hw_markevich_3

poniedziałek, 21 października 2024 01:17

Упр 1

1. Базовая задача МНК.

Теоретический вопрос 1.

Пусть дана выборка точек y_i. Решите аналитически задачу МНК, моделируя данные постоянной величиной $\hat{y}$, что отвечает минимизации функции потерь

$$\mathcal{L} = \sum_{i=1}^{l} (y_i - \hat{y})^2 \to \min_{\hat{y}}.$$

Теоретический вопрос 2.

Покажите, что прямая, построенная по методу МНК, всегда проходит через точку $(\bar{x}, \bar{y})$, где $\bar{x}$ и $\bar{y}$ - выборочные средние. Обобщите на случай многомерной регрессии.

1) Обозн. $L = \sum_{i=1}^{l} (y_i - \hat{y})^2 \to \min$

$$\frac{\partial L}{\partial \hat{y}} = -\sum_{i=1}^{l} 2(y_i - \hat{y}) = 0 \Rightarrow \sum_{i=1}^{l} (y_i - \hat{y}) = 0 = \sum_{i=1}^{l} y_i - l\hat{y} = 0 \Rightarrow \hat{y} = \frac{\sum_{i}^{l} y_i}{l} = \bar{y}$$

2) $L = \sum_{i=1}^{l} (y_i - kx_i - b)^2$

$$\frac{\partial L}{\partial k} = \sum_{i=1}^{l} -2(y_i - kx_i - b)x_i = 0 \; ; \; \frac{\partial L}{\partial b} = \sum_{i=1}^{l} -2(y_i - kx_i - b) = 0 \Rightarrow \begin{cases} \sum_{1}^{l} y_i - lb - k\sum_{1}^{l} x_i = 0 & \Big| \cdot \frac{1}{l} \\ \sum_{i=1}^{l} y_i x_i - lb\sum x_i - k\sum^{l} x_i^2 = 0 & \Big| \cdot \frac{1}{l} \end{cases}$$

$\bar{y} - b - k\bar{x} = 0 \Rightarrow$ проходит ч/з $(\bar{x}, \bar{y})$

многомерная регрессия: $L = \sum_{0}^{l} (y_i - \sum_{j=1}^{n} k_j x_{ji} - b)^2 \Rightarrow \frac{\partial L}{\partial b} = -2\sum_{i=1}^{l} (y_i - \sum_{j=1}^{n} k_j x_{ji} - b) = 0$

$$\sum_{i=1}^{l} y_i - lb - \sum_{j=1}^{n} \left(\sum_{i=1}^{l} k_j x_{ji}\right) = 0 \; \Big| \cdot \frac{1}{l}$$

$\bar{y} - b - \sum_{j=1}^{n} k_j \bar{x}_j = 0 \Rightarrow$ проходит ч/з $(\bar{x}_1, \dots, \bar{x}_n, \bar{y})$

Упр 2

2. Централизация признаков и МНК.

Покажите, что следующие две процедуры приводят к одинаковому результату:

1. В матрице объект-признак X из каждого столбца вычитается среднее по столбцу (централизация признаков). После этого вычисляется $(X^T X)^{-1}$.
2. К матрице X дописывается в конец столбец, состоящий из одних единиц. Вычисляется $(X^T X)^{-1}$ и в получившейся матрице вычеркивается последний столбец и последняя строка.

1. Рассмотрим $\mu = \frac{1}{n} \begin{pmatrix} \sum_{i=1}^{n} x_{i1} \\ \vdots \\ \sum_{i=1}^{n} x_{im} \end{pmatrix}$ – транспон. строка со средними знач. по столбцу

$X_1 = X - \mathbb{1}_n \mu^T \Rightarrow X_1^T = X^T - \mu \mathbb{1}_n^T$ ← ф. строка

$\bullet\; X_1^T X_1 = (X^T - \mu \mathbb{1}_n^T)(X - \mathbb{1}_n \mu^T) = X^T X - X^T \mathbb{1}_n \mu^T - \mu \mathbb{1}_n^T X + \mu \mathbb{1}_n^T \mathbb{1}_n \mu^T = X^T X - n\mu\mu^T - n\mu\mu^T + n\mu\mu^T =$

$= X^T X - n\mu\mu^T$

$\bullet\; (X_1^T X_1)^{-1} = (X^T X - n\mu\mu^T)^{-1}$

2. Рассм $X_2 = (X \mid \mathbb{1}_n)$; X_i – столбцы X

$$X_2^T X_2 = \begin{pmatrix} X^T \\ \mathbb{1}_n^T \end{pmatrix} (X \; \mathbb{1}_n) = \begin{pmatrix} X^T X & X^T \mathbb{1}_n \\ \mathbb{1}_n^T X & \mathbb{1}_n^T \mathbb{1}_n \end{pmatrix} = \begin{pmatrix} X^T X & \sum_{1}^{n} X_i \\ \sum_{1}^{n} X_i^T & n \end{pmatrix}$$

