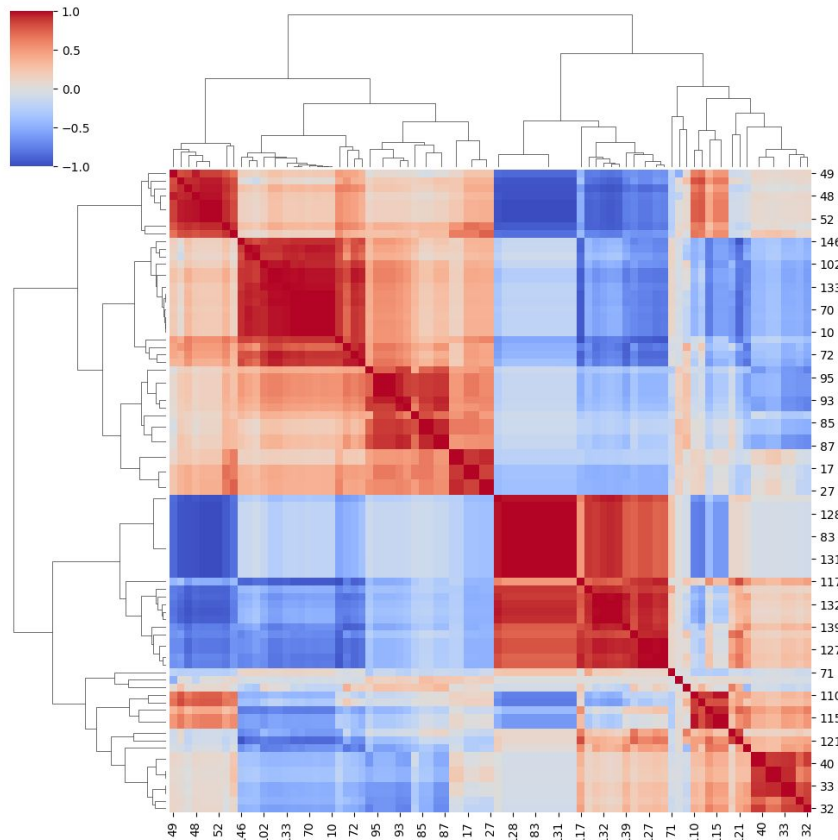


# Как манипулировать пространством, чтобы повышать качество работы методов машинного обучения

Устюгов В.Н., Институт математики им. С.Л. Соболева СО РАН (Омский филиал)

Исследование выполнено за счет гранта Российского научного фонда № 22-71-10015-П, <https://rscf.ru/en/project/22-71-10015-П/>

# Зачем?



Ограничения:

- Threads строго 8;
- Minrelnodes строго 0.

Параметры выбранные для настройки:

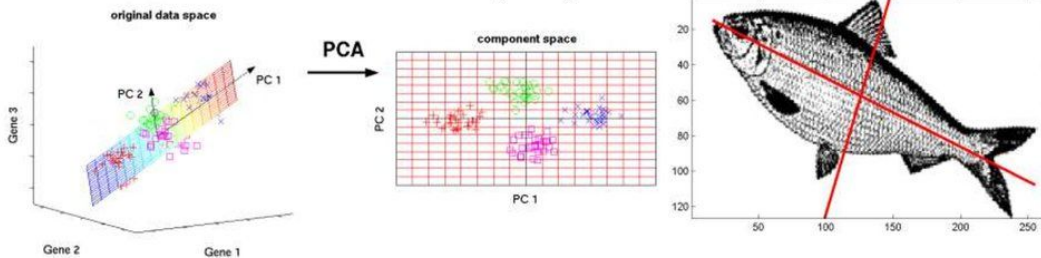
- Gomorypasses;
- Method;
- Mipfocus;
- Presolve.

# Методы

- Principal component analysis (PCA)
- Vector Quantization (VQ)

# Principal component analysis

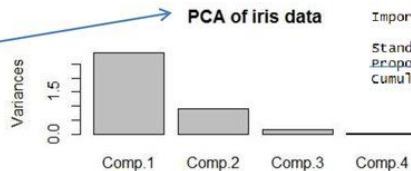
## Методы понижения размерности: анализ главных компонент (PCA)



Идея. Каждый объект – точка в  $n$ -мерном Евклидовом пространстве, весь массив данных – облако точек. Требуется найти новые оси, которые будут наилучшим образом объяснять изменчивость.

1я главная компонента – прямая, секущая облако в направлении его максимальной изменчивости (линия регрессии, по сути). 2я главная компонента перпендикулярна 1й в наиболее «широком» месте.

Служебный  
график осыпи  
(scree plot)



