Решение теоретических заданий
Вычислительная математика (пилотный курс)

Коцевич Андрей Витальевич, группа Б02-920

5 семестр, 2021

Содержание

В папках А2 и А3 было 2 листика с домашним заданием. Я выбирал те, где задач больше.

1 А2

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

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

Решение.

$$\frac{\partial \mathcal{L}}{\partial \tilde{y}} = -2 \sum_{i=1}^{l} (y_i - \tilde{y}) = 0 \tag{2}$$

$$\sum_{i=1}^{l} y_i - l\tilde{y} = 0 \tag{3}$$

$$\boxed{\tilde{y} = \frac{\sum\limits_{i=1}^{l} y_i}{l} = \bar{y}} \tag{4}$$

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.
Решение.
Линейная регрессия:

$$g_w(x) = xw + b \tag{5}$$

Формула МНК:

$$b = \bar{y} - w\bar{x} \tag{6}$$

$$g_w(x) = xw + \bar{y} - w\bar{x} = w(x - \bar{x}) + \bar{y} \to \boxed{g_w(\bar{x}) = \bar{y}} \tag{7}$$

По аналогии, в случае многомерной регрессии гиперплоскость, построенная по методу МНК, всегда проходит через точку $(\bar{x}_1, ..., \bar{x}_n, \bar{y}_1, ..., \bar{y}_m)$.
5. 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).
Решение. Для поиска решения с минимальной l_2-нормой весов $||w||^2 = w^T w$ (условного экстремума) воспользуемся методом множителей Лагранжа:

$$\mathcal{L} = \lambda^T (Xw - y) + w^T w \to \min_{w} |_{w \to \min} \tag{8}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda} = (Xw - y)^T = 0 \to y = Xw \tag{9}$$

$$\frac{\partial \mathcal{L}}{\partial w} = \lambda^T X + 2w^T = 0 \to w^T = -\frac{1}{2}\lambda^T X \tag{10}$$

Транспонируем левую и правую части:

$$w = -\frac{1}{2} X^T \lambda \tag{11}$$

$$y = Xw = -\frac{1}{2} X X^T \lambda \to \lambda = -2(XX^T)^{-1} y \tag{12}$$

$$\lambda = -2(X^{-1})^T X^{-1} y = -2(XX^T)^{-1} y \tag{13}$$

Отсюда получаем

$$\boxed{w = X^T (XX^T)^{-1} y} \tag{14}$$

где $R = X^T(XX^T)^{-1}$ – правая псевдообратная матрица.

7. На лекции обсуждался учет влияния систематической погрешности путем усреднения решения задачи МНК по гауссовому нормальному распределению для y-координат точек выборки: $\tilde{y}_i \sim \mathcal{N}(y_i, s^2)$, где погрешность по оси ординат считалась равной s. Обобщите этот вывод на случай, когда каждая точка имеет свою y-погрешность s_i. Для этого проведите усреднение по многомерному нормальному распределению для $\tilde{y}_i$ с произвольной симметричной матрицей ковариации A[1]:

$$\tilde{y} \sim \frac{1}{(2\pi)^{\frac{l}{2}} \sqrt{\det A}} \exp\left(-\frac{(\tilde{y} - y)^T A^{-1} (\tilde{y} - y)}{2}\right), \tag{15}$$

где $y = (y_1, ..., y_l)^T$, а $\tilde{y} = (\tilde{y}_1, ..., \tilde{y}_l)^T$.

1. Покажите,что распределение правильно нормированно. Указание: Выполните замену координат $\tilde{y} - y = Sz$, где матрица S диагонализует A.

2. Вычислите неприводимые парные корреляторы $\ll \tilde{y}_i \tilde{y}_j \gg$, усредняя по распределению. Указание: Сделайте замену $\tilde{y} - y = Y$. Для вычисления гауссового интеграла с предэкспонентой вычислите интеграл $\int d^l Y \exp(-Y^T A^{-1} Y/2 + J^T Y)$ и выполните дифференцирование по параметрам J_i (компоненты вектора J).

3. Оцените погрешности параметров модели w_α, следуя вычислению, приведенному на лекции, и используя корреляторы, полученные в предыдущем пункте.

