин. Назовём её матрицей перевозки. Также обозначим $\mathbf{C} = ||c_{ij}|| \in \mathbb{R}^{n \times m}$, $\mathbf{T} = ||t_{ij}|| \in \mathbb{R}^{n \times m}$.

Тогда ограничения из условия можно переписать в виде $\sum\limits_{i=1}^{n} \mathbf{S}_{ij} = b_j$, $\sum\limits_{j=1}^{m} \mathbf{S}_{ij} \leqslant a_i$. Для проверки корректности задачи нужно убедиться, что $\sum\limits_{i=1}^{n} a_i \geqslant \sum\limits_{j=1}^{m} b_j$.

Суммарная стоимость перевозки будет равна $\sum\limits_{i,j} \mathbf{S}_{ij}\mathbf{C}_{ij} = \mathrm{trace}(\mathbf{S}^\top \mathbf{C})$, а суммарное время – $\sum\limits_{i,j} \mathbf{S}_{ij}\mathbf{T}_{ij} = \mathrm{trace}(\mathbf{S}^\top \mathbf{T})$.

В зависимости от того, что нам важнее: уменьшить стоимость доставки или уменьшить время, можно по-разному ставить задачу. Можно ввести коэффициенты важности α, β, такие что $\alpha, \beta \geqslant 0$ и $\alpha + \beta = 2$. Тогда нам нужно минимизировать сумму $\alpha \cdot \mathrm{trace}(\mathbf{S}^\top \mathbf{C}) + \beta \cdot \mathrm{trace}(\mathbf{S}^\top \mathbf{T}) = \mathrm{trace}(\mathbf{S}^\top(\alpha\mathbf{C} + \beta\mathbf{T}))$ по матрице $\mathbf{S}$. При этом параметры α, β выбираем сами в зависимости от того, что нам нужнее в данном случае: чем больше α, тем нам важнее минимизировать стоимость, чем больше β, тем дороже время. В частности, если нам без разницы на стоимость, то $\alpha = 0$, если не имеет значение время, то $\beta = 0$.

Задача выглядит так:

$$\min_{\mathbf{S}} \operatorname{trace}(\mathbf{S}^\top(\alpha\mathbf{C} + \beta\mathbf{T}))$$

s.t.:

$$\sum_{i=1}^{n} \mathbf{S}_{ij} = b_j,\ \sum_{j=1}^{m} \mathbf{S}_{ij} \leqslant a_i$$

$$\mathbf{S} \in (\mathbb{N} \cup \{0\})^{n\times m}$$

4. Задача четвёртая

Уравнение эллипса: $\mathbf{E} = \{(\mathbf{x}-\mathbf{x}_0)^\top \mathbf{P}^{-1}(\mathbf{x}-\mathbf{x}_0) = 1, \mathbf{x} \in \mathbb{R}^2\}$, где $\mathbf{P}$ – некоторая симметричная положительно определённая матрица. Нам нужно выбрать эллипс так, чтобы станция тратила минимальное количество энергии, то есть эллипс с минимальной площадью, покрывающий все базовые станции (координаты базовых станций равны $\mathbf{x}_i \in bbR^2$, $i \in \overline{1,m}$). Площадь эллипса $\mathbf{E}$ равна $\frac{\pi}{\sqrt{\det \mathbf{P}}}$. Значит, нужно максимизировать $\det \mathbf{P}$.

Задача выглядит так:

$$\max_{\mathbf{x}_0, \mathbf{P}} \det \mathbf{P}$$

s.t.:

$$(\mathbf{x}_i - \mathbf{x}_0)^\top \mathbf{P}^{-1}(\mathbf{x}_i - \mathbf{x}_0) \leqslant 1,\ \forall i \in \overline{1,m}$$

$$\mathbf{P} \succ 0$$

5. Задача пятая

Пусть $||.||$ – некоторая норма $\mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{n\times n}$ – невырожденная матрица. Докажем, что $\rho(\mathbf{x},\mathbf{y}) = ||\mathbf{A}(\mathbf{x}-\mathbf{y})||$ – метрика.

1. $\rho(\mathbf{x},\mathbf{y}) \geq 0$, причём $\rho(\mathbf{x},\mathbf{y}) = 0 \Leftrightarrow ||\mathbf{A}(\mathbf{x}-\mathbf{y})|| = 0 \Leftrightarrow \mathbf{x} = \mathbf{y}$
2. $\rho(\mathbf{x},\mathbf{y}) = ||\mathbf{A}(\mathbf{x}-\mathbf{y})|| = ||\mathbf{A}(\mathbf{y}-\mathbf{x})|| = \rho(\mathbf{y},\mathbf{x})$
3. $\rho(\mathbf{x},\mathbf{y}) = ||\mathbf{A}(\mathbf{x}-\mathbf{y})|| = ||\mathbf{A}(\mathbf{x}-\mathbf{z})+\mathbf{A}(\mathbf{z}-\mathbf{x})|| \leq ||\mathbf{A}(\mathbf{x}-\mathbf{z})||+||\mathbf{A}(\mathbf{z}-\mathbf{y})|| = \rho(\mathbf{x},\mathbf{z})+\rho(\mathbf{y},\mathbf{z})$

Таким образом можно задать параметризацию метрики.

6. Задача шестая

6.1 Пункт 1

Норма это функция, заданная на некотором линейном пространстве $\mathcal{X}$:

$$||.||: \ \mathcal{X} \to \mathbb{R}_+$$

Удовлетворяет следующим свойствам:

1. $\forall \mathbf{x} \in \mathcal{X} \ \ ||\mathbf{x}|| \geq 0, ||\mathbf{x}|| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$
2. $\forall \alpha \in \mathbb{R} \forall \mathbf{x} \in \mathcal{X} \ \ ||\alpha \mathbf{x}|| = \alpha ||\mathbf{x}||$
3. $\forall \mathbf{x}, \mathbf{y} \in \mathcal{X} \ \ ||\mathbf{x}|| + ||\mathbf{y}|| \geq ||\mathbf{x} + \mathbf{y}||$

Говорят, что нормы $||.||_1$ и $||.||_2$ (это любые нормы) заданные на линейном пространстве $\mathcal{X}$ эквивалентны, если $\exists C_1, C_2 \in \mathbb{R}$, такие что $\forall \mathbf{x} \in \mathcal{X} \ \ C_1||\mathbf{x}||_1 \leq ||\mathbf{x}||_2 \leq C_2||\mathbf{x}||_1$.

Если $||\mathbf{x}||_2 \leq C_2||\mathbf{x}||_1$, то говорят, что норма $||.||_2$ подчинена норме $||.||_1$.

Нормы ℓ_1, ℓ_2, ℓ_∞ заданы на пространстве $\mathbb{R}^n$. Далее будем их обозначать просто соответствующими индексами.

$$||\mathbf{x}||_1 = \sum_{i=1}^{n} |x_i|$$

$$||\mathbf{x}||_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

$$||\mathbf{x}||_\infty = \max_i |x_i|$$

Докажем, что первая норма подчинена второй.

$$\sqrt{n}||\mathbf{x}||_2 = \sqrt{n}\sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{\sum_{i=1}^{n} 1^2}\sqrt{\sum_{i=1}^{n} x_i^2}$$

По неравенству КБШ

$$\sqrt{\sum_{i=1}^{n} 1^2}\sqrt{\sum_{i=1}^{n} x_i^2} \geq \sum |x_i| \ |1| = ||\mathbf{x}||_1$$

Теперь покажем, что вторая норма подчинена норме ∞

$$\sqrt{n}||\mathbf{x}||_\infty = \sqrt{n}\max_i |x_i| = \sqrt{\sum_{j=1}^{n} (\max_i |x_i|)^2} \geq \sqrt{\sum_{j=1}^{n} x_j^2} = ||\mathbf{x}||_2$$

Ну и норма ∞ подчинена первой, так как $||\mathbf{x}||_1 = \sum_{i=1}^{n} |x_i| \geq \max_i |x_i| = ||\mathbf{x}||_\infty$.

