Прикладная статистика в машинном обучении
Семинар 2

И. К. Козлов
(*Мехмат МГУ*)

2022

Упражнения с прошлого занятия

1. Упражнения с прошлого занятия
2. Вопросы про ML
3. Псевдообратная матрица
4. Линейная регрессия

Семинар 2

Докажем смещённость выборочной дисперсии

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 .$$

Семинар 2

Докажем смещённость выборочной дисперсии

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.$$

Доказательство:

$$\mathbb{E}S^2 = \mathbb{E}\left(\frac{1}{n}X_i^2 - \frac{1}{n^2}\sum_{i,j}X_i, X_j\right) =$$

Семинар 2

Докажем смещённость выборочной дисперсии

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.$$

Доказательство:

$$\mathbb{E}S^2 = \mathbb{E}\left(\frac{1}{n}X_i^2 - \frac{1}{n^2}\sum_{i,j} X_i, X_j\right) =$$

$$= \frac{1}{n}\mathbb{E}X_i^2 - \frac{1}{n^2}\mathbb{E}X_i^2 - \frac{1}{n^2}\sum_{i \neq j}\mathbb{E}X_i\,\mathbb{E}X_j =$$

Семинар 2

Докажем смещённость выборочной дисперсии

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 .$$

Доказательство:

$$\mathbb{E}S^2 = \mathbb{E}\left(\frac{1}{n}X_i^2 - \frac{1}{n^2}\sum_{i,j}X_i, X_j\right) =$$

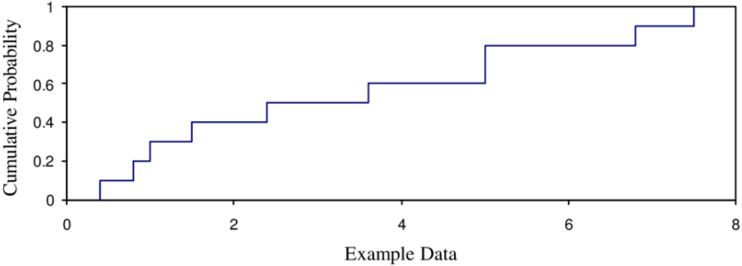
$$= \frac{1}{n}\mathbb{E}X_i^2 - \frac{1}{n^2}\mathbb{E}X_i^2 - \frac{1}{n^2}\sum_{i\neq j}\mathbb{E}X_i\,\mathbb{E}X_j =$$

$$\frac{1}{n}\mathbb{E}X_1^2 - \frac{1}{n}\mathbb{E}X_i^2 - \frac{n-1}{n}\left(\mathbb{E}X_1\right)^2 = \frac{n-1}{n}\mathbb{V}X_1 .$$

Эмпирическая функция распределений

Эмпирическая функция распределения

$$\hat{F}(x) = \hat{F}_n(x) = \frac{\text{кол-во эл-тов в выборке} \leq \text{x}}{n} = \frac{1}{n}\sum_{i=1}^{n} I\left(X_i \leq x\right).$$



Семинар 2

Утверждение:

$$\mathbb{E}\left(\hat{F}_n(x)\right) = F(x),$$

Утверждение:

$$\mathbb{E}\left(\hat{F}_n(x)\right) = F(x),$$

Доказательство: Нужно вспомнить стандартные формулы.

- Формула для эмпирической функции распределения

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I\left(X_i \leq x\right).$$

Семинар 2

- Вспоминаем формулу для матожидания

$$\mathbb{E}\left(a(x)\right) = \int a(x)dF(x) = \int a(x)f(x)dx$$

Семинар 2

- Вспоминаем формулу для матожидания

$$\mathbb{E}\left(a(x)\right) = \int a(x)dF(x) = \int a(x)f(x)dx$$

Используем линейность матожидания:

$$\mathbb{E}\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\, I\left(X_i \le x\right) = \mathbb{E}\, I\left(X_1 \le x\right).$$

Семинар 2

- Вспоминаем формулу для матожидания

$$\mathbb{E}\left(a(x)\right) = \int a(x)dF(x) = \int a(x)f(x)dx$$

Используем линейность матожидания:

$$\mathbb{E}\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\, I\left(X_i \le x\right) = \mathbb{E}\, I\left(X_1 \le x\right).$$

Считаем матожидание:

$$\mathbb{E}\, I\left(X \le x\right) = \int_{-\infty}^{\infty} I(z \le x)f(z)dx = \int_{-\infty}^{x} f(z)dx$$

Семинар 2

- Осталось вспомнить формулу для функции распределения

$$F(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f(z)dz$$

Утверждение доказано.