4. Запишите решение в частном случае диагональной матрицы $A = \mathrm{diag}(A_1, ..., A_l)$. Как следует выбирать величины A_i для моделирования y-погрешности i-ой точки, равной s_i?

Решение.

1. Поскольку матрица A^{-1} симметрична, то $\exists S : S^T A^{-1} S = E$ и $\det S = \sqrt{\det A}$. Выполним замену координат $\tilde{y} - y = Sz$.

$$\int \tilde{y} d^l \tilde{y} = \frac{\det S}{(2\pi)^{\frac{l}{2}} \sqrt{\det A}} \int d^l z \exp\left(-\frac{zz^T}{2}\right) = \frac{1}{(2\pi)^{\frac{l}{2}}} (2\pi)^{\frac{l}{2}} = 1 \tag{16}$$

Распределение правильно нормированно.

2. Сделаем замену $\tilde{y} - y = Y$.

$$\ll \tilde{y}_i \tilde{y}_j \gg = \langle (\tilde{y}_i - y_i)(\tilde{y}_j - y_j) \rangle = \langle Y_i Y_j \rangle \tag{17}$$

Обозначим интеграл с током

$$\begin{aligned} I_J = \frac{1}{(2\pi)^{\frac{l}{2}}\sqrt{\det A}} \int d^l Y \exp\left(-\frac{Y^T A^{-1} Y}{2} + J^T Y\right) = \\ = \frac{1}{(2\pi)^{\frac{l}{2}}\sqrt{\det A}} \int d^l Y \exp\left(\frac{(Y - AJ)^T A^{-1}(Y - AJ)}{2} + \frac{(AJ)^T A^{-1} AJ}{2}\right) = \\ = \exp\left(\frac{J^T A^T J}{2}\right) = \exp\left(\frac{J^T A J}{2}\right), \end{aligned} \tag{18}$$

поскольку A – симметричная.

$$\frac{\partial I_J}{\partial J_i}|_{J=0} = \frac{1}{(2\pi)^{\frac{l}{2}}\sqrt{\det A}} \int d^l Y Y_i \exp\left(-\frac{Y^T A^{-1} Y}{2}\right) = \langle Y_i \rangle = 0 \tag{19}$$

$$\frac{\partial^2 I_J}{\partial J_i \partial J_j}|_{J=0} = \frac{1}{(2\pi)^{\frac{l}{2}}\sqrt{\det A}} \int d^l Y Y_i Y_j \exp\left(-\frac{Y^T A^{-1} Y}{2}\right) = \langle Y_i Y_j \rangle \tag{20}$$

$$\frac{\partial^2 I_J}{\partial J_i \partial J_j}|_{J=0} = A_{ij} \tag{21}$$

Таким образом, неприводимые парные корреляторы

$$\boxed{\ll \tilde{y}_i \tilde{y}_j \gg = A_{ij}} \tag{22}$$

3.

$$w = Qy, \quad Q = (X^T X)^{-1} X^T \tag{23}$$

В индексных обозначениях:

$$w_\alpha = Q_{\alpha i} y_i, \quad \tilde{w}_\alpha = Q_{\alpha i} \tilde{y}_i \tag{24}$$

$$\mathrm{var}\tilde{w}_\alpha = \ll \tilde{w}_\alpha \tilde{w}_\alpha \gg = Q_{\alpha i} Q_{\alpha j} \ll \tilde{y}_i \tilde{y}_j \gg = Q_{\alpha i} Q_{\alpha j} A_{ij} \tag{25}$$

$$\boxed{\mathrm{var}\tilde{w}_\alpha = (QAQ^T)_{\alpha\alpha}} \tag{26}$$

4. Из предыдущего пункта:

$$\boxed{\mathrm{var}\tilde{w}_\alpha = \sum_{i=1}^{n} Q_{\alpha i}^2 A_i} \tag{27}$$