Следовательно, эти нормы попарно эквивалентны.

6.2 Пункт 2

Норма матрицы, порождённая векторной нормой $||.||$:

$$||\mathbf{A}|| = \sup_{||\mathbf{x}||=1} ||\mathbf{Ax}||$$

Норма матрицы L_1:

$$||\mathbf{A}||_1 = \sup_{||x_i||_1=1} \sum_i \left| (\mathbf{Ax})_i \right|$$

Выведем выражение для этой формулы

$$\sum_i \left| (\mathbf{Ax})_i \right| = \sum_i \left| \sum_j |a_{ij}| \ |x_j| \right| \leq \sum_{i,j} |a_{ij}| \ |x_j| = \sum_j \left(\sum_i |a_{ij}| \right) |x_j| \leq$$

$$\leq \sum_j \left(\max_j \sum_i |a_{ij}| \right) |x_j| = \max_j \sum_i |a_{ij}| \cdot \sum_j |x_j| = \max_j \sum_i |a_{ij}|$$

Теперь покажем, что равенство достигается. Пусть $\max\limits_j \sum_i |a_{ij}| = \sum_i |a_{ik}|$. Тогда рассмотрим вектор $\mathbf{x}$, у которого k-тая компонента равна 1, а остальные 0. Тогда $||\mathbf{Ax}||_1 = \sum_i |a_{ik}|$.

Норма матрицы L_2:

$$||\mathbf{A}||_2 = \sup_{||x_i||_2=1} \sqrt{\sum_i (\mathbf{Ax})_i^2}$$

Теперь выведем формулу:

$$\sqrt{\sum_i (\mathbf{Ax})_i^2} = \sqrt{(\mathbf{Ax}, \mathbf{Ax})} = \sqrt{(\mathbf{A}^*\mathbf{Ax}, \mathbf{x})} = \sqrt{(\mathbf{A}^\top\mathbf{Ax}, \mathbf{x})}$$

Здесь $\mathbf{A}^*$ – сопряжённая матрица. В ОНБ она равна просто $\mathbf{A}^\top$. $\mathbf{A}^\top\mathbf{A}$ – симметричная матрица, поэтому у неё все собственные значения $\lambda_1, \ldots, \lambda_n$ действительные и из её собственных векторов можн составить ОНБ $\omega_1, \ldots, \omega_n$. Поэтому любой вектор $\mathbf{x}$ раскладывается по этим векторам. $\mathbf{x} = \sum\limits_i \xi_i\omega_i$. $(\mathbf{A}^\top\mathbf{Ax}, \mathbf{x}) = \left(\sum\limits_i \lambda_i\xi_i\omega_i, \xi_i\omega_i \right) = \sum\limits_i \lambda_i\xi_i^2$. Причём $||\mathbf{x}||_2 = \sqrt{\left(\sum\limits_i \xi_i\omega_i, \sum\limits_i \xi_i\omega_i \right)} = \sqrt{\xi_i^2} = 1$.

Из этого следует, что $||\mathbf{A}||_2 \leq \sqrt{\sum\limits_i \lambda_i\xi_i^2} \leq \sqrt{\lambda_{max} \cdot 1^2} = \sqrt{\lambda_{max}}$. Причём максимум достигается на векторе ω_k, где $\lambda_k = \lambda_{max}$.

Норма матрицы L_∞:

$$L_\infty = \sup_{||x_i||_2=1} \max_i \left| (\mathbf{Ax})_i \right|$$

Выведем её:

$$\max_i \left|(\mathbf{Ax})_i\right| = \max_i \left|\sum_j |a_{ij}|\ |x_j|\right| \leq \left(\max_i \sum |a_{ij}|\right) \max_j |x_j| = \max_i \sum |a_{ij}|$$

Несложно проверить, что максимум достигается на векторе $\mathbf{x} = \mathbf{1}^\top$

Итак,

$$||\mathbf{A}||_1 = \max_j \sum_i |a_{ij}|$$

$$||\mathbf{A}||_2 = \sqrt{\lambda_{max}(\mathbf{A}^\top \mathbf{A})}$$

$$||\mathbf{A}||_\infty = \max_i \sum_j |a_{ij}|$$

Фробениусова (или евклидова) норма это норма $||\mathbf{A}||_F = \sqrt{\sum\limits_{i,j} a_{ij}^2}$. Это аналог евклидовой векторной нормы $||\mathbf{x}|| = \sqrt{\sum\limits_i x_i^2}$.

6.3 Пункт 3

6.4 Пункт 4

Пусть $\mathbf{A} \in \mathbb{R}^{m\times n},\ \mathbf{B} \in \mathbb{R}^{n\times k},\ \mathbf{C} \in \mathbb{R}^{k\times l}$. Рассмотрим ij-тый элемент матрицы $(\mathbf{AB})\mathbf{C}$.

$$(\mathbf{AB})\mathbf{C} = \sum_{\beta=1}^{k} (\mathbf{AB})_{i\beta}\mathbf{C}_{\beta j} = \sum_{\beta=1}^{k}\left(\sum_{\alpha=1}^{n} \mathbf{A}_{i\alpha}\mathbf{B}_{\alpha\beta}\right)\mathbf{C}_{\beta j} =$$

$$= \sum_{1\leq\alpha\leq n,\ 1\leq\beta\leq k} \mathbf{A}_{i\alpha}\mathbf{B}_{\alpha\beta}\mathbf{C}_{\beta j} = \sum_{\alpha=1}^{n} \mathbf{A}_{i\alpha} \sum_{\beta=1}^{k} (\mathbf{B}_{\alpha\beta}\mathbf{C}_{\beta j}) = \sum_{\alpha=1}^{n} \mathbf{A}_{i\alpha}(\mathbf{BC})_{\alpha j}$$

Последнее число и есть ij элемент матрицы $\mathbf{A}(\mathbf{BC})$. Это верно для произвольной пары ij, где $i \in \overline{1,m},\ j \in \overline{1,l}$. Поэтому верно равенство $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.

7. Задача седьмая

7.1 Пункт 1

Множество точек, равноудалённых от $\mathbf{x}_0$ и $\mathbf{x}_i$, есть гиперплоскость, проходящая через середину отрезка, соединяющего эти 2 точки и перпендикулярная этому отрезку. Следовательно, множество $\{\mathbf{x} \mid ||\mathbf{x} - \mathbf{x}_0||_2 \leq ||\mathbf{x} - \mathbf{x}_i||_2\}$ есть полуплоскость, лежащая в той же части гиперплоскости, что и $\mathbf{x}_0$.

Вектор $\mathbf{x}_0 - \mathbf{x}_i$ параллелен отрезку, соединяющему точки $\mathbf{x}_0$ и $\mathbf{x}_i$. Поэтому искомая гиперплоскость имеет вид $\langle \mathbf{x}, \mathbf{x}_0 - \mathbf{x}_i\rangle = b_i$, где b_i – некоторое число. Его мы можем определить, зная, что середина отрезка, то есть точка $\frac{\mathbf{x}_0+\mathbf{x}_i}{2}$, принадлежит данной гиперплоскости. $\langle \frac{\mathbf{x}_0+\mathbf{x}_i}{2}, \mathbf{x}_0 - \mathbf{x}_i\rangle = \frac{1}{2}\left(||\mathbf{x}_0||_2^2 - ||\mathbf{x}_i||_2^2\right)$. Итак, уравнение гиперплоскости $\langle \mathbf{x}, \mathbf{x}_0 - \mathbf{x}_i\rangle = \frac{1}{2}\left(||\mathbf{x}_0||_2^2 - ||\mathbf{x}_i||_2^2\right)$. Нам нужны точки, которые ближе к $\mathbf{x}_0$, то есть, та полуплоскость, в которой лежит $\mathbf{x}_0$. Несложно проверить, что $\mathbf{x}_0$ принадлежит именно этой полуплоскости: $\langle \mathbf{x}, \mathbf{x}_0 - \mathbf{x}_i \rangle \geq \frac{1}{2}\left(||\mathbf{x}_0||_2^2 - ||\mathbf{x}_i||_2^2\right)$ или $(\mathbf{x}_0 - \mathbf{x}_i)^\top \mathbf{x} \geq \frac{1}{2}\left(||\mathbf{x}_0||_2^2 - ||\mathbf{x}_i||_2^2\right)$.

