Homework 1

Lygina Oksana
878

23 февраля 2021 г.

Matrix calculus

1. **solution**:
f(x) = $||Ax||_2 = \sqrt{< Ax, Ax >}$
df = $\frac{d<Ax,Ax>}{||Ax||_2}$ = $\frac{<Adx,Ax>+<Ax,Adx>}{2||Ax||_2}$ = $\frac{<Ax,Adx>}{||Ax||_2}$ = <$\frac{A^T Ax}{||Ax||_2}$,dx> → ∇ f(x) = <$\frac{A^T Ax}{||Ax||_2}$ >
Ответ: ∇ f(x) = <$\frac{A^T Ax}{||Ax||_2}$ >

2. **solution**:
С семинара: $g(X) = detX \to g'(X) = detXX^{-T}$
Тогда df(X) = d $\log detX$ = $\frac{<detXX^{-T},dX>}{detX}$ $=< X^{-T}, dX > \to f'(X) = X^{-T}$
h(X) = X^{-T}. Тогда найдем h'(X):
Дифференцируем тождество: $X^T X^{-T} = I \to dX^T X^{-T} + X^T d(X^{-T}) = 0$
dh(X) = $-X^{-T} dX^T X^{-T}$
Обозначим через E_{ij} матрицу, в которой везде нули, кроме элемента (i, j) - там стоит единица.
Тогда: $dX = E_{ij} dX_{ij}$
$dh(X) = -X^{-T} E_{ij}^T X^{-T} dX_{ij} \to \frac{\partial h(X)}{\partial x_{ij}} = -X^{-T} E_{ji} X^{-T}$
Ответ: $f''(X) = -X^{-T} E_{ji} X^{-T}$

3. **solution**:
$f(X) = ||X||_F^2 = tr(X^T X) =< X, X >$
$df(X) = d < X, X >= 2 < X, dX >=< 2X, dX >\to \frac{\partial}{\partial X}||X||_F^2 = 2X$
Ответ: 2X

4. **solution**:
f(x) = $||y{-}$y$_o||_2^2$ = <(y - y_o), (y - y_o)> = <(y - Wx - b), (y - Wx - b)>, где y_o = Wx + b
$\frac{\partial L}{\partial W}$ = 2<y - Wx - b, $\frac{\partial}{\partial W}$(y - Wx - b)> = 2<y - Wx - b, -x dW> = (-2x$^T(y - Wx - b), dW >\to \frac{\partial L}{\partial W}$ = -2x$^T(y - Wx - b)$

$\frac{\partial L}{\partial b}$ = 2<y - Wx - b, $\frac{\partial}{\partial b}$(y - Wx - b)> = 2<y - Wx - b, -db> = (-2(y - Wx - b), db>
$\rightarrow \frac{\partial L}{\partial b}$ = -2(y - Wx - b)

Convex sets

1. **solution**:
Пусть множество S - выпукло и $x_o, y_o \in int(S) \rightarrow x_o, y_o \in S$
Так как точки x_o, y_o - внутренние $\rightarrow \exists\epsilon > 0$: $\forall x$: ||x - x_o|| $\leq \epsilon \rightarrow$ x $\in S$
И $\forall y: ||y - y_o|| \leq \epsilon \rightarrow y \in S$
Домножим первое неравенство на α, а второе на (1 - α), далее сложим и используем неравенство треугольника, тогда получим:
||z - z_o|| = $|\alpha x$ + (1 - α)y - (αx_o + (1 - α)y_o)|| $\leq \alpha$||x - x_o|| + (1 - α)||y - y_o|| $\leq \epsilon$
$z =\alpha$ x + (1 - α)y, z_o= αx_o + (1 - α)y_o, $0 \leq \alpha \leq 1$
Так как x, x_o, y, $y_o \in$ S и S - выпукло $\rightarrow$ z, $z_o \in$ S
Это значит, что z_o содержится в S вместе со своей окрестностью и значит in S $\rightarrow \in int(S) \rightarrow$ int(S) - выпукло.
Обратно неверно - приведем контрпример.
Возьмем квадрат ABCD на плоскости, границы AB и CD входят в множество, включая точки A, B, C, D, но при этом границы BC и AD не входят.
Это множество невыпукло, так как, например, отрезок между D и A не принадлежит этому множеству. Но множество внутренних точек квадрата выпукло. ✓