Для моделирования y-погрешности i-ой точки, равной s_i, нужно выбрать A_i:

$$\boxed{A_i = \ll \tilde{y}_i \tilde{y}_i \gg = s_i^2} \tag{28}$$

Необязательное дополнение. Случай линейной регрессии:

$$X = \begin{pmatrix} x_1 & 1 \\ ... & ... \\ x_l & 1 \end{pmatrix}, \quad Q = (X^T X)^{-1} X^T = \frac{1}{l(\text{var}x)^2} \begin{pmatrix} x_1 - \bar{x} & ... & x_l - \bar{x} \\ \bar{x^2} - \bar{x}x_1 & ... & \bar{x^2} - \bar{x}x_l \end{pmatrix} \tag{29}$$

$$\text{var}\tilde{w}_1 = \frac{1}{l(\hat{\text{var}}x)^2} \left(\overline{x^2 A^2} - 2\bar{x}\overline{xA^2} + \bar{x}^2 \overline{A^2} \right), \quad \text{var}\tilde{w}_0 = \frac{1}{l(\hat{\text{var}}x)^2} \left(\overline{x^2}^2 \overline{A^2} - 2\bar{x}\overline{x^2}\, \overline{xA^2} + \bar{x}^2 \overline{x^2 A^2} \right)$$

$$\boxed{\text{var}\tilde{w}_1 = \frac{1}{l(\hat{\text{var}}x)^2} \left(\overline{x^2 s^2} - 2\bar{x}\overline{xs^2} + \bar{x}^2 \overline{s^2} \right), \quad \text{var}\tilde{w}_0 = \frac{1}{l(\hat{\text{var}}x)^2} \left(\overline{x^2}^2 \overline{s^2} - 2\bar{x}\overline{x^2}\, \overline{xs^2} + \bar{x}^2 \overline{x^2 s^2} \right)} \tag{30}$$

При $s_i = s = \text{const}$ получаем формулы с лекции:

$$\text{var}\tilde{w}_1 = \frac{s^2}{l\hat{\text{var}}x}, \quad \text{var}\tilde{w}_0 = \frac{s^2 \overline{x^2}}{l\hat{\text{var}}x} \tag{31}$$

2 A3

1. Пусть дана выборка точек на прямой $\{x_i\}$. Максимизируйте правдоподобие (или его логарифм) в гауссовой вероятностной модели:

$$\prod_i p(x_i) \to \max_{\mu,\sigma}, \quad p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right) \tag{32}$$

Решение. Пусть при измерении величины μ мы получили выборку $\{x_i\}$. Плотность вероятности того, что величина μ примет значение x_i:

$$p(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2} \right) \tag{33}$$

Плотность вероятности того, что величина μ примет значения x_1, ..., x_n:

$$p(x_1, ..., x_n) = \prod_{i=1}^n \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2} \right) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^{\frac{n}{2}} \exp\left(-\sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2} \right) \tag{34}$$

Логарифм правдоподобия:

$$\log p(x_1, ..., x_n) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2 \tag{35}$$

Условия максимума правдоподобия:

$$\frac{\partial(\log p(x_1, ..., x_n))}{\partial\mu} = \frac{\sum\limits_{i=1}^n (x_i-\mu)}{\sigma^2} = 0 \to \boxed{\hat{\mu} = \frac{\sum\limits_{i=1}^n x_i}{n} = \bar{x}_i} \tag{36}$$

$$\frac{\partial(\log p(x_1, ..., x_n))}{\partial\sigma} = -\frac{n}{\sigma} + \frac{\sum\limits_{i=1}^n (x_i-\mu)^2}{\sigma^3} = 0 \to \hat{\sigma}^2 = \frac{\sum\limits_{i=1}^n (x_i-\mu)^2}{n} = \frac{\sum\limits_{i=1}^n (x_i-\bar{x})^2}{n} \tag{37}$$

$$\boxed{\hat{\sigma}^2 = \hat{\text{var}}\ x} \tag{38}$$

2. Количество срабатываний счетчика Гейгера за минуту n подчиняется распределению Пуассона:

$$P_\lambda(n) = \frac{\lambda^n}{n!}e^{-\lambda} \tag{39}$$

1. В ходе эксперимента счетчик Гейгера сработал за минуту m раз. С помощью теоремы Байеса определите апостериорное распределение на λ. Указание: априорную плотность вероятности λ можно считать постоянной (так как мы изначально ничего не знаем про λ).

2. Эксперимент повторили ещё раз, в этот раз счетчик Гейгера сработал за минуту m' раз. Как обновилось апостериорное распределение на λ?

Решение.