Так как полученное неравенство должно выполняться для всех точек одновременно, то получаем систему $\mathbf{A}\mathbf{x} = \mathbf{b}$, где i-тая строка матрицы $\mathbf{A}$ есть строка $(\mathbf{x}_0 - \mathbf{x}_i)^\top$, а $b_i = \frac{1}{2}\left(||\mathbf{x}_0||_2^2 - ||\mathbf{x}_i||_2^2\right)$. То есть, область Вороного это выпуклый многогранник.

7.2 Пункт 2

Теперь покажем, как можно получить точки $\mathbf{x}_0, \ldots, \mathbf{x}_k$ имея уравнение выпуклого многогранника $\mathbf{A}\mathbf{x} \geq \mathbf{b}$. Для удобства обозначим x_j^i – это j-тая компонента i-того вектора.

Помним, что в предыдущем пункте i-тая строка матрицы $\mathbf{A}$ – это $(\mathbf{x}_0 - \mathbf{x}_i)^\top$. Поэтому $x_j^0 - x_j^i = a_{ij}$. Также имеем $2b_i = ||\mathbf{x}_0||_2^2 - ||\mathbf{x}_i||_2^2 = \sum\limits_j (x_j^0)^2 - \sum\limits_j (x_j^i)^2$.

Подставим значения $x_j^i = x_j^0 - a_{ij}$ в выражение для b_i. Получим

$$2b_i = \sum_j (x_j^0)^2 - \sum_j (x_j^0 - a_{ij})^2 = \sum_{j=1}^n (2x_j^0 a_{ij} - a_{ij}^2)$$

Здесь индекс i пробегает значения от 1 до k. То есть мы получили систему из k уравнений и n неизвестных $x_1^0, \ldots, x_n^0$. Если у этой системы есть решения (возможно не одно), то, подставив найденный вектор $\mathbf{x}_0$ в равенства $x_j^i = x_j^0 - a_{ij}$, получим искомые точки $\mathbf{x}_0, \ldots, \mathbf{x}_k$. То есть, $\mathbf{x}_0$, в зависимости от системы, можно выбрать одним или множеством способов, а зная $\mathbf{x}_0$, можно однозначно определить $\mathbf{x}_1, \ldots, \mathbf{x}_k$. Если же у системы нет решений, то и область Вороного построить нельзя.

7.3 Пункт 3

Обобщение диаграммы Вороного на $p > 1$ точек можно определить следующим образом:

Даны n точек $\mathbf{x}_1, \ldots, \mathbf{x}_n$ в $\mathbb{R}^n$. Нужно найти разбиение пространства на области, в каждой из которых находятся точки те и только те, которые наиболее близки к какой-то из $\mathbf{x}_i$ чем к любой другой точке.

Пусть дано разбиение пространства на многоугольники. Покажем, как можно восстановить множество точек, для которого это разбиение является разбиением Вороного.

Каждый многоугольник из разбиения задаётся системой неравенств $\mathbf{A}\mathbf{x} \leq \mathbf{b}$. Поэтому найти точку

8. Задача восьмая

8.1 Пункт 1

Множество $A = \{\mathbf{x} \in \mathbb{R} \mid \alpha \leqslant \mathbf{a}^\top \mathbf{x} \leqslant \beta\}$ выпукло:

Если $\mathbf{x}_1, \mathbf{x}_2 \in A, a \in [0,1]$, то $\langle \mathbf{a}, a\mathbf{x}_1 + (1-a)\mathbf{x}_2 \rangle = a\langle \mathbf{a}, \mathbf{x}_1 \rangle + (1-a)\langle \mathbf{a}, \mathbf{x}_2 \rangle \geqslant a\alpha + (1-a)\alpha = \alpha$. И аналогично $\langle \mathbf{a}, a\mathbf{x}_1 + (1-a)\mathbf{x}_2 \rangle \leqslant \beta$. Значит, $a\mathbf{x}_1 + (1-a)\mathbf{x}_2 \in A$, и множество выпукло.

Исследуем на аффинность:

Если $\alpha = \beta$, то $A = \{\mathbf{x} \in \mathbb{R} \mid \mathbf{a}^\top \mathbf{x} = \beta\}$. Оно аффинно по критерию аффинности.

Если же $\alpha < \beta$, то рассмотрим точки $\frac{\alpha}{||\mathbf{a}||^2}\mathbf{a}$ и $\frac{\beta}{||\mathbf{a}||^2}\mathbf{a}$. Они принадлежат множеству A, так как $\langle \mathbf{a}, \frac{\alpha}{||\mathbf{a}||^2}\mathbf{a} \rangle = \alpha \frac{\langle \mathbf{a}, \mathbf{a} \rangle}{||\mathbf{a}||^2} = \alpha$ и $\langle \mathbf{a}, \frac{\beta}{||\mathbf{a}||^2}\mathbf{a} \rangle = \beta \frac{\langle \mathbf{a}, \mathbf{a} \rangle}{||\mathbf{a}||^2} = \beta$. Но в то же время рассмотрим точку

$\frac{2\beta-\alpha}{||\mathbf{a}||^2}\mathbf{a}$, лежащую на одной прямой с теми двумя. $\langle \mathbf{a}, \frac{2\beta-\alpha}{||\mathbf{a}||^2}\mathbf{a}\rangle = (2\beta-\alpha)\frac{\langle \mathbf{a},\mathbf{a}\rangle}{||\mathbf{a}||^2} = 2\beta-\alpha$. Так как $\alpha < \beta$, то $2\beta - \alpha > \beta$, значит, эта точка не принадлежит множеству A и A не аффинно.

Исследуем на коничность:

Пусть одно из чисел α или β (например α) отлично от 0. Если $\alpha < 0$, то рассмотрим точку $\frac{\alpha}{||\mathbf{a}||^2}\mathbf{a}$. Она лежит в A, но так как $2\alpha < \alpha$, то $\langle \mathbf{a}, \frac{2\alpha}{||\mathbf{a}||^2}\mathbf{a}\rangle < \alpha$ и точка $\frac{\alpha}{||\mathbf{a}||^2}\mathbf{a}$ уже не лежит в A. Если $\alpha > 0$, то и $\beta > 0$, $2\beta > \beta$. $\frac{\beta}{||\mathbf{a}||^2}\mathbf{a} \in A$, $\frac{2\beta}{||\mathbf{a}||^2}\mathbf{a} \notin A$. Следовательно, множество не является конусом.

Аналогично если $\beta \neq 0$, то множество не конус.

Если же $\alpha = \beta = 0$, то $A = \{\mathbf{x} \in \mathbb{R} \mid \mathbf{a}^\top \mathbf{x} = 0\}$. Если $\mathbf{x}_1 \in A$, то $\langle \mathbf{a}, \mathbf{x}_1\rangle = 0$. Но тогда и $\forall a > 0$ $\langle \mathbf{a}, a\mathbf{x}_1\rangle = 0$. Следовательно $a\mathbf{x}_1 \in A$ и множество является конусом.

8.2 Пункт 2

$A = \{\mathbf{x} \mid \mathbf{a}_1^\top \mathbf{x} \leqslant b_1, \mathbf{a}_2^\top \mathbf{x} \leqslant b_2\}$

Докажем выпуклость: аналогично первому пункту доказываем, что если $\mathbf{x}_1, \mathbf{x}_2 \in A$, то $\mathbf{a}_1^\top \mathbf{x}_1 \leqslant b_1$, $\mathbf{a}_1^\top \mathbf{x}_2 \leqslant b_1$, $\mathbf{a}_2^\top \mathbf{x}_1 \leqslant b_2$ и $\mathbf{a}_2^\top \mathbf{x}_2 \leqslant b_2$. И отсюда следует, что

$\forall \alpha \in [0,1]$ верно $\mathbf{a}_1^\top(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leqslant b_1$ и $\mathbf{a}_2^\top(\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2) \leqslant b_2$. То есть A выпукло.

