**НИУ ВШЭ в Нижнем Новгороде. Факультет информатики, математики и компьютерных наук. Дисциплина «Методы анализа данных».**

Charge de cours: Калягин В.А.

**Снижение размерности. Метод главных компонент** (PCA=Principal Component Analysis).

Банком проведено исследование с целью понять, следует ли разрабатывать специальные программы маркетинга для нескольких ключевых сегментов. Один из вопросов исследования касался отношения к банковским операциям. С использованием шкалы «согласен - несогласен» (от 0 до 9) респонденты оценивали следующие утверждения:

1. Мелкие банки взимают за свои услуги меньше, чем крупные

2. Крупные банки делают ошибки чаще, чем мелкие

3. Не обязательно, чтобы банковские служащие были подчеркнуто вежливыми и дружелюбными; достаточно, чтобы они вели себя воспитанно.

4. Я хочу, чтобы в моем банке меня знали лично и чтобы ко мне относились с особой вежливостью

5. Если в финансовом институте я встречу холодное или небрежное отношение, я перестану быть клиентом этой организации.

**Данные:** ответы представлены в виде матрицы. Каждая строка матрицы содержит ответы одного из респондентов.

9 6 9 2 2

4 6 2 6 7

0 0 5 0 0

2 2 0 9 9

6 9 8 3 3

3 8 5 4 7

4 5 6 3 6

8 6 8 2 2

4 4 0 8 8

2 8 4 5 7

1 2 6 0 0

6 9 7 3 5

6 7 1 7 8

2 1 7 1 1

9 7 9 2 1

**Вопрос:** какую информацию мы можем извлечь из этих данных? Что управляет ответами респондентов, Можно ли утверждать, что существует небольшое число (один или два) скрытых признаков (факторов), которые определяют ответы респондентов? Как их найти и какой они имеют смысл? Выделение таких признаков может дать наглядное представление о клиентах (карта клиентов).

**Проблема:** формализовать задачу (математическая постановка) и найти алгоритм решения.

**Подход к решению:** задача аппроксимации заданной матрицы матрицей меньшего ранга, SVD разложение (singular value decomposition), теорема Эккарта-Юнга.