1. $P(\lambda) = P_0 = \text{const}$ – априорное распределение, $P(\lambda|m)$ – апостериорное. По теореме Байеса:

$$P(\lambda|m) = \frac{P(m|\lambda)P(\lambda)}{P(m)} = \frac{P(m|\lambda)P(\lambda)}{\int d\lambda P(m|\lambda)P(\lambda)} = \frac{P_\lambda(m)}{\int\limits_0^\infty d\lambda P_\lambda(m)} = P_\lambda(m) \tag{40}$$

$$\boxed{P(\lambda|m) = \frac{\lambda^m}{m!}e^{-\lambda}} \tag{41}$$

2. Теперь $P(\lambda) = P(\lambda|m) = P_\lambda(m)$ - априорное распределение, $P(\lambda|m,m')$ - апостериорное. По теореме Байеса:

$$P(\lambda|m,m') = \frac{P(m'|\lambda)P(\lambda)}{P(m')} = \frac{P(m'|\lambda)P(\lambda)}{\int d\lambda P(m'|\lambda)P(\lambda)} = \frac{P_\lambda(m')P_\lambda(m)}{\int\limits_0^\infty d\lambda P_\lambda(m')P_\lambda(m)} \tag{42}$$

$$\boxed{P(\lambda|m,m') = \frac{2^{1+m+m'}\lambda^{m+m'}}{(m+m')!}e^{-2\lambda}} \tag{43}$$

3. Ультрачувствительный тест от короновируса ошибается в 1%случаев (как в одну, так и в другую сторону). В данный момент в популяцiидоля заболевших 10^{-5}. Петя получил положительный тест на коронавирус.С какой вероятностью он действительно болеет коронавирусом? **Решение.** Обозначим через событие A - болезнь коронавирусом, через B -положительный результат теста. Тогда

$$P(A) = 10^{-5}, \quad P(B|A) = 0,99, \quad P(B|\neg A) = 0,01 \tag{44}$$

По теореме Байеса:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\neg A)P(\neg A)} \tag{45}$$

$$\boxed{P(A|B) = \frac{0,99 \cdot 10^{-5}}{0,99 \cdot 10^{-5} + 0,01 \cdot (1 - 10^{-5})} = 9,89 \cdot 10^{-4}} \tag{46}$$

Как видно, Пете следует сделать повторный тест.

4. Пусть имеется априорное распределение на вектор $\vec{x}$,задаваемое матрицей A:

$$p_0(\vec{x}) = \frac{1}{Z} e^{-\frac{\vec{x}^T A \vec{x}}{2}} \tag{47}$$

Было произведено измерение величин $\vec{x}$, которое дало значение $\vec{x}_1$. Найдите апостериорное распределение на $\vec{x}$.

Решение.

Распишем априорное распределение на вектор $\vec{x}$:

$$p_0(\vec{x}) = \frac{1}{Z} e^{-\frac{\sum\limits_{i,j=1}^{n} x_i A_{ij} x_j}{2}} = \frac{1}{Z} \prod_{i,j=1}^{n} e^{-\frac{x_i A_{ij} x_j}{2}} \tag{48}$$

$P(A) = P_0 = \text{const}$ – априорное распределение A, $P(A|\vec{x}_1)$ – апостериорное. По теореме Байеса:

$$P(A|\vec{x}_1) = \frac{P(\vec{x}_1|A)P(A)}{P(\vec{x}_1)} = \frac{P(\vec{x}_1|A)P(A)}{\int\limits_{-\infty}^{\infty} P(\vec{x}_1|A)P(A)dA} = \frac{\prod\limits_{i,j=1}^{n} e^{-\frac{x_{1i} A_{ij} x_{1j}}{2}} P_0}{\prod\limits_{i,j=1}^{n} \int\limits_{0}^{\infty} e^{-\frac{x_{1i} A_{ij} x_{1j}}{2}} P_0 dA_{ij}} \tag{49}$$

$$P(A|\vec{x}_1) = \frac{\prod\limits_{i,j=1}^{n} e^{-\frac{x_{1i} A_{ij} x_{1j}}{2}}}{\prod\limits_{i,j=1}^{n} \frac{2}{x_{1i} x_{1j}}} = \frac{\prod\limits_{i=1}^{n} x_{1i}^{n^2}}{2^{n^2}} e^{-\frac{\sum\limits_{i,j=1}^{n} x_{1i} A_{ij} x_{1j}}{2}} \tag{50}$$