8.3 Пункт 3

8.4 Пункт 4

8.5 Пункт 5

Пусть $\mathcal{C} = \bigcap\limits_{i \in I} \mathcal{C}_i$ – пересечение конусов $\mathcal{C}_i$. Пусть $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}$. Тогда $\forall i \in I$ $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}_i$. Так как $\mathcal{C}_i$ – выпуклые, то $\forall i \in I$ $\forall \alpha \in [0,1]$ $\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 \in \mathcal{C}_i$. Это верно для всех конусов из I. Значит, $\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2$ принадлежит и пересечению этих конусов, то есть множнству $\mathcal{C}$. Оно выпукло. А так как $\mathcal{C}_i$ –конусы, то из определения конуса $\forall \beta \geqslant 0$ выполнено $\beta\mathbf{x}_1 \in \mathcal{C}_i$. Так как i – любой элемент из I, то $\beta\mathbf{x}_1$ принадлежит пересечению этих конусов, то есть множеству $\mathcal{C}$. В итоге имеем, что $\forall \beta \geqslant 0$, $\forall \mathbf{x} \in \mathcal{C}$ верно $\beta\mathbf{x} \in \mathcal{C}$. То есть, $\mathcal{C}$ – конус.

8.6 Пункт 6

Рассуждения аналогичны пункту 5.

Пусть $\mathcal{A} = \bigcap\limits_{i \in I} \mathcal{A}_i$ – пересечение аффинных множеств. Пусть $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{A}$. Тогда

$\forall i \in I$ $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{A}_i$. Так как $\mathcal{A}_i$ – аффинные, то $\forall i \in I$ $\forall \alpha \in \mathbb{R}$ $\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 \in \mathcal{A}_i$. То есть $\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 \in \mathcal{A}$. Отсюда следует, что $\mathcal{A}$ – аффинно.

9. Задача девятая

9.1 Пункт 1

Докажем, что множеству M могут принадлежать только вектора, для которых $1 \leqslant ||\mathbf{x}||_2 \leqslant 3$ (далее в этой задаче везде будем подразумевать вторую норму).

Пусть $\mathbf{x} \in M$. Тогда $\exists \mathbf{a} \in M_1, \mathbf{b} \in M_2 : \mathbf{x} = \mathbf{a} + \mathbf{b}$. Поработаем с этим равенством:

$$\mathbf{x} = \mathbf{a} + \mathbf{b}$$

$$||\mathbf{x}|| = ||\mathbf{a} + \mathbf{b}||$$

$$||\mathbf{x}||^2 = ||\mathbf{a}||^2 + ||\mathbf{b}||^2 - 2||\mathbf{a}|| \; ||\mathbf{b}|| \cos\alpha$$

Здесь подразумеваем $\cos\alpha = \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{||\mathbf{a}|| \; ||\mathbf{b}||}$

$$||\mathbf{x}||^2 = 1^2 + 2^2 - 2 \cdot 1 \cdot 2 \cdot \cos\alpha$$

$$\cos\alpha = \frac{5 - ||\mathbf{x}||^2}{4}$$

Так как $-1 \leqslant \cos\alpha \leqslant 1$, то отсюда получаем $-1 \leqslant \frac{5-||\mathbf{x}||^2}{4} \leqslant 1$ и $1 \leqslant ||\mathbf{x}||_2 \leqslant 3$.

А теперь покажем, как получить разложение $\mathbf{x} = \mathbf{a} + \mathbf{b}$ для любого $\mathbf{x}$ что $1 \leqslant ||x|| \leqslant 3$, такое что $\mathbf{a} \in M_1$, $\mathbf{b} \in M_2$.

$$||\mathbf{x} - \mathbf{a}|| = ||\mathbf{b}||$$

$$||\mathbf{x}||^2 + ||\mathbf{a}||^2 - 2\langle \mathbf{a}, \mathbf{x} \rangle = ||\mathbf{b}||^2$$

$$||\mathbf{x}||^2 + 1 - 2a_1x_1 - 2a_2x_2 = 4$$

Учитывая, что $a_1^2 + a_2^2 = ||\mathbf{a}||^2 = 1$ и зная $||\mathbf{x}||$, из последних двух уравнений мы можем найти a_1 и a_2, то есть сам вектор $\mathbf{a}$. Аналогично находим и вектор $\mathbf{b}$.

9.2 Пункт 2

Рассмотрим множества $M_1 = \{\mathbf{x} \mid ||\mathbf{x}||_2 \leqslant 1\}$ и $M_2 = \{\mathbf{x} \mid 1 \leqslant ||\mathbf{x}||_2 \leqslant 2\} \cup \{0\}$. Второе из них невыпукло. Но покажем, что их сумма $M = M_1 + M_2$ есть выпуклое множество $\{\mathbf{x} \mid ||\mathbf{x}||_2 \leqslant 3\}$.

- Вектора, нормы которых больше 3 не могут лежать в сумме Минковского, так как В множестве M_1 норма элемента не больше 1, а в M_2 – не больше 2. Норма суммы не больше суммы норм.

- Элементы с нормой от 1 до 3 могут быть представлены в виде суммы элементов из M_1 и M_2 как показано в пункте 2 (множества M_1 и M_2 из нашего примера содержат в себе соответствующие множества M_1 и M_2 из пункта 1, и поэтому содержат и их сумму Минковского).

- Все вектора с нормой меньше 1 лежат в M_1, и поэтому любой элемент $\mathbf{x}$ с нормой меньше 1 может быть представлен в виде суммы $\mathbf{x} = \mathbf{0} + \mathbf{x}$, где $\mathbf{0} \in M_2,\ \mathbf{x} \in M_1$.

Таким образом, сумма множеств M_1 и M_2 есть выпуклое множество M.

10. Задача десятая

10.1 Пункт 1

Так как $\forall \mathbf{p} \in \mathbb{R}^n \ \langle \mathbf{p}, 0 \rangle = 0 < 1$, то $\{0\}^* = \mathbb{R}$

10.2 Пункт 2

Условие $x_i \geq 0, \ \mathbf{x} \in \mathbb{R}^n$ можно записать как $\mathbf{x} \in \mathbb{R}^n_+$. Если $\mathbf{x} \in \mathbb{R}^n_+$, то, очевидно, $\forall \alpha \geq 0 \ \ \alpha\mathbf{x} \in \mathbb{R}^n_+$ Тогда $\forall \alpha \geq 0$ и $\forall \mathbf{y} \in \mathcal{G}$ верно $\alpha\mathbf{y} = \alpha\mathbf{A}\mathbf{x} = \mathbf{A}(\alpha\mathbf{x}) \in \mathcal{G}$. То есть $\mathcal{G}$ – конус.

Пусть $\mathbf{p} \in \mathcal{G}$. Тогда $\forall \mathbf{x} \in \mathbb{R}^n_+$ верно $\langle \mathbf{p}, \mathbf{A}\mathbf{x} \rangle = \mathbf{p}^\top \mathbf{A}\mathbf{x} \geqslant 0$. $\mathbf{p}^\top \mathbf{A}\mathbf{x} = \langle \left(\mathbf{p}^\top \mathbf{A}\right)^\top, \mathbf{x} \rangle = \langle \mathbf{A}^\top \mathbf{p}, \mathbf{x} \rangle \geq 0$. На семинаре было показано, что $\forall \mathbf{x} \in \mathbb{R}^n_+ \ \ \langle \mathbf{A}^\top \mathbf{p}, \mathbf{x} \rangle \geq 0 \Leftrightarrow \mathbf{A}^\top \mathbf{p} \in \mathbb{R}^n_+$. Значит, $\mathcal{G}^* = \{\mathbf{p} \mid \mathbf{A}^\top \mathbf{p} \in \mathbb{R}^n_+\}$.