вычеркиваются последние

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \Big/\!\!\Big/ \text{где } A, B, C, D - \text{подматрицы} \Big/\!\!\Big/ = \begin{pmatrix} A^{-1} + A^{-1} B H^{-1} C A^{-1} & -A^{-1} B H^{-1} \\ -H^{-1} C A^{-1} & H^{-1} \end{pmatrix}, \text{ где } H = D - CA^{-1}B$$

У нас $A = X^T X$, $B = X^T \mathbb{1}_n$, $C = \mathbb{1}_n^T X$, $D = n$, $H = n - \mathbb{1}_n^T X (X^T X)^{-1} X^T \mathbb{1}_n \in \mathbb{R}$

Тогда $$\left((X^T X)^{-1} + \frac{(X^T X)^{-1} X^T \mathbb{1}_n \mathbb{1}_n^T X (X^T X)^{-1}}{n - n^2 \mu^T (X^T X)^{-1} \mu} \right) \times (X^T X - n\mu\mu^T) = E - n(X^T X)^{-1} \mu\mu^T +$$

$+\; n(X^TX)^{-1} \dots$ [illegible]

Тогда $\left((X^TX)^{-1} + \frac{\ldots}{n - n^2\mu^T(X^TX)^{-1}\mu}\right) \times (X^TX - n\mu\mu^T) = E - n(X^TX)^{-1}\mu\mu^T +$

$$+ \frac{n(X^TX)^{-1}\mu\mu^T}{1 - n\mu^T(X^TX)^{-1}\mu} - \frac{n^2(X^TX)^{-1}\mu(\mu^T(X^TX)^{-1}\mu)\mu^T}{1 - n\mu^T(X^TX)^{-1}\mu} = E + \frac{-n(X^TX)^{-1}\mu\mu^T + n(X^TX)^{-1}\mu\mu^T(X^TX)^{-1}\mu}{1 - n\mu^T(X^TX)^{-1}\mu} +$$

$$+ \frac{n(X^TX)^{-1}\mu\mu^T - n^2(X^TX)^{-1}\mu\mu^T(X^TX)^{-1}\mu\mu^T}{1 - n\mu^T(X^TX)^{-1}\mu} = E \quad \text{ч.т.д.}$$

$f_1 = x^2 \quad f_2 = \sin x$

$x_1 = 1,\ x_2 = 2,\ x_3 = 3$

$$X = \begin{matrix} 1 & \sin 1 & 1 \\ 4 & \sin 2 & 1 \\ 9 & \sin 3 & 1 \end{matrix}$$

3. Геометрический смысл псевдообратной матрицы.

На лекции обсуждалось, что метод наименьших квадратов - это способ поставить задачу о решении переопределенной системы $Xw = y$, которая имеет явный ответ, выражающийся через левую псевдообратную матрицу для X. Для недоопределенной системы $Xw = y$ (имеющей бесконечно много решений) можно поставить задачу о поиске решения с минимальной l_2-нормой весов $\|w\|^2 = w^T w$. Решите такую задачу и покажите, что ответ выражается через правую псевдообратную матрицу для X. Считайте, что прямоугольная матрица X имеет полный ранг (максимально возможный).

$\underset{n\times m}{X}\underset{m\times 1}{w} = y^{\,n\times 1}$

$\|w\|^2 = w^T w = w^\alpha w_\alpha$. Хотим найти min $G = \|w\|^2$ при усл $Xw = y$; λ — столбец мн. Лагранжа

$$\mathcal{L}(w;) = w_\alpha^2 + \underset{1\times m}{\lambda^T}(Xw - y) \Rightarrow \begin{cases} \mathcal{L}'_w = 2w + X^T\lambda = 0 \\ \mathcal{L}'_\lambda = Xw - y = 0 \end{cases} \Rightarrow \begin{matrix} w = -\frac{1}{2}X^T\lambda \\ -\frac{1}{2}XX^T\lambda = y \end{matrix}$$

$$\lambda = -2(XX^T)^{-1}y \Rightarrow w = X^T(XX^T)^{-1}y$$

4. Матрица объект-признак.

Теоретический вопрос 1.

Пусть X - матрица объект-признак (размерность $l \times F$), для которой сингулярное разложение имеет вид $X = V\sqrt{\Lambda}U^T$. После понижения размерности данных с помощью метода главных компонент, в диагональной матрице $\Lambda = \operatorname{diag}\{\lambda_1 \geq \cdots \geq \lambda_F\}$ оставляются только $\tilde{F}$ наибольших сингулярных чисел: $\tilde{\Lambda} = \operatorname{diag}\{\lambda_1 \geq \cdots \geq \lambda_{\tilde{F}}\}$. При этом данные, как правило, можно восстановить только с некоторой ошибкой: $\tilde{X} = V\sqrt{\tilde{\Lambda}}U^T \neq X$. Покажите, что L_2 норма ошибки выражается через сумму по оставшимся сингулярным числам:

$$\frac{1}{l}\|X - \tilde{X}\|^2 = \sum_{i=\tilde{F}+1}^{F} \lambda_i$$

Теоретический вопрос 2.