Апостериорное распределение на $\vec{x}$:

$$P(\vec{x}) = \int\limits_{0}^{\infty} P(\vec{x}|A)P(A|\vec{x}_1)dA = \frac{\prod\limits_{i=1}^{n} x_{1i}^{n^2}}{2^{n^2} Z} \prod_{i,j=1}^{n} \int\limits_{0}^{\infty} e^{-\frac{x_i A_{ij} x_j}{2}} e^{-\frac{x_{1i} A_{ij} x_{1j}}{2}} dA_{ij} \tag{51}$$

$$\boxed{P(\vec{x}) = \frac{\prod\limits_{i=1}^{n} x_{1i}^{n^2}}{Z} \prod_{i,j=1}^{n} \frac{1}{x_i x_j + x_{1i} x_{1j}}} \tag{52}$$

7. Покажите, что задача минимизации квадратичной функции потерь с дополнительным ограничением (лассо Тибширани):

$$\mathcal{L} = ||Xw - y||^2 \to \min_{w}, \quad \sum_{\alpha} |w_\alpha| < C \tag{53}$$

эквивалентна L_1-регуляризации. *Указание:* можно воспользоваться условиями Каруша-Куна-Таккера (обобщение метода Лагранжа).

Доказательство. Для исследования дополнительного ограничения в задаче минимизации квадратичной функции потерь (лассо Тибширани) воспользуемся множителем Лагранжа μ:

$$\mathcal{L}' = ||Xw - y||^2 + \mu\left(\sum_\alpha |w_\alpha| - C\right) \tag{54}$$

Достаточные условия Каруша - Куна - Таккера:

$$\begin{cases} \min_w \mathcal{L}'(w) = \mathcal{L}'(\hat{w}), \\ \mu\left(\sum\limits_\alpha |\hat{w}_\alpha| - C\right) = 0, \\ \mu > 0 \end{cases} \tag{55}$$

L_1-регуляризация:

$$\mathcal{L}'' = ||Xw - y||^2 + \mu\sum_\alpha |w_\alpha| \to \min_w \tag{56}$$

Решение задачи минимизации удовлетворяет L_1-регуляризации, поскольку при выполнении условий Каруша - Куна - Таккера при $w_a = \hat{w}_a$: $\mathcal{L}'' = \mu C \to$ достигает минимума. Обратно, устремление к минимуму $\mathcal{L}''$ при L_1-регуляризации приводит к условиям Каруша-Куна-Таккера. □

8.* Bias-Variance decomposition

Воспользуемся вероятностной моделью данных, в которой предполагается, что каждый элемент выборки не зависимо от других поступает из распределения $p(x, y)$. Тогда вероятность получить какой-то конкретный набор данных $(X_l, y_l) = (x_1, ...x_l; y_1, ..., y_l)$ в обучающей выборке равна $p(X_l, y_l) = \prod\limits_{i=1}^{l} p(x_i, y_i)$. В дальнейшем будем обозначать как (x, y) элемент тестовой выборки, который не входит в (X_l, y_l).

В выбранной модели $\breve{y} = g_\theta(x)$ параметры θ определяются с помощью фиттирования по обучающей выборке: $\theta = \theta(X_l, y_l)$, поэтому $\breve{y}$ зависит от x, X_l и y_l. Тогда формальное выражение для функции потерь (соответствующее пределу бесконечной большой тестовой выборки) можно записать как

$$L = \mathbb{E}_{X_l,y_l}[\mathbb{E}_{x,y}(y - \breve{y})^2] \tag{57}$$

В этом выражении квадратичная функция потерь усредняется по элементу тестовой выборки (x, y) и по обучающей выборке (X_l, y_l).

Покажите, что справедливо разложение этой величины на шум, смещение и разброс:

$$L = \mathbb{E}_{x,y}(y - \mathbb{E}(y|x))^2 + \mathbb{E}_{x,y}(\mathbb{E}_{X_l,y_l}(\breve{y}) - \mathbb{E}(y|x))^2 + \mathbb{E}_{x,y}(\mathbb{E}_{X_l,y_l}(\breve{y} - \mathbb{E}_{X_l,y_l}\breve{y})^2) \tag{58}$$

где $\mathbb{E}_{x,y}(y - \mathbb{E}(y|x))^2$ – noise, $\mathbb{E}_{x,y}(\mathbb{E}_{X_l,y_l}(\breve{y}) - \mathbb{E}(y|x))^2$ – bias, $\mathbb{E}_{x,y}(\mathbb{E}_{X_l,y_l}(\breve{y} - \mathbb{E}_{X_l,y_l}\breve{y})^2)$ – variance.