11. Задача одиннадцатая

$\mathbf{P}^\top \mathbf{1} = \mathbf{1}$. Это означает, что $\sum\limits_i a_{ij} = 1 \ \forall i \in \overline{1, n}$.

Покажем, что у матрицы $\mathbf{P}$ наибольшее собственное значение есть 1. Рассмотрим матрицу $\mathbf{A} = \mathbf{P} - \mathbf{E}$. $a_{ij} = p_{ij} \ (i \neq j)$ и $a_{ii} = p_{ii} - 1$. Для матрицы $\mathbf{A}$ сумма элементов в каждом столбце равна $\sum\limits_i a_{ij} = \sum\limits_i p_{ij} - 1 = 0$. Поэтому если мы сложим все строки этой матрицы, то получим строку из нулей. То есть, строки $\mathbf{A}$ ЛЗ и она вырождена, $\det(\mathbf{P} - \mathbf{E}) = 0$. А это означает, что $\lambda = 1$ – собственное значение этой матрицы.

Предположим, что у матрицы есть собственное значение $\lambda > 1$. Рассмотрим матрицу $\mathbf{A} = \mathbf{P} - \lambda\mathbf{E}$. $a_{ij} = p_{ij} \ (i \neq j)$ и $a_{ii} = p_{ii} - \lambda$. Поэтому $\forall i \in \overline{1, n} \ \ \sum\limits_j a_{ji} = 1 - \lambda < 0$. Рассмотрим нетривиальную линейную комбинацию строк матрицы $\mathbf{A}$:

$$\alpha_1 \begin{pmatrix} a_{11} \\ . \\ . \\ . \\ a_{1n} \end{pmatrix} + \ \ldots \ + \alpha_n \begin{pmatrix} a_{n1} \\ . \\ . \\ . \\ a_{nn} \end{pmatrix} = \mathbf{z}$$

Пусть α_k – наибольший из всех коэффициентов. Если он положительный, то для всех $i \neq k \ \ a_{ik} > 0$, а значит, $\alpha_k a_{ik} \geq \alpha_i a_{ik}$ тогда можно оценить k-тый элемент столбца $\mathbf{z}$: $z_k = \sum\limits_i \alpha_i a_{ik} \leq \alpha_k \sum\limits_i a_{ik} = \alpha_k(1 - \lambda) < 0$ (здесь мы для всех $i \neq k$ пользуемся неравенством $\alpha_k a_{ik} > \alpha_i a_{ik}$, а слагаемое $\alpha_k a_{kk}$ оставляем без изменений). Если же максимальный элемент неположительный, то рассмотрим, наоборот, наименьший α_m. Он должен быть отрицательным, так как линейная комбинация нетривиальная. Тогда для всех $i \neq k$ верно $\alpha_k a_{ik} \leq \alpha_i a_{ik}$ Аналогично оцениваем $z_k = \sum\limits_i \alpha_i a_{ik} \geq \alpha_k \sum\limits_i a_{ik} = \alpha_k(1 - \lambda) > 0$. То есть, $\mathbf{z} \neq \mathbf{0}$. Значит, строки матрицы $\mathbf{A} = \mathbf{P} - \lambda\mathbf{E}$ ЛНЗ при $\lambda > 1$ и у матрицы $\mathbf{P}$ нет собственных значений, больших 1. Если

Теперь докажем, что существует решение системы $\mathbf{P}\mathbf{x} = \mathbf{x}$, $\mathbf{x}^\top \mathbf{1} = 1$, $x_i \geq 0$.

Прежде всего заметим, что $\mathbf{1}^\top \mathbf{x} = \mathbf{x}^\top \mathbf{1} = 1$. Теперь можно представить систему равенств в виде:

$$\begin{cases} (\mathbf{P}-\mathbf{E})\mathbf{x} = 0 \\ \mathbf{1}^\top \mathbf{x} = 1 \end{cases} \tag{1}$$

Обозначим $\mathbf{A} \in \mathbb{R}^{(n+1)\times n}$ – матрица, состоящая из строк матрицы $\mathbf{P}-\mathbf{E}$ с приписанной снизу строкой из единиц. $\mathbf{b} \in \mathbb{R}^{(n+1)}$ – вектор-столбец, такой что $b_{p+1} = 1$, а все остальные нули. Тогда система, написанная выше, примет вид $\mathbf{Ax} = \mathbf{b}$.

По лемме Фаркаша, имеет решение одна и только одна из систем

1. $\mathbf{Ax} = \mathbf{b}$, $x_i \geq 0$
2. $\mathbf{y}^\top \mathbf{A} \geq \mathbf{0}$, $\langle \mathbf{y}, \mathbf{b} \rangle < 0$

Покажем, что вторая система не имеет решений.

$\langle \mathbf{y}, \mathbf{b} \rangle = y_{n+1} < 0$. А теперь перепишем выражение $\mathbf{y}^\top \mathbf{A}$.

$$\mathbf{y}^\top \mathbf{A} = y_1 \begin{pmatrix} a_{11} \\ . \\ . \\ . \\ a_{1n} \end{pmatrix} + \ \ldots\ + y_n \begin{pmatrix} a_{n1} \\ . \\ . \\ . \\ a_{nn} \end{pmatrix} + y_{n+1} \begin{pmatrix} 1 \\ . \\ . \\ . \\ 1 \end{pmatrix} \tag{2}$$

$y_{n+1} < 0$, поэтому должно выполняться

$$\mathbf{y}^\top \mathbf{A} = y_1 \begin{pmatrix} a_{11} \\ . \\ . \\ . \\ a_{1n} \end{pmatrix} + \ \ldots\ + y_n \begin{pmatrix} a_{n1} \\ . \\ . \\ . \\ a_{nn} \end{pmatrix} > \begin{pmatrix} 0 \\ . \\ . \\ . \\ 0 \end{pmatrix} \tag{3}$$

Пусть y_k наибольший из всех коэффициентов. Тогда для любого $i \neq k$ верно $y_i a_{ik} \leq y_k a_{ik}$. Значит, $\sum\limits_i y_i a_{ik} \leq y_k \sum\limits_i a_{ik} = y_k \cdot 0 = 0$. То есть, всегда найдётся неположительная компонента. Значит, вторая система из леммы Фаркаша не имеет решение, и имеет решение первая, ч.т.д.

11.1 Пункт 1

Последовательность случайных величин с конечным числом исходов $\{X_n\}_{n\in\mathbb{N}}$. называется Марковской цепью с конечным числом состояний, если

$\mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_{n+1} = i_{n+1} \mid X_n = i_n)$. Таким образом, условное распределение последующего состояния цепи Маркова зависит только от текущего состояния и не зависит от всех предыдущих состояний (в отличие от цепей Маркова высших порядков).

Область значений случайных величин $\{X_n\}$ называется пространством состояний цепи, а номер n – номером шага.

Матрица $\mathbf{P}(n)$, где $\mathbf{P}_{ij}(n) \equiv \mathbb{P}(X_{n+1} = j \mid X_n = i)$ называется матрицей переходных вероятностей на n-том. Вектор $\mathbf{p} = (p_1, p_2, \ldots)^\top$, где $p_i \equiv \mathbb{P}(X_0 = i)$ называется начальным распределением цепи Маркова. Матрица обладает следующими свойствами $\sum\limits_j P_{ij}(n) = 1, \quad \forall n \in \mathbb{N}$ и $\mathbf{P}(n)_{ij} \geq 0$. То есть, матрица $\mathbf{P}^\top$ из условия может являться матрицей переходных вероятностей.

11.2 Пункт 2

Пусть имеется n объектов, некоторые из них ссылаются на другие объекты (но не на себя). С помощью алгоритма PageRank можно отсортировать эти объекты по важности. Таким примером может быть сортировка сайтов в поисковой системе, учитывая, какие сайты на какие ссылаются. Или может быть граф из документов, где некоторые документы ссылаются на другие работы из этого графа.