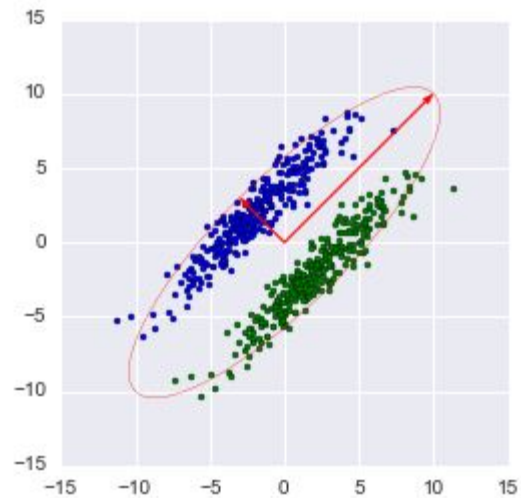
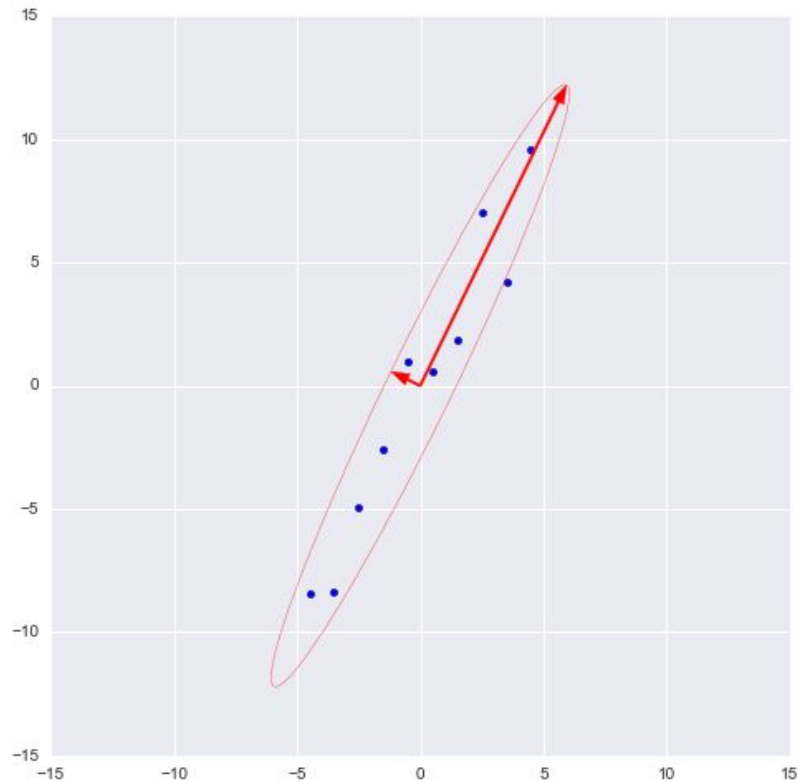
PCA of iris data

Importance of components:

	Comp.1	Comp.2	Comp.3	Comp.4
Standard deviation	1.7026571	0.9528572	0.38180950	0.143445939
Proportion of variance	0.7296245	0.2285076	0.03668922	0.005178709
Cumulative Proportion	0.7296245	0.9581321	0.99482129	1.000000000

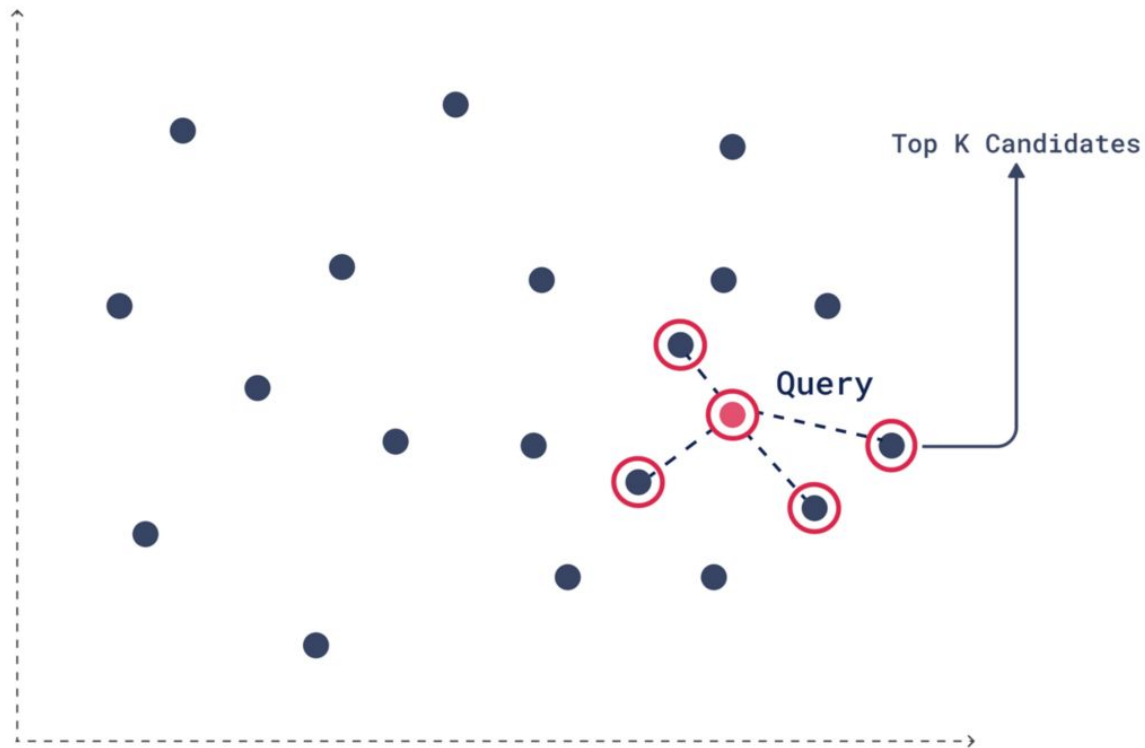
Доля объяснённой дисперсии. Первые 2 гл.компоненты объясняют почти 96% дисперсии!

# Principal component analysis