Указание: сначала покажите, что

$$\mathbb{E}_{x,y}(y - \breve{y})^2 = \mathbb{E}_{x,y}(y - \mathbb{E}(y|x))^2 + \mathbb{E}_{x,y}(\mathbb{E}(y|x) - \breve{y})^2 \tag{59}$$

Доказательство. Пусть $y = f(x) + \varepsilon$ – функция с шумом, где шум ε имеет $\mathbb{E}(\varepsilon) = 0$ и $\operatorname{var}\varepsilon = \mathbb{E}_{x,y}(y - \mathbb{E}(y|x))^2$.

$$\mathbb{E}_{x,y}\, y = \mathbb{E}_{x,y}(f(x) + \varepsilon) = \mathbb{E}_{x,y}(f(x)) = f(x) \tag{60}$$

$$\operatorname{var} y = \mathbb{E}_{x,y}(y - \mathbb{E}_{x,y}(y))^2 = \mathbb{E}_{x,y}(y - f)^2 = \mathbb{E}_{x,y}\varepsilon^2 = (\mathbb{E}_{x,y}\varepsilon)^2 + \operatorname{var}\varepsilon = \operatorname{var}\varepsilon \tag{61}$$

Поскольку ε и f независимы, то запишем

$$\mathbb{E}_{x,y}(y-\breve{y})^2 = \mathbb{E}_{x,y}(y^2+\breve{y}^2-2y\breve{y})^2 = \text{var } y + \mathbb{E}_{x,y}y^2 + \text{var}\breve{f} + \mathbb{E}_{x,y}\breve{f}^2 - 2f\mathbb{E}_{x,y}\breve{y} \tag{62}$$

$$\mathbb{E}_{x,y}(y-\breve{y})^2 = \text{var } y + \text{var}\breve{f} + (f^2 - 2f\mathbb{E}_{x,y}\breve{y} + \mathbb{E}_{x,y}\breve{y}^2) = \text{var } y + \text{var}\breve{y} + (f-\mathbb{E}_{x,y}\breve{y})^2 \tag{63}$$

где 1 слагаемое – прообраз noise, второе – variance, третье – bias. Усредняя также по обучающей выборке, получаем требуемое. □

3 A4

2. 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 $\Lambda = \text{diag}\{\lambda_1 \geq ... \geq \lambda_F\}$ оставляются только $\tilde{F}$ наибольших сингулярных чисел: $\tilde{\Lambda} = \text{diag}\{\lambda_1 \geq ... \geq \lambda_{\tilde{F}}\}$. 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:

$$\frac{1}{l}||X-\tilde{X}||^2 = \sum_{i=\tilde{F}+1}^{F}\lambda_i \tag{64}$$

Доказательство.

$$X = V\sqrt{\Lambda}U^T, \quad \tilde{X} = V\sqrt{\tilde{\Lambda}}U^T \tag{65}$$

$$X - \tilde{X} = V\left(\sqrt{\Lambda}-\sqrt{\tilde{\Lambda}}\right)U^T \tag{66}$$

$$\begin{aligned}||X-\tilde{X}||^2 &= \text{Tr}((X-\tilde{X})(X-\tilde{X})^T) = \text{Tr}\left(V\left(\sqrt{\Lambda}-\sqrt{\tilde{\Lambda}}\right)U^T\left(V\left(\sqrt{\Lambda}-\sqrt{\tilde{\Lambda}}\right)U^T\right)^T\right) = \\ &= \text{Tr}\left(V\left(\sqrt{\Lambda}-\sqrt{\tilde{\Lambda}}\right)U^TU\left(\sqrt{\Lambda}-\sqrt{\tilde{\Lambda}}\right)^TV^T\right) = \text{Tr}\left(\left(\sqrt{\Lambda}-\sqrt{\tilde{\Lambda}}\right)\left(\sqrt{\Lambda}-\sqrt{\tilde{\Lambda}}\right)^T\right)\end{aligned} \tag{67}$$