После работы алгоритм выдаёт для каждоого объекта некоторое число, и чем оно больше, тем важнее сам объект.

Опишем работу упрощённой версии алгоритма.

Пусть $\mathbf{P} \in \mathbb{R}^{n\times n}$ – матрица, значения которой: $p_{ij} = \frac{1}{N(j)}$, если объект j ссылается на документ i. Иначе $p_{ij} = 0$. Здесь $N(j)$ – количество ссылок из вершины j. Ранг объектов рассчитывается таким образом, что $x_j = \sum\limits_{i\in S(j)} \frac{x_i}{N(i)}$, где $S(j)$ – множество объектов, ссылающихся на j. Заметим, что $x_j = \sum\limits_{i=1}^{n} p_{ji} \cdot x_i$. Поэтому вектор рангов $\mathbf{x}$ есть решение системы $\mathbf{Px} = \mathbf{x}$. Алгоритм находит приближённое решение этой системы следующим образом:

Пусть $\mathbf{x}^0$ – некоторый начальный ненулевой вектор (выбирается почти произвольно, можно выбрать, например, $\mathbf{1}$.

На i-той итерации сначала вычисляем $\mathbf{x}^{i+1} = \mathbf{Px}^i$, затем вычисляем параметр $d = ||\mathbf{x}^i||_1 - ||\mathbf{x}^{i+1}||_1$. После этого присваиваем вектору $\mathbf{x}^{i+1}$ новое значение: $\mathbf{x}^{i+1} \leftarrow \mathbf{x}^{i+1} + dE$. Вычисляем параметр $\delta = ||\mathbf{x}^i - \mathbf{x}^{i+1}||_1$. Затем сравниваем значения этого параметра с требуемой точностью ε. Если $\delta > \varepsilon$, переходим к $i+1$ итерации. Иначе выходим из цикла и выдаём вектор $\mathbf{x}$ в качестве ответа.

11.3 Пункт 3

Алгоритм называется PageRank, так как впервые он был предложен именно для ранжирования страниц в поисковой системе Google.

11.4 Пункт 4

Алгоритм ищет приближённое решение системы $\mathbf{Px} = \mathbf{x}$, причём такое, что $x_i \geq 0$. А также заметим, что сумма чисел в каждом столбце матрицы $\mathbf{P}$ равна 1, да ещё и все элементы матрицы $\mathbf{P}$ неотрицательные. Это подозрительно похоже на задачу нахождения равновесного распределения вероятностей Марковской цепи с n состояниями и матрицей переходов $\mathbf{P}$.

12. Задача двенадцатая

Система $\mathbf{Ax} = \mathbf{b}$ ($\mathbf{A} \in \mathbb{R}^{m\times n},\ \mathbf{b} \in bbR^m,\ \mathbf{x} \in \mathbb{R}^n$ имеет решение тогда и только тогда, когда решение $\mathbf{x}$ этой системы представимо в виде разности $\mathbf{u} - \mathbf{v}$, где $u_i \geq 0,\ v_i \geq 0$. В свою очередь, последнее утверждение равносильно тому, что система $\mathbf{Bz} = \mathbf{b}$, где $\mathbf{B} = [\mathbf{A} \quad -\mathbf{A}]$, а

$$\mathbf{z} = \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \end{pmatrix} \tag{4}$$

имеем решение такое что $z_i \geq 0$.

По лемме Фаркаша эта система имеет решение тогда и только тогда, когда не имеет решения система $\mathbf{B}^\top\mathbf{y} \geq \mathbf{0}$, $\langle\mathbf{b},\mathbf{y}\rangle < 0$ (здесь $\mathbf{y} \in \mathbb{R}^m$. Но система $\mathbf{B}^\top\mathbf{y} \geq \mathbf{0}$ означает $\mathbf{A}^\top\mathbf{y} \geq \mathbf{0}$, $-\mathbf{A}^\top\mathbf{y} \geq \mathbf{0}$, то есть равносильна системе $\mathbf{A}^\top\mathbf{y} = \mathbf{0}$. Для последней системы если $\mathbf{y}$ является её решением, то и $-\mathbf{y}$ – тоже решение. Поэтому условие $\langle\mathbf{b},\mathbf{y}\rangle < 0$ равносильно просто условию $\langle\mathbf{b},\mathbf{y}\rangle \neq 0$. То есть, получили систему

$$\begin{cases} \mathbf{A}^\top\mathbf{y} = \mathbf{0} \\ \langle\mathbf{b},\mathbf{y}\rangle \neq 0 \end{cases} \tag{5}$$

Итак, доказали, что имеет решение одна и только одна из систем

$$\begin{cases} \mathbf{A}^\top\mathbf{y} = \mathbf{0} \\ \langle\mathbf{b},\mathbf{y}\rangle \neq 0 \end{cases} \tag{6}$$

и

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

13. Задача тринадцатая

13.1 Пункт 1

$$J(\mathbf{U},\mathbf{V}) = ||\mathbf{U}\mathbf{V} - \mathbf{Y}||_F^2 + \frac{\lambda}{2}\left(||\mathbf{U}||_F^2 + ||\mathbf{V}||_F^2\right) = \sum_{i,j}\left(\sum_{\alpha=1}^{n} u_{i\alpha}v_{\alpha j} + y_{ij}\right)^2 + \frac{\lambda}{2}\left(\sum_{i,j} u_{ij}^2 + \sum_{i,j} v_{ij}^2\right)$$

$$\frac{\partial J}{\partial u_{ij}} = \sum_{k=1}^{n} 2v_{jk}\left(\sum_{\alpha=1}^{n} u_{i\alpha}v_{\alpha k} + y_{ik}\right) + \lambda u_{ij}$$

$$\frac{\partial J}{\partial v_{ij}} = \sum_{k=1}^{n} 2u_{ki}\left(\sum_{\alpha=1}^{n} u_{k\alpha}v_{\alpha j} + y_{kj}\right) + \lambda v_{ij}$$

$$\frac{\partial J}{\partial \mathbf{U}} = \left|\left|\frac{\partial J}{\partial u_{ij}}\right|\right|_{ij}$$

$$\frac{\partial J}{\partial \mathbf{V}} = \left|\left|\frac{\partial J}{\partial v_{ij}}\right|\right|_{ij}$$

13.2 Пункт 2

Пользуясь правилом дифференцирования сложной функции:

$$\frac{\partial(\log(1 + e^{-y_i\mathbf{w}^\top\mathbf{x}}))}{\partial w_j} = \frac{-y_i x_i^j e^{-y_i\mathbf{w}^\top\mathbf{x}_i}}{1 + e^{-y_i\mathbf{w}^\top\mathbf{x}_i}}$$

Здесь x_i^j – j-тая компонента i-того вектора.

$$\frac{\partial f}{\partial w_j} = \sum_{i=1}^{m} \frac{-y_i x_i^j e^{-y_i\mathbf{w}^\top\mathbf{x}_i}}{1 + e^{-y_i\mathbf{w}^\top\mathbf{x}_i}}$$

$$\frac{\partial(\log(1+e^{-y_i\mathbf{w}^\top\mathbf{x}_i}))}{\partial\mathbf{w}} = \left|\left|\frac{\partial f}{\partial w_j}\right|\right|_j$$

Теперь найдём компоненты Гессиана

$$\frac{\partial}{\partial w_k}\left(\frac{\partial f}{\partial w_j}\right) = \frac{\partial}{\partial w_k}\left(\sum_{i=1}^m \frac{-y_i x_i^j e^{-y_i\mathbf{w}^\top\mathbf{x}_i}}{1+e^{-y_i\mathbf{w}^\top\mathbf{x}_i}}\right) =$$

$$= \sum_{i=1}^m \frac{(1+e^{-y_i\mathbf{w}^\top\mathbf{x}_i})y_i^2 x_i^j x_i^k e^{-y_i\mathbf{w}^\top\mathbf{x}_i} - (-y_i x_i^j e^{-y_i\mathbf{w}^\top\mathbf{x}_i})(-y_i x_i^k e^{-y_i\mathbf{w}^\top\mathbf{x}_i})}{(1+e^{-y_i\mathbf{w}^\top\mathbf{x}_i})^2} = \sum_{i=1}^m \frac{y_i^2 x_i^j x_i^k e^{-y_i\mathbf{w}^\top\mathbf{x}_i}}{(1+e^{-y_i\mathbf{w}^\top\mathbf{x}_i})^2}$$