2. **solution**:
Возьмем две квадратные симметричные положительные определенные матрицы A и B, то есть A, B $\succ 0$
Это означает, что $\forall x \in R^n$ (x $\neq$ 0) выполнено: $x^T A x > 0$, $x^T B x > 0$
Пусть $\alpha \in [0\ ,\ 1]$. Тогда C = α A + (1 - α)B
Домножим слева на x^T и справа на x:
$x^T C x = x^T(\alpha$A + (1 - α)B)x = $\alpha x^T A x$ + (1 - α)$x^T B x > 0 \rightarrow$ C $\succ 0$ и $\mathcal{S}^n_{++}$ выпукло.✓

3. **solution**:
Пусть x, y $\in$ S; $0 \leq \alpha \leq 1$
Проверим принадлежность S произвольной выпуклой комбинации x и y ($\prod\limits_{i=1}^{n} x_i \geq 1$, $\prod\limits_{i=1}^{n} y_i \geq 1$):
$\prod\limits_{i=1}^{n} z_i = \prod\limits_{i=1}^{n}$ (αx_i + (1 - α)y_i) $\geq \prod\limits_{i=1}^{n} x_i^{\alpha} y_i^{1-\alpha} = (\prod\limits_{i=1}^{n} x_i)^{\alpha}(\prod\limits_{i=1}^{n} y_i)^{1-\alpha} \geq 1^{\alpha}1^{1-\alpha} = 1 \rightarrow$ z $\in$ S
$\rightarrow$ S - выпукло. ✓

4. **solution**:
$\Rightarrow$ Пусть выполнено ($\alpha + \beta$)S = αS + βS

Тогда $\forall x_1, x_2 \in S$ $\forall \alpha, \beta \geq 0 \rightarrow \exists x \in S : \alpha x_1 + \beta x_2 = (\alpha + \beta)$x
Возьмем $\alpha =$ a, $\beta =$ 1 - a, a $\in$ [0, 1] $\rightarrow \forall x_1, x_2 \in S$ $\forall$a $\in$ [0, 1] $\rightarrow$ a$x_1 + (1-a)x_2 = x \in$ S $\rightarrow$ S - выпукло
$\Leftarrow$ Пусть S - выпукло, α, $\beta \geq 0$. Считаем $\alpha + \beta > 0$, так как случай $\alpha = \beta = 0$ тривиален.
1) Пусть x $\in (\alpha + \beta)$S. Тогда $\exists$y $\in$ S: x $= (\alpha + \beta)$y. Тогда αy $\in \alpha$S, βy $\in \beta$S и выполнено: x $= \alpha$y $+ \beta$y $\rightarrow$ x $\in \alpha$S $+ \beta$S
2) Пусть x $\in \alpha$S $+ \beta$S. Тогда $\exists$y$_1, y_2 \in$S: x $= \alpha y_1 + \beta y_2$
S выпукло, значит, $\frac{x}{\alpha+\beta} = \frac{\alpha}{\alpha+\beta} y_1 + \frac{\beta}{\alpha+\beta} y_2 \in$ S $\rightarrow$ x $\in (\alpha + \beta)$S. ✓