Покажите, что сингулярный вектор матрицы X, отвечающий наибольшему сингулярному числу, является решением задачи

$$\boldsymbol{u} = \operatorname{argmax}_{||\boldsymbol{u}||=1}(X\boldsymbol{u})^2,$$

где подразумевается матричное умножение X на $\boldsymbol{u}$.

1) $X - \tilde{X} = V(\sqrt{\Lambda} - \sqrt{\tilde{\Lambda}})U^T = V\begin{pmatrix} 0 & & & \\ & \ddots & & \\ & & \sqrt{\lambda_{\tilde{F}+1}} & \\ & & & \ddots \ \sqrt{\lambda_F} \end{pmatrix}U^T \Rightarrow \frac{1}{l}\|F - \tilde{F}\|^2 = \sum_{i=\tilde{F}+1}^{F}\lambda_i$

2) $u = \underset{\|u\|=1}{\arg\max}(Xu)^2$. $\exists\ X^TXu = \lambda_{max}u$

$(Xu)^2 = u^TX^TXu = u^T\lambda_{max}u = \lambda_{max}$, а больше не может быть т.к $\sqrt{\lambda_{max}}$ - макс увел нормы вектора от оператора X

5. Геометрический смысл сингулярного разложения.

Теоретический вопрос 1.

Пусть дан набор точек на плоскости (x_i, y_i), для которых выборочные средние x_i и y_i равны нулю. Покажите, что сингулярный вектор для матрицы объект-признак, отвечающий наибольшему сингулярному числу, задает прямую a (проходящую через начало координат), которая является решением следующей задачи оптимизации:

$$L' = \sum_{i=1}^{N} \operatorname{distance}^2\left[(x_i, y_i); a\right] \quad \longrightarrow \quad \min_a,$$

где distance $[(x_i, y_i); a]$ - расстояние от точки (x_i, y_i) до прямой a (равное длине перпендикуляра). Обратите внимание, что такая задача отличается от задачи МНК, в которой расстояние от точки до аппроксимирующей прямой вычисляется не по перпендикуляру, а вдоль оси y, отвечающей целевой переменной.

где distance $[(x_i, y_i)\,; a]$ - расстояние от точки (x_i, y_i) до прямой a (равное длине перпендикуляра). Обратите внимание, что такая задача отличается от задачи МНК, в которой расстояние от точки до аппроксимирующей прямой вычисляется не по перпендикуляру, а вдоль оси y, отвечающей целевой переменной.

Теоретический вопрос 2.

Пусть дан набор из N точек в трехмерном пространстве $X_{i\alpha}, i \in \{1, \dots, N\}, \alpha \in \{1, 2, 3\}$. Покажите, что задача нахождения сингулярных чисел матрицы X эквивалентна нахождению главных моментов инерции твердого тела, составленного из набора точечных масс, расположенных в точках (X_{i1}, X_{i2}, X_{i3}) (можно представлять себе, что точечные массы соединены между собой невесомыми и абсолютно жесткими стержнями).

2) $$\mathcal{J} = \begin{pmatrix} \mathcal{J}_x & -\mathcal{J}_{xy} & -\mathcal{J}_{xz} \\ -\mathcal{J}_{yx} & \mathcal{J}_y & -\mathcal{J}_{yz} \\ -\mathcal{J}_{zx} & -\mathcal{J}_{zy} & \mathcal{J}_z \end{pmatrix}, \quad \mathcal{J}_x = \sum_{j=1}^{N} (y_j^2 + z_j^2), \quad \mathcal{J}_{xy} = \sum_{j=1}^{N} x_j y_j \Rightarrow \mathcal{J}_{ij} = \left(\sum_{k=1}^{N} r_k^2\right)\delta_{ij} + X_{ki} X_{kj} \Rightarrow \mathcal{J} = \sum_{k=1}^{N} r_k^2 E - X^T X = \|X\|_F^2 E - X^T X$$

Тогда $\lambda_{m\mathcal{J}} = \|X\|_F^2 - \lambda_{m X^T X} = \|X\|_F^2 - \sigma_{m_X}^2$

1) $\lambda_{m_{\mathcal{J}}} = \|X\|_F^2 - \lambda_{m_{X^T X}} \Rightarrow$ максимальному спектр. числу соотв-ет миним. гл. момент инерции

в главных осях расстояние до оси $x = y^2 + z^2 = \mathcal{J}_x \Rightarrow$ и т.д.