Получаем гессиан

$$\mathbf{H} = \left|\left|\frac{\partial^2 f}{\partial w_k \partial w_j}\right|\right|_{ij}$$

13.3 Пункт 3

Если $i \neq j$, то

$$\frac{\partial f_i}{\partial w_j} = -\frac{e^{w_i}e^{w_j}}{\left(\sum e^{w_k}\right)^2}$$

Если $i = j$

$$\frac{\partial f_i}{\partial w_j} = \frac{-e^{w_i}e^{w_j} + e^{w_i}\left(\sum e^{w_k}\right)}{\left(\sum e^{w_k}\right)^2}$$

В общем случае

$$\frac{\partial f_i}{\partial w_j} = \frac{-e^{w_i}e^{w_j} + \delta_{ij}e^{w_i}\left(\sum e^{w_k}\right)}{\left(\sum e^{w_k}\right)^2}$$

Обозначим $\mathbf{z} = (e^{w_1}\dots e^{w_n})^\top$. Тогда $\left(\sum e^{w_k}\right)^2 = (\mathbf{1}^\top\mathbf{z})^2$ и

$$\frac{\partial\mathbf{f}}{\partial\mathbf{w}} = \frac{1}{(\mathbf{1}^\top\mathbf{z})^2}\left(-\mathbf{z}\mathbf{z}^\top + \operatorname{diag}(z_1,\dots,z_n)(\mathbf{1}^\top\mathbf{z})^2\right)$$

13.4 Пункт 4

a)

$$f(\mathbf{X}) = \sum_{i=1}^n \lambda_i(\mathbf{X}) = \text{trace}X = \sum_{i=1}^n a_{ii}$$

$$\frac{\partial f}{\partial\mathbf{X}} = \mathbf{I}$$

$\mathbf{I}$ – единичная матрица.

b)

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

На семинаре было получено выражение для градиента детерминанта:

$$\frac{\partial f}{\partial \mathbf{X}} = \mathbf{X}^{-\top} \det \mathbf{X}$$

Здесь $\mathbf{X}^{-\top} = (\mathbf{X}^{-1})^{\top}$.

14. Задача четырнадцатая

14.1 Пункт 1

14.2 Пункт 2

Доказательство в следующем пункте.

14.3 Пункт 3

Так как при всех $x \geq -1$ верно $y \geq \log(y+1)$ и $p(x) \geq 0$

$$-D_{KL}(p \,||\, q) = -\int p(x) \log \frac{p(x)}{q(x)} dx = \int p(x) \log \frac{q(x)}{p(x)} dx \leq \int p(x) \left(\frac{q(x)}{p(x)} - 1 \right) dx = \int (q(x) - p(x)) dx = 0$$

Из свойств функций плотности вероятности следует $\int p(x)dx = \int q(x)dx = 1$. Отсюда и получаем последнее равенство.

А это означает, что $D_{KL}(p \,||\, q) \geq 0$.

Из неравенства $\int p(x) \log \frac{p(x)}{q(x)} dx \geq 0$ и из того, что $p(x) \geq 0$, $\log \frac{p(x)}{q(x)} \geq 0$ следует, что равенство достигается только если $p(x) \log \frac{p(x)}{q(x)} \equiv 0$ почти всюду. Интеграл мы рассматриваем только на таких множествах, на которых $p(x) \neq 0$ почти всюду, чтобы значение логарифма было определено. Поэтому $\log \frac{p(x)}{q(x)} \equiv 0$ почти всюду. А это означает, что $p(x) \equiv q(x)$ почти всюду.

В обратную сторону понятно, что если $p(x) \equiv q(x)$, то $\log \frac{p(x)}{q(x)} \equiv 0$ и $D_{KL}(p \,||\, q) = \int p(x) \log \frac{p(x)}{q(x)} dx = 0$.

Итак, мы доказали, что $p(x) \equiv q(x) \;\Leftrightarrow\; D_{KL}(p \,||\, q) = 0$.

14.4 Пункт 4

Рассмотрим 2 распределения:

$$p(x) = \begin{cases} 3 & x \in \left[0, \frac{1}{4}\right] \\ 1 & x \in \left(\frac{1}{4}, \frac{1}{2}\right] \end{cases} \tag{7}$$

$$q(x) = 2,\; x \in \left[0, \frac{1}{2}\right]$$

Посчитаем дивергенцию Кульбака-Лейбнера:

$$D_{KL}(p \mid\mid q) = \int p(x) \log \frac{p(x)}{q(x)} dx = \frac{1}{4}\left(3 \log \frac{3}{2} + \log \frac{1}{2}\right) = \frac{1}{4} \log \frac{3^3}{2^4}$$

$$D_{KL}(q \mid\mid p) = \int q(x) \log \frac{q(x)}{p(x)} dx = \frac{1}{4} \cdot 2 \left(\log \frac{2}{3} + \log 2\right) = \frac{1}{4} \log \frac{2^4}{3^2}$$

Видим, что дивергенция Кульбака-Лейбнера несимметрична.

15. Задача пятнадцатая

Пусть f – выпуклая функция.

Сначала заметим, что в силу того, что λ_1 – единственное положительное число, а сумма всех коэффициентов равна 1, то $\lambda_1 \geq 1$, а также $|\lambda_1| > |\lambda_i|$ для любого $i = 2, \dots, n$.

Рассмотрим набор коэффициентов

$alpha_1 = \frac{1}{\lambda_1}, \alpha_2 - \frac{\lambda_2}{\lambda_1}, \alpha_3 - \frac{\lambda_3}{\lambda_1}, \dots, \alpha_n - \frac{\lambda_n}{\lambda_1}$. Их сумма равна $\frac{1}{\lambda_1} + \sum\limits_{i=2}^{n} \frac{-\lambda_i}{\lambda_1} = \frac{1}{\lambda_1} + \frac{\lambda_1 - 1}{\lambda_1} = 1$. При этом из условия и из сказанного в предыдущем абзаце следует, что все $0 \leq \alpha_i \leq 1$.

Пусть $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathrm{dom} f$ Рассмотрим точки $\mathbf{y}_1 = \sum\limits_{i=1}^{n} \lambda_i \mathbf{x}_i,\ \mathbf{y}_i = \mathbf{x}_i,\ i = 2, \dots, n$. Если $\mathbf{y}_i \in \mathrm{dom} f$, то выпуклость функции f даёт нам неравенство

$$\sum_{i=1}^{n} \alpha_i f(\mathbf{y}_i) \geq f\left(\sum_{i=1}^{n} \alpha_i \mathbf{y}_i\right)$$

А теперь подставим сюда $\mathbf{x}_i$ и λ_i:

$$\sum_{i=1}^{n} \alpha_i \mathbf{y}_i = \frac{1}{\lambda_1}(\lambda_1 \mathbf{x}_1 + \dots + \lambda_n \mathbf{x}_n) + \sum_{i=2}^{n} \frac{-\lambda_i}{\lambda_1} \mathbf{x}_i = \frac{\lambda_1 \mathbf{x}_1 + \dots + \lambda_n \mathbf{x}_n + \sum_{i=2}^{n} -\lambda_i \mathbf{x}_i}{\lambda_1} = \mathbf{x}_1$$

$$\sum_{i=1}^{n} \alpha_i f(\mathbf{y}_i) = \frac{f(\lambda_1 \mathbf{x}_1 + \dots + \lambda_n \mathbf{x}_n) + (-\lambda_2) f(\mathbf{x}_2) + \dots + (-\lambda_n) f(bx_n)}{\lambda_1}$$

Неравенство примет вид

$$\frac{f(\lambda_1 \mathbf{x}_1 + \dots + \lambda_n \mathbf{x}_n) - \lambda_2 f(\mathbf{x}_2) - \dots - \lambda_n f(bx_n)}{\lambda_1} \geq f(\mathbf{x}_1)$$

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

16. Задача шестнадцатая

Докажем, что функция вогнута.