$$\boxed{||X-\tilde{X}||^2 = \sum_{i=\tilde{F}+1}^{F}\lambda_i} \tag{68}$$

□

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

$$\vec{u} = \text{argmax}_{||\vec{u}||=1}(X\vec{u})^2 \tag{69}$$

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

Доказательство. Сингулярное разложение:

$$X = V\sqrt{\Lambda}U^T, \quad \sqrt{\Lambda} = \text{diag}\{\lambda_1 \geq ... \geq \lambda_F\} \tag{70}$$

$$XU = V\sqrt{\Lambda} \to X\vec{u} = \sqrt{\lambda}\vec{v} \tag{71}$$

Как видно, максимум $(X\vec{u})^2$ достигается при максимальном $\lambda = \lambda_1$, этому сингулярному числу соотвветствует вектор $\vec{u}$, удовлетворяющий

$$\boxed{\vec{u} = \text{argmax}_{||\vec{u}||=1}(X\vec{u})^2} \tag{72}$$

$\square$

4. 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, задаёт прямую a (проходящую через начало координат), которая является решением следующей задачи оптимизации:

$$L' = \sum_{i=1}^{N} \text{distance}^2[(x_i, y_i); a] \to \min_a \tag{73}$$

где $\text{distance}[(x_i, y_i); a]$ – 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.

Доказательство. По теореме Пифагора:

$$\sum_{i=1}^{N} \text{distance}^2[(x_i, y_i); a] = \sum_{i=1}^{N}(r_i^2 - (\vec{r}_i, \vec{a})^2) \to \min_a \tag{74}$$

где $\vec{r}_i$ – радиус-вектор i-й точки. $\sum\limits_{i=1}^{N} r_i^2$ фиксирована и не зависит от направления прямой, поэтому

$$\sum_{i=1}^{N}(\vec{r}_i, \vec{a})^2 = (X\vec{a})^T(X\vec{a}) = (X\vec{a})^2 \to \max_a \tag{75}$$

где столбцы матрицы X – векторы $\vec{r}_i$. Из предыдущей задачи известно, что максимум достигается на сингулярном векторе $\vec{a}$, отвечающем наибольшему сингулярному числу. $\square$

5. Пусть дан набор из N точек в трёхмерном пространстве $X_{i\alpha}, i \in \{1, ..., N\}, \alpha \in \{1, 2, 3\}$. Покажите, что задача нахождения сингулярных чисел матрицы X эквивалентна нахождению *главных моментов инерции* твёрдого 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 жёсткими стержнями).

Доказательство. Распишем X:

$$X = \begin{pmatrix} x_1 & y_1 & z_1 \\ \dots & \dots & \dots \\ x_l & y_l & z_l \end{pmatrix} \tag{76}$$

$$X^{\mathrm{T}}X \equiv \hat{\mathrm{Cov}}X = \begin{pmatrix} \sum\limits_i x_i^2 & \sum\limits_{i,j} x_i y_j & \sum\limits_{i,j} x_i z_j \\ \sum\limits_{i,j} x_i y_j & \sum\limits_i y_i^2 & \sum\limits_{i,j} y_i z_j \\ \sum\limits_{i,j} x_i z_j & \sum\limits_{i,j} y_i z_j & \sum\limits_i z_i^2 \end{pmatrix} \tag{77}$$

Момент инерции

$$\hat{I} = \begin{pmatrix} \sum\limits_i y_i^2 + z_i^2 & -\sum\limits_{i,j} x_i y_j & -\sum\limits_{i,j} x_i z_j \\ -\sum\limits_{i,j} x_i y_j & \sum\limits_i x_i^2 + z_i^2 & -\sum\limits_{i,j} y_i z_j \\ -\sum\limits_{i,j} x_i z_j & -\sum\limits_{i,j} y_i z_j & \sum\limits_i x_i^2 + y_i^2 \end{pmatrix} \tag{78}$$

Матрицы $X^T X$ и I диагонализуемы совместно. В системе координат, где они обе диагональные, $\sum\limits_{i,j} x_i y_j = 0$. Собственные числа $X^T X$ (которые являются сингулярными числами X) и главные моменты инерции совпадают. $\square$