

Задача

Построить рейтинг на основе предоставленных признаков и экспертных оценок объектов. Предложить метод оценки критической массы экспертных мнений, то есть метод принятия решения о достаточности текущего числа экспертов.

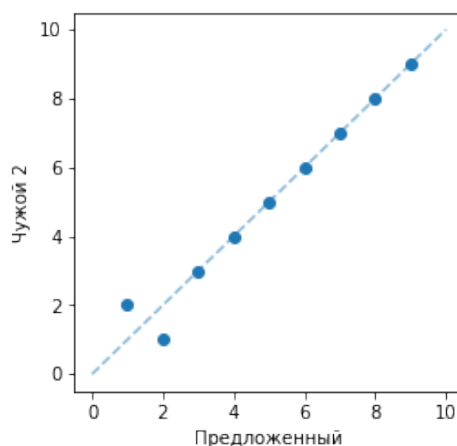
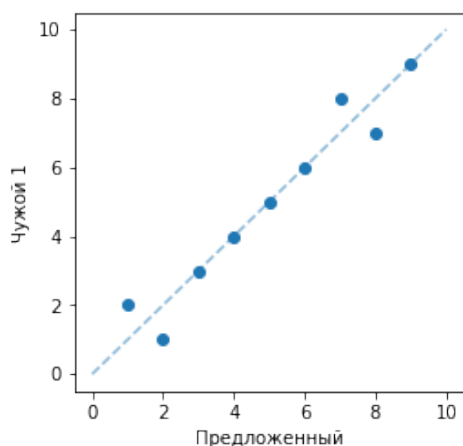
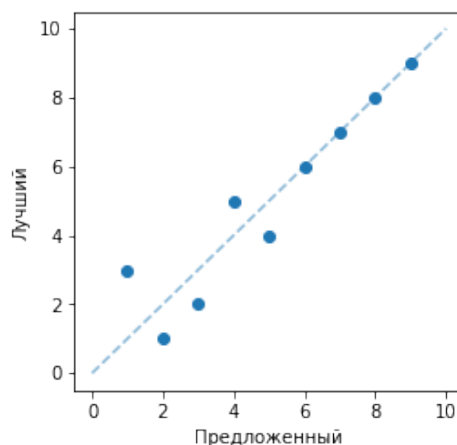
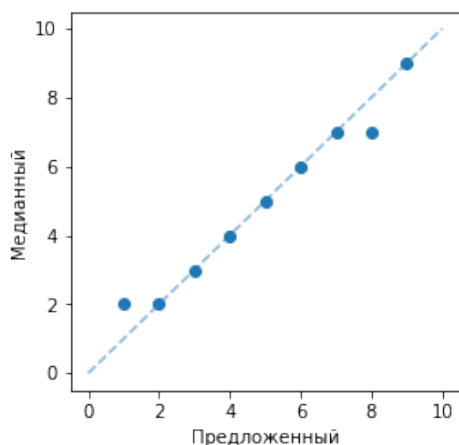
Предлагаемое решение

В качестве основной модели построения рейтинга используется метод попарных сравнений Кемени-Янга. В решении рассматриваются исключительно экспертные оценки, признаки объектов не учитываются при построении рейтинга.

Метод попарных сравнений имеет стандартный вид: вычисление матрицы Талли для каждого эксперта, сложение матриц, вычисление статистики, на основе которой производится сортировка всевозможных перестановок объектов. В виду того, что размерность модельной выборки мала, решение искалось полным перебором: статистика вычислялась для каждой из возможных перестановок.

1. Шоколад
2. Сырок «РОСТАГРОЭКСПОРТ»
3. Гречневая крупа
4. Арахис соленый
5. Молоко 2,5%
6. Макароны
7. Овсяное молоко 3.2%'
8. Кефир 3.2%
9. Устрицы

Сравнение рейтингов



Задача определения достаточного числа экспертов решалась в «онлайн» режиме, то есть по мере поступления новых экспертов. Метод основывается на построении зависимости числа смен «лидера» в алгоритме Кемени-Янга от числа экспертов. «Лидером» называем перестановку, соответствующую полученному рейтингу, то есть с наибольшим значением статистики.

При допущении «рациональности» экспертов, то есть их сэмплировании из некоторого фиксированного распределения, реализации статистики (среднее значение по экспертам), используемой при сортировке в алгоритме Кемени-Янга, будет сходиться к среднему значению при увеличении числа экспертов.

Предлагается считать, что число экспертов N достаточно, если при добавлении $N+1$ эксперта «лидер» в алгоритме Кемени-Янга не менялся на протяжении k предыдущих добавлений экспертов. Параметр k является гиперпараметром модели.

Определение достаточного числа экспертов

Рейтинг, предоставленный экспертом, $\mathbf{r}_i = \text{argsort}(z_1, \dots, z_5)^T$, где $z_j \sim N(j, 4)$, то есть argsort от вектора, сэмплированного из многомерного нормального, с диагональной матрицей ковариации.

Случайные пропуски. Матрицы экспертов $\mathbf{R} = (\mathbf{r}_1, \dots, \mathbf{r}_K)$. Рассматриваем «временные» срезы вида $\mathbf{R}_N = (\mathbf{r}_1, \dots, \mathbf{r}_N)$, $N \leq K$.

Всю процедуру повторяем 5 раз, меняя сиды генератора случайных чисел.