Пусть при наборе весов $\mathbf{c}_1$ минимум достигается на пути P, при наборе $\mathbf{c}_2$ – на пути Q, при $\alpha \mathbf{c}_1 + (1 - \alpha)\mathbf{c}_2$ – на пути R. При этом пусть веса этих путей на наборе $\mathbf{c}_1$ равны $p_1,\ q_1$ и r_1 соответственно, а при наборе $\mathbf{c}_2$ – $p_2,\ q_2$ и r_2. То есть, $p_{ij}(\mathbf{c}_1) = p_1,\ p_{ij}(\mathbf{c}_2) = q_2$, а

$p_{ij}(\alpha\mathbf{c}_1+(1-\alpha)\mathbf{c}_2) = \alpha r_1+(1-\alpha)r_2$. Так как путь P минимальный на первом наборе, то $r_1 \geq p_1$, и аналогично $r_2 \geq q_2$. Следовательно, $\alpha r_1+(1-\alpha)r_2 \geq \alpha p_1+(1-\alpha)q_2 = \alpha p_{ij}(\mathbf{c}_1)+(1-\alpha)p_{ij}(\mathbf{c}_2)$. Итак, $p_{ij}(\alpha\mathbf{c}_1 + (1-\alpha)\mathbf{c}_2) \geq \alpha p_{ij}(\mathbf{c}_1) + (1-\alpha)p_{ij}(\mathbf{c}_2)$, и функция p_{ij} вогнута.

17. Задача семнадцатая

18. Задача восемнадцатая

18.1 Пункт 1

Обозначим i-тую строчка матрицы $\mathbf{A}$ за вектор $\mathbf{a}_i$.

$$||\mathbf{A}\mathbf{x}+\mathbf{b}||_1 = \sum_i \Big|\langle \mathbf{a}_i, \mathbf{x}\rangle + b_i\Big| =$$

$$= \sum_i \max\left(\langle \mathbf{a}_i, \mathbf{x}\rangle + b_i, -\langle \mathbf{a}_i, \mathbf{x}\rangle + b_i\right)$$

По формуле для субдифференциала максимума

$$\partial \max\left(\langle \mathbf{a}_i, \mathbf{x}\rangle + b_i, -\langle \mathbf{a}_i, \mathbf{x}\rangle + b_i\right) = \begin{cases} \mathbf{a}_i & \langle \mathbf{a}_i, \mathbf{x}\rangle + b_i > 0 \\ \operatorname{conv}\left(\mathbf{a}_i, -\mathbf{a}_i\right) & \langle \mathbf{a}_i, \mathbf{x}\rangle + b_i > 0 \\ -\mathbf{a}_i & \langle \mathbf{a}_i, \mathbf{x}\rangle + b_i < 0 \end{cases} \tag{8}$$

По формуле для субдифференциала суммы

$$\partial||\mathbf{A}\mathbf{x}+\mathbf{b}||_1 = \sum_i \partial \max\left(\langle \mathbf{a}_i, \mathbf{x}\rangle + b_i, -\langle \mathbf{a}_i, \mathbf{x}\rangle + b_i\right) = \sum_i \left(\begin{cases} \mathbf{a}_i & \langle \mathbf{a}_i, \mathbf{x}\rangle + b_i > 0 \\ \operatorname{conv}\left(\mathbf{a}_i, -\mathbf{a}_i\right) & \langle \mathbf{a}_i, \mathbf{x}\rangle + b_i > 0 \\ -\mathbf{a}_i & \langle \mathbf{a}_i, \mathbf{x}\rangle + b_i < 0 \end{cases}\right) \tag{9}$$

18.2 Пункт 2

$$\partial\left(\frac{1}{2}||\mathbf{w}||_2^2\right) = \mathbf{w}$$

$$\partial \max(0,\ 1 - y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b_i)) = \begin{cases} \mathbf{0} & 1 < y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b_i) \\ \operatorname{conv}\left(\mathbf{0}, -y_i\mathbf{w}\right) & 1 = y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b_i) \\ -y_i\mathbf{w} & 1 > y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b_i) \end{cases} \tag{10}$$

По формуле субдифференциала суммы получаем

$$\partial L(\mathbf{w}) = \mathbf{w} + \sum_i \left(\begin{cases} \mathbf{0} & 1 < y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b_i) \\ \operatorname{conv}\left(\mathbf{0}, -y_i\mathbf{w}\right) & 1 = y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b_i) \\ -y_i\mathbf{w} & 1 > y_i(\langle \mathbf{w}, \mathbf{x}_i\rangle + b_i) \end{cases}\right) \tag{11}$$

18.3 Пункт 3

В точках из интервалов $(-\infty, -2)$ и $(2, +\infty)$ субдифференциал не определён, так как по определению в одномерном случае это такое число a, что $f(y) - f(x) \geq a(y - x)$. При этом если $x \in (-\infty, -2) \cup (2, +\infty)$, то $f(x) = \infty$, а вычитать бесконечность мы не можем.

Для остальных точек несложно найти:

$$\partial f(x) = \begin{cases} (-\infty, -1] & x = -2 \\ -1 & x \in (-2, -1) \\ [-1, 0] & x = -1 \\ 0 & x \in (-1, 1) \\ [0, 1] & x = 1 \\ 1 & x \in (1, 2) \\ [1, +\infty) & x = 2 \end{cases} \tag{12}$$

18.4 Пункт 4

$$\partial f_X(\mathbf{x}_0) = \partial f(\mathbf{x}_0) + N(\mathbf{x}_0 \mid X)$$

Множество $\{\mathbf{x} \mid ||\mathbf{x}||_2^2 = 2\}$ есть окружность. Поэтому нормальный конус к этому множеству есть

$$N(\mathbf{x}_0 \mid X) = \begin{cases} \mathbf{0} & ||\mathbf{x}_0||_2^2 < 2 \\ \alpha \mathbf{x}_0 & \alpha \geq 0,\ ||\mathbf{x}_0||_2^2 < 2 \end{cases} \tag{13}$$

По формуле субдифференциала для модулей:

$$\partial f(\mathbf{x}_0) = \begin{cases} (1, -1) & x_1 - x_2 > 0 \\ \operatorname{conv}((1, -1), (-1, 1)) & x_1 - x_2 = 0 \\ (-1, 1) & x_1 - x_2 < 0 \end{cases} + \begin{cases} (1, 1) & x_1 + x_2 > 0 \\ \operatorname{conv}((1, 1), (-1, -1)) & x_1 + x_2 = 0 \\ (-1, -1) & x_1 + x_2 < 0 \end{cases} \tag{14}$$

В итоге

$$\partial f_X(\mathbf{x}_0) = \begin{cases} \mathbf{0} & ||\mathbf{x}_0||_2^2 < 2 \\ \alpha \mathbf{x}_0 & \alpha \geq 0,\ ||\mathbf{x}_0||_2^2 < 2 \end{cases} + \begin{cases} (1, -1) & x_1 - x_2 > 0 \\ \operatorname{conv}((1, -1), (-1, 1)) & x_1 - x_2 = 0 + \\ (-1, 1) & x_1 - x_2 < 0 \end{cases} \tag{15}$$

$$+ \begin{cases} (1, 1) & x_1 + x_2 > 0 \\ \operatorname{conv}((1, 1), (-1, -1)) & x_1 + x_2 = 0 \\ (-1, -1) & x_1 + x_2 < 0 \end{cases} \tag{16}$$