Семинар 2

Утверждение:

$$\mathbb{V}\left(\hat{F}_n(x)\right) = \frac{F(x)\left(1 - F(x)\right)}{n}$$

Семинар 2

Утверждение:

$$\mathbb{V}\left(\hat{F}_n(x)\right) = \frac{F(x)\,(1 - F(x))}{n}$$

Доказательство.

- Вспоминаем формулу дисперсии
$$\mathbb{V}(X) = \mathbb{E}\,(X)^2 - (\mathbb{E}X)^2\,.$$

- Мы знаем, что $\mathbb{E}\left(\hat{F}_n(x)\right) = F(x)$, осталось посчитать $\mathbb{E}\left(\hat{F}_n(x)\right)^2$.

Семинар 2

- Подставляем формулу для эмпирической функции распределения

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I\left(X_i \le x\right).$$

Для краткости обозначим

$$I_i = I\left(X_i \le x\right)$$

Получаем

$$\mathbb{E}\left(\hat{F}_n(x)\right)^2 = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} I_i\right)^2 = \frac{1}{n^2}\left(\sum_{i=1}^{n}\mathbb{E}I_i^2 + \sum_{i\neq j}\mathbb{E}\left(I_i I_j\right)\right).$$

Семинар 2

- В 1ом слагаемом $I_i^2 = I_i$, т.к. это индикатор:

$$\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}I_i^2 = \frac{1}{n^2}\sum\mathbb{E}I_i = \frac{1}{n}F(x).$$

Семинар 2

Во 2ом слагаемом $\sum_{i \neq j} \mathbb{E}\left(I_i I_j\right)$ вспоминаем 2 факта:

- (Борелевские) функции от независимых случайных величин независимы.
- Если X и Y — независимы, то
$$\mathbb{E}(XY) = \mathbb{E}X \, \mathbb{E}Y.$$

Семинар 2

Во 2ом слагаемом $\sum_{i \neq j} \mathbb{E}\left(I_i I_j\right)$ вспоминаем 2 факта:

- (Борелевские) функции от независимых случайных величин независимы.
- Если X и Y — независимы, то
$$\mathbb{E}(XY) = \mathbb{E}X\,\mathbb{E}Y.$$

Поэтому

$$\frac{1}{n^2}\sum_{i \neq j} \mathbb{E}\left(I_i I_j\right) = \frac{1}{n^2}\mathbb{E}I_i\,\mathbb{E}I_j = \frac{n^2 - n}{n^2}F(x)^2.$$

Подставляем все вычисленные значения в формулу для дисперсии.

Получаем требуемый ответ:

$$\mathbb{V}\left(\hat{F}_n(x)\right) = \frac{1}{n}F(x) + \frac{n^2 - n}{n^2}F(x)^2 - F(x)^2 =$$

$$= \frac{F(x)}{n} - \frac{F(x)^2}{n} = \frac{F(x)\,(1 - F(x))}{n}$$

Логистическая регрессия

Logloss = Cross Entropy:

$$L_{\log}(y, p) = -\sum_i \left(y_i \log p_i + (1 - y_i) \log(1 - p_i)\right)$$

- $y_i \in \{0, 1\}$ — истинные метки классов,
- $p_i = \mathbb{P}(y_i = 1)$ — вероятностные предсказания модели,

Q: Откуда взялся этот лосс?