5. **solution**:
Пункт a:
1) Если $\alpha > a_i$, то набор будет пустым и будет выпуклым.
2) Если $\alpha \leq a_1$, то набор векторов совпадает с вероятностным пространством векторов и тоже выпуклый.
3) Если $a_{k-1} \leq \alpha \leq a_k$, то: $P(x > \alpha) = \sum\limits_{i=k}^{n} a_i p_i \geq \beta$
Пусть p, q $\in$ S: $\forall \lambda \in$ [0, 1]
Тогда для их выпуклой комбинации выполнено: $\sum\limits_{i=k}^{n} a_i(\lambda p_i + (1 - \lambda) q_i) = \lambda \sum\limits_{i=k}^{n} a_i p_i + (1 - \lambda) \sum\limits_{i=k}^{n} a_i q_i \geq \lambda\beta + (1 - \lambda)\beta = \beta \rightarrow$ S - выпукло. ✓
Пункт b:
Вектор p принадлежит набору $\Leftrightarrow E$ $|$x$^{201}| = \sum\limits_{i=1}^{n} p_i |a_i^{201}| \leq \alpha \sum\limits_{i=1}^{n} p_i |a_i|$
Рассмотрим p $= \lambda$c $+ (1 - \lambda)$b, где c, b принадлежат данному набору, $\lambda \in$ [0, 1]
$\sum\limits_{i=1}^{n} p_i |\alpha_i^{201}| = \lambda \sum\limits_{i=1}^{n} c_i$ $|$a$_i^{201}| + (1-\lambda) \sum\limits_{i=1}^{n} b_i$ $|$a$_i^{201}| \leq \lambda\alpha \sum\limits_{i=1}^{n} c_i |a_i| + (1-\lambda)$ $\alpha \sum\limits_{i=1}^{n} b_i |a_i| = \alpha \sum\limits_{i=1}^{n}$ $(\lambda c_i + (1 - \lambda)$ b$_i |$a$_i| = \alpha \sum\limits_{i=1}^{n} p_i |a_i| \rightarrow$ вектор p принадлежит данному набору $\rightarrow$ данный набор выпуклый. ✓
Пункт c:
Аналогично пунктам a и б:
Вектор p принадлежит набору $\Leftrightarrow E$ $|$x$^2| = \sum\limits_{i=1}^{n} p_i$ $|$a$_i^2| \geq \alpha$
Рассмотрим p $= \lambda$c $+ (1 - \lambda)$b, где c, b принадлежат данному набору, $\lambda \in$ [0, 1]
$\sum\limits_{i=1}^{n} p_i |\alpha_i^2| = \lambda \sum\limits_{i=1}^{n} p_i |c_i^2| + (1-\lambda) \sum\limits_{i=1}^{n} b_i |a_i^2| \geq \lambda\alpha + (1 - \lambda)$ $\alpha \geq \alpha \rightarrow$ вектор p принадлежит данному набору $\rightarrow$ данный набор выпуклый. ✓

Пункт d:
Распишем дисперсию: $\bigvee$x $= E$x - $(E$x$)^2 = \sum\limits_{i=1}^{n} a_i^2 p_i$ - $(\sum\limits_{i=1}^{n} a_i p_i)^2$
Пусть p, q $\in$ S, $\lambda \in$ [0,1]. Тогда их выпуклой комбинации будет: z $= \lambda$p $+ (1 - \lambda)$q

$\sum\limits_{i=1}^{n} a_i^2 p \geq (\sum\limits_{i=1}^{n} a_i^2 p)^2 + \alpha$
$\sum\limits_{i=1}^{n} a_i^2 q \geq (\sum\limits_{i=1}^{n} a_i^2 q)^2 + \alpha$
$\bigvee(z) = \lambda \sum\limits_{i=1}^{n} a_i^2 p + (1 - \lambda)\sum\limits_{i=1}^{n} a_i^2 q - \lambda^2(\sum\limits_{i=1}^{n} a_i p)^2 - (1 - \lambda)^2(\sum\limits_{i=1}^{n} a_i q)^2 - 2\lambda(\lambda - 1)\sum\limits_{i=1}^{n} a_i p \sum\limits_{i=1}^{n} a_i q \geq$
$\lambda\alpha + \lambda(1 - \lambda)((\sum\limits_{i=1}^{n} a_i q)^2 + (\sum\limits_{i=1}^{n} a_i p)^2 - 2\sum\limits_{i=1}^{n} a_i \mathrm{p} \sum\limits_{i=1}^{n} a_i \mathrm{q}) \geq \alpha + (\sum\limits_{i=1}^{n} a_i \mathrm{p} + \sum\limits_{i=1}^{n} a_i \mathrm{q})^2 \geq \alpha \rightarrow$
z∈S → S - выпуклое множество. ✓

