

Пусть (x_1, x_2, \dots, x_n) последовательность векторов:

$$x_i = \begin{pmatrix} x_i^1 \\ \vdots \\ x_i^m \end{pmatrix}, \quad i = 1, \dots, n.$$

В статистике часто приходится вычислять вектор **выборочного среднего**

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

и **выборочную ковариационную матрицу**

$$\bar{V} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T,$$

где x^T - транспонирование столбца x .

1. Какую **каноническую** форму информации, Вы бы предложить, чтобы представить последовательность (x_1, x_2, \dots, x_n) для того, чтобы вычислить вектор выборочного среднего и выборочную ковариационную матрицу?

Убедитесь в том, что все “желательные” свойства канонической информации выполнены:

- (a) **Существование и единственность.** Любая последовательность исходных данных (x_1, x_2, \dots, x_n) может быть представлена в канонической форме единственным образом. Зависит ли это представление от порядка векторов x_i в последовательности данных?
- (b) **Элементарная** каноническая информация. Существует ли каноническая информация для одного наблюдения?
- (c) **Пустая** каноническая информация. Существует ли каноническая информация для пустой последовательности наблюдений?
- (d) **Объединение** (комбинация, композиция). Как бы Вы определили объединение информации в канонической форме? Выполняются ли аксиомы коммутативного моноида? (Коммутативность, ассоциативность, нейтральный элемент)
- (e) **Обновление.** Как обновляется каноническая информация, когда поступает новый вектор наблюдений ?
- (f) **Полнота.** Каноническая информация должна сохранять всю информацию, которая присутствовала в исходных необработанных данных. В частности, вычисления, использующие каноническую информацию должны давать те же результаты, что и оригинальный алгоритм примененный к исходным необработанным данным. Удовлетворяет ли Ваше каноническое представление данных этому требованию? Как вычислить \bar{x} и \bar{V} используя только собранную канонической информацией?
- (g) **Компактность и эффективность.** Что Вы можете сказать о компактности (или минимальности) и эффективности вашей канонической формы информации с точки зрения требований к хранению и сложности объединения, обновления и использования?
- (h) Каково **минимальное число** наблюдений n для которого \bar{x} и \bar{V} определены?

2. (**Необязательно**) Какую **явную** форму информации Вы могли бы предложить, чтобы представить последовательность векторов (x_1, x_2, \dots, x_n) ? Это представление должно содержать \bar{x} и \bar{V} и, возможно, что-то еще.

Проанализируйте все вопросы задачи 1 для предложенной явной информации.