Логлосс

A: Вспомним функцию правдоподобия для $\mathrm{Bernoulli}\,(p)$:

$$\mathcal{L} = p^x (1-p)^{1-x}.$$

Логарифм правдоподобия даёт минус логлосс:

$$\ell = x \log p + (1-x)\log(1-p).$$

Для сравнения логлосс:

$$L_{\log}(y, p) = -\sum_i \left(y_i \log p_i + (1-y_i)\log(1-p_i)\right).$$

SVD-разложение:

$$A = U \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} V^{T}$$

A	U	Σ	V^T
Любая матрица m x n	Ортогональная m x m	Диагональная m x n	Ортогональная n x n

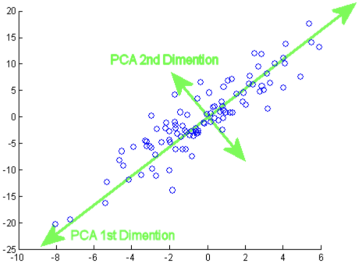
$$\sigma_1 \geq \cdots \geq \sigma_r > 0$$

Напомним, матрица U ортогональная, если $U^T U = E$.

Сингулярное разложение

Q: Для чего в ML может применять сингулярное разложение? Кроме линейной/логистической регрессии

A: *Уменьшение размерности*. См. Метод главных компонент **PCA**



Сингулярное разложение

Пусть SVD-разложение:

$$A = U\Sigma V^T,$$

$$\Sigma = (\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)\,.$$

Если u_i — столбцы U, а v_i — строки V^T, то

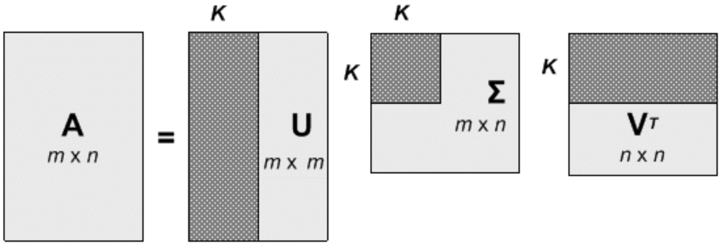
$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T\,.$$

Сингулярное разложение

В диагональной матрице можно обнулить последние коэффициенты.

$$\sigma_{k+1} \to 0, \qquad \dots \quad \sigma_r \to 0$$

Получится экономия по памяти.



Псевдообратная матрица

1. Упражнения с прошлого занятия
2. Вопросы про ML
3. Псевдообратная матрица
4. Линейная регрессия

Псевдообратная матрица

Формулы для псевдообратной матрицы A^+:

- Если столбцы A линейно независимы.

$$A^+ = (A^T A)^{-1} A^T.$$

- Если строчки A линейно независимы:

$$A^+ = A^T (AA^T)^{-1}.$$

Логистическая регрессия

Q: Почему

$$\operatorname{rk} A = \operatorname{rk} A^T A?$$

Что такое матрица $A^T A$?

Логистическая регрессия

Q: Почему
$$\operatorname{rk} A = \operatorname{rk} A^T A?$$
Что такое матрица $A^T A$?

A: $A^T A$ — это матрица Грамма для столбцов.

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a^2 + c^2 & ac + bd \\ ac + bd & b^2 + d^2 \end{pmatrix}$$

Логистическая регрессия

Q: Почему

$$\operatorname{rk} A = \operatorname{rk} A^T A?$$

Что такое матрица $A^T A$?

A: $A^T A$ — это матрица Грамма для столбцов.

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a^2 + c^2 & ac + bd \\ ac + bd & b^2 + d^2 \end{pmatrix}$$

Вспоминаем линейную алгебру:

Ранг матрицы Грамма векторов $v_1, \ldots, v_N$ = размерности линейного подпространства, порождённого векторами v_i.

Логистическая регрессия

Q: Найти

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \dots \\ x_n & 1 \end{pmatrix}^+$$

Логистическая регрессия

Q: Найти

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \dots \\ x_n & 1 \end{pmatrix}^+$$

A: Применяем формулу

$$A^+ = (A^T A)^{-1} A^T.$$

Получаем

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \dots \\ x_n & 1 \end{pmatrix}^+ = \begin{pmatrix} n\overline{x^2} & n\bar{x} \\ n\bar{x} & n \end{pmatrix}^{-1} \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ 1 & 1 & \dots & 1 \end{pmatrix} =$$

Логистическая регрессия

Q: Найти

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \dots \\ x_n & 1 \end{pmatrix}^+$$