Projection

1. **solution**:
Норма Фробениуса:
Пусть SVD матрицы: X = UDV,где D сингулярные значения стоят на диагонали порядке убывания: $\sigma_1 \geq \sigma_2 \geq ...$ Покажем, что $X_k = \sum\limits_{i=1}^{k} \sigma_i u_i v_i^T$, где u_i, v_i i-ые столбцы V, U есть искомая проекция.
$||\mathrm{X} - X_k||_F^2 = ||\sum\limits_{i=1}^{k+1} \sigma_i u_i v_i^T||_F^2 = \sum\limits_{i=1}^{k+1} \sigma_i^2$
Покажем, что для любой матрицы Y ранга k выполняется: $||\mathrm{X} - \mathrm{Y}||_F^2 \geq ||\mathrm{X} - \mathrm{X}_k||_F^2$
Учтем неравенство сингулярных разложений: $\sigma_{i+j-1}(\mathrm{A} + \mathrm{B}) \leq \sigma_i(A) + \sigma_j(B)\ \forall \mathrm{j} \leq$ n, i + j - 1 ≤ n
Так как у Y ранг k,то $\sigma_{k+1}(Y) = 0. j = k+1, B = Y, A = X - Y$:$\sigma_{i+k}(\mathrm{X}) \leq \sigma_i^2(X - Y)$ $\forall i$: 1 ≤ n - k → $||\mathrm{X} - \mathrm{Y}||_F^2 \geq \sum\limits_{i=1}^{n-k} \sigma_i^2(X - Y) \geq \sum\limits_{i=k+1}^{n} \sigma_i^2(X) = ||\mathrm{X} - \mathrm{X}_k||_F^2$
Значит, X_k.✓
Спектральная норма:
Покажем, что для любой матрицы Y ранга k выполняется: $||X - Y||_2^2 \geq ||X - X_k||_2^2 = ||\sum\limits_{i=k+1}^{n} \sigma_i u_i v_i^T||_2^2 = \sigma_{k+1}^2$
Из неравенства сингулярных значений: $\sigma_{k+1}(X) \leq \sigma_1(X - Y) \rightarrow ||X - Y||_2^2 \geq \sigma_1^2(X - Y) \geq \sigma_{k+1}(X) = ||X - X_k||_2^2$
А значит, $X_k = \sum\limits_{i=1}^{k} \sigma_i u_i v_i^T$ искомая проекция по спектральной норме. ✓

2. **solution**:
Используем теорему: Пусть S - выпуклое замкнутое множество в R^n. Тогда проекция $\pi_S(a)$ любой точки a ∈ R^n на S существует и единственна. Тогда $<\pi_S(x_1) - \mathrm{x}_1, y - \pi_S(x_1)> \geq 0$
$<\pi_S(x_2) - \mathrm{x}_2, y - \pi_S(x_2)> \geq 0$
Так как у любой, то возьмем вместо него проекции противоложных точек, то есть $<\pi_S(x_1) - \mathrm{x}_1, \pi_S(x_2) - \pi_S(x_1)> \geq 0$

<$\pi_S(x_2)$ - x_2,$\pi_S(x_1)$ - $\pi_S(x_2)$> ≥ 0
Из этих неравенств получаем требуемое неравенство $||\pi_S(x_2)$ - $\pi_S(x_1)||_2 \leq ||x_2 - x_1||_2$✓

Matrix calculus

1. **solution**:
   a) Воспользуемся способом ограничения на прямую. Пусть X $\in S^n_{++}$, Y $\in R^{nxn}$. Покажем, что функция h(t) = f(X + tY) = $tr(X+tY)^{-1}$ выпукла на множестве T = t| X + tY $\in S^n_{++}$. Чтобы T не равнялось пустому множеству, необходимо, чтобы матрица Y была симметричной.
   h(t) = $tr(X + tY)^{-1}$ = $tr(I$+t$X^{-1})^{-1}X^{-1}$ = trX^{-1}(I + tP$\bigwedge P^T)^{-1}$, где $X^{-1}Y$ = P$\bigwedge P^T$ - диагонализация симметричной матрицы X^{-1}Y. Еще здесь используется, что след произведения симметричных матриц не зависит от порядка их перемножения.
   Далее учтем, что преобразования не меняют след.
   h(t) = tr(X^{-1}P$(I$+$t\bigwedge)^{-1}P^T$) = tr(p^TX^{-1}P$(I$+$t\bigwedge)^{-1}$) = $\sum\limits_{k=1}^{n}$(P^TX^{-1}P$)_{kk}$ $\frac{1}{1+t}\lambda_k$=$\sum\limits_{k=1}^{n} \frac{c_k}{1+\lambda_k t}$
   Используем тот факт,что если A, B $\succeq 0$ и AB - симметричная матрица, то AB $\succeq 0$
   Тогда у нас A = $X^{-1} \succeq 0$, B = X + tY $\succeq 0$. Значит, AB = I + tX^{-1}Y $\succeq 0$, то есть выполнено 1 + λ_kt $\geqslant 0$ $\forall$k $\forall$t $\in$T $\rightarrow$ функции $\frac{c_k}{1+}\lambda_k$t выпуклы на T
   Тогда и h(t) выпукла как сумма выпуклых функций. ✓
   b) Воспользуемся способом ограничения на прямую. Пусть X $\in S^n_{++}$, Y $\in R^{nxn}$. Покажем, что функция h(t) = -g(X + tY) = -(det(X + tY)$)^{\frac{1}{n}}$.,Y.
   $h(t) = -(detX)^{\frac{1}{n}}(det(I + t$X^{-1}Y$))^{\frac{1}{n}} = -C\prod\limits_{k=1}^{n}$(1 + λ_kt$))^{\frac{1}{n}}$
   $a)$1+λ_kt > 0.
   Есть такой факт, что геометрическое среднее F(x) = $\prod\limits_{k=1}^{n} {x_k}^{\frac{1}{n}}$$R^n_{++}$
   Тогда функция h(t) является выпуклой. ✓

2. **solution**:
   f(p) = $\sum p_i log(p_i) \rightarrow f'_{p_i}$ = log(p_i) + 1$\rightarrow$ $f''_{p_ip_i} = \frac{1}{p_i} \rightarrow \nabla^2$f = diag(1$p_i$) $\succ 0$
   По второму дифференциальному критерию f выпукла
   Тогда первый дифференциальный критерий: f(p) = $\geq$ f(q) + $\nabla f^T(q)$(p - q) $\rightarrow$ D(p, q) ≥ 0 $\forall p, r \in R^n_{++}$ и D(p, q) = 0 $\leftrightarrow p = q$.✓

3. **solution**:
   1) Математическое ожидание линейная функция E(αx_1 + (1 - $\alpha$$x_2$) =$\alpha$E x_1 + (1$-\alpha$)E $x_2$$\rightarrow$ это одновременно и выпуклая и вогнутая функция. ✓
   2) Если $\alpha > a_n$, то P(p) = 0 $\forall$p выпуклая функция. Аналогично при $\alpha < a_1$. Пусть a_i $< \alpha < a_i$ + 1, тогда P(p) = $\sum\limits_{j=i+1}^{n} p_j$ - линейная функция, значит, является выпуклой и вогнутой. ✓

3) Аналогично предыдущего пункту: $a_i < \alpha \leq a_i + 1$, $a_j \leq \beta < a_j + 1$

P(p) = $\sum\limits_{k=i+1}^{j} p_k$ - линейная функция, а значит является выпуклой и вогнутой. ✓

4) Из первой части задачи 2 по второму критериб следует, что строго выпуклая функция. ✓

5) v(x) = E(x^2)$-$E(x)^2

0 <λ < 1

V(λp_1 + (1 - λp_2)) = $\sum\limits_{i=1}^{n}$a$_i^2$(λp_{1i} + (1 - λ)p_{2i})) - ($\sum\limits_{i=1}^{n}$a$_i$(λp_{1i} + (1 - λp_{2i}))^2