A: Применяем формулу

$$A^+ = (A^T A)^{-1} A^T.$$

Получаем

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \dots \\ x_n & 1 \end{pmatrix}^+ = \begin{pmatrix} n\overline{x^2} & n\bar{x} \\ n\bar{x} & n \end{pmatrix}^{-1} \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ 1 & 1 & \dots & 1 \end{pmatrix} =$$

$$\frac{1}{n} \frac{1}{\overline{x^2} - \bar{x}^2} \begin{pmatrix} 1 & -\bar{x} \\ -\bar{x} & \overline{x^2} \end{pmatrix} \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ 1 & 1 & \dots & 1 \end{pmatrix} = \frac{1}{n} \frac{1}{\overline{x^2} - \bar{x}^2} \begin{pmatrix} x_1 - \bar{x} & x_2 - \bar{x} & \dots & x_n - \bar{x} \\ \overline{x^2} - x_1\bar{x} & \overline{x^2} - x_2\bar{x} & \dots & \overline{x^2} - x_n\bar{x} \end{pmatrix}.$$

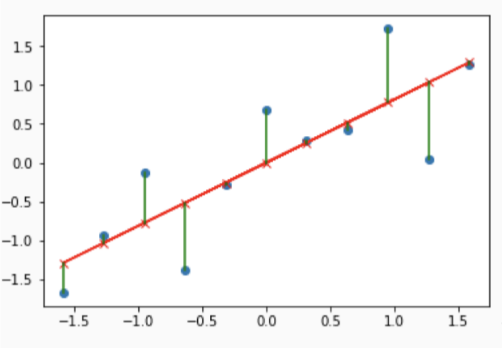
Линейная регрессия

Семинар 2

Простейшая линейная регрессия.

Пусть даны точки (x_i, y_i), где $i = 1, \ldots, N$. Аппроксимируем решение линейной функцией

$$y_i = kx_i + b$$



Семинар 2

Задача. Найти минимум MSE:

$$J = \frac{1}{N}\sum_i (y_i - (kx_i + b))^2.$$

Решение. Сумма выпуклых функций — выпуклая функция. Поэтому у лосса существует минимум, и его можно найти как решение

$$\frac{\partial J}{\partial k} = 0, \qquad \frac{\partial J}{\partial b} = 0$$

Семинар 2

Вычисляем производные:

$$\frac{\partial J}{\partial k} = \frac{1}{N}\sum_i (y_i - (kx_i + b))(-x_i),$$
$$\frac{\partial J}{\partial b} == \frac{1}{N}\sum_i (y_i - (kx_i + b))(-1)$$

Из 2го уравнения находим

$$b = \bar{y} - k\bar{x}$$

Семинар 2

Вычисляем производные:

$$
\frac{\partial J}{\partial k} = \frac{1}{N}\sum_i \left(y_i - (kx_i + b)\right)(-x_i),
$$
$$
\frac{\partial J}{\partial b} == \frac{1}{N}\sum_i \left(y_i - (kx_i + b)\right)(-1)
$$

Из 2го уравнения находим

$$
b = \bar{y} - k\bar{x}
$$

Подставляем в 1ое уравнение

$$
0 = -\overline{xy} + k\overline{x^2} - b\bar{x} = -\overline{xy} + \bar{x}\bar{y} + k(\overline{x^2} - \bar{x}^2)
$$

Получаем **ответ**

$$
k = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}, \qquad b = \bar{y} - k\bar{x}.
$$

Семинар 2

Задача. Доказать, что это решение — это псевдорешение

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \dots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} k \\ b \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}.$$

Семинар 2

Псевдорешение:

$$\begin{pmatrix} k \\ b \end{pmatrix} = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \dots & \dots \\ x_n & 1 \end{pmatrix}^{+} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} =$$

$$= \frac{1}{n} \frac{1}{\overline{x^2} - \bar{x}^2} \begin{pmatrix} 1 & -\bar{x} \\ -\bar{x} & \overline{x^2} \end{pmatrix} \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ 1 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} =$$

$$\frac{1}{n} \frac{1}{\overline{x^2} - \bar{x}^2} \begin{pmatrix} 1 & -\bar{x} \\ -\bar{x} & \overline{x^2} \end{pmatrix} \begin{pmatrix} n\overline{xy} \\ n\bar{y} \end{pmatrix}.$$

Легко видеть, что

$$k = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}$$

Равенство $b = \bar{y} - k\bar{x}$ оставим как упражнение.