λV(p$_1$) + (1$-\lambda$)V(p$_2$) =$\lambda \sum\limits_{i=1}^{n} a_i^2 p_{1i}$ - $\lambda(\sum\limits_{i=1}^{n} a_i p_{1i})^2$ + (1 - λ)$\sum\limits_{i=1}^{n} a_{i2} p_{2i}$ - (1 - λ)($\sum\limits_{i=1}^{n} a_i p_{2i})^2$

λV(p$_1$) + (1$-\lambda$)V(p$_2$) $- V(\lambda p_1$ + (1 - λ)p_2) = ($\sum\limits_{i=1}^{n} a_i(\lambda p_{1i}$ + (1 - λ)p_{2i}))$^2-\lambda(\sum\limits_{i=1}^{n} a_i p_{1i})^2$ $-$ (1$-\lambda$)($\sum\limits_{i=1}^{n} a_i p_{2i})^2$ = $\lambda(\lambda$ - 1)(($\sum\limits_{i=1}^{n} a_i p_{1i})^2$+($\sum\limits_{i=1}^{n} a_i p_{2i})^2-2\sum\limits_{i=1}^{n} a_i p_{1i} \sum\limits_{i=1}^{n} a_i p_{2i}$) = ($\sum\limits_{i=1}^{n} a_i p_{1i}$ - $\sum\limits_{i=1}^{n} a_i p_{2i})^2 \leq$ 0

Значит, функция вогнутая. ✓

6) Заметим, что в случае дискретной случайной величины квартилем является:

f(p) = quartile(x) = min$\{a_k | \sum\limits_{i=1}^{k} p_i \geq \frac{1}{4}\}$

Квартиль, в случае дискретной случайной величины, являетя ступенчатой функцией. Ступенчатая функция не может быть ни выпуклой, ни вогнутой, так как ее надграфик и подграфик не являются выпуклыми множествами.

Теперь докажем это:

Покажем, что f не является выпуклой:

p = ($\frac{1}{4}$, $\frac{3}{4}$), q = (0, 1), a = (0, 1000), λ = $\frac{1}{2}$

Тогда при z = λp + (1 - λ)q = ($\frac{1}{8}$, $\frac{7}{8}$):

f(z) = 1000, λf(p) + (1 - λ)f(q) = 500 $\rightarrow$ нарушается определение выпуклости

Покажем, что f не является вогнутой:

p = (1, 0), q = (0, 1), a = (0, 1000), λ = $\frac{1}{2}$

Тогда при z = λp + (1 - λ)q = ($\frac{1}{2}$, $\frac{1}{2}$):

f(z) = 0, λf(p) + (1 - λ)f(q) = 500 $\rightarrow$ нарушается определение вогнутости. ✓

4. **solution**:

a) Это линейная функция, ее гесиан равен 0, значит, a(x) выпукла и вогнута. ✓

b) g(x) - вогнутая функция. Покажем это:

По дифференциальному критерию первого порядка, достаточно показать, что $\forall$x, y $\in R^n_{++}$:

g(y) $\leq$ g(x) + ∇g(x)$^T(y - x)$

$g(x)\neq$ 0

$(\prod\limits_{k=1}^{n} \frac{y_k}{x_k})^{\frac{1}{n}} \leq 1 + \frac{1}{n} \sum\limits_{k=1}^{n} \frac{y_k - x_k}{x_k}$

$(\prod\limits_{k=1}^{n} \frac{y_k}{x_k})^{\frac{1}{n}} \leq \frac{1}{n} \sum\limits_{k=1}^{n} \frac{y_k}{x_k}$

Последнее неравенство - это неравенство Коши о среднем: среднее геометрическое не превосходит среднего арифметического. ✓

5. **solution**:
Функция f(x) = -xlnx - (1 - x)ln(1 - x) определена на (0, 1)
f'(x) = -x$\frac{1}{x}$ - lnx + ln(1 - x) + 1 = ln(1 - x) - lnx
f''(x) = -$\frac{1}{1-x}$ - $\frac{1}{x}$ = $\frac{1}{(1-x)x}$ < 0
Значит, f строго вогнута на (0, 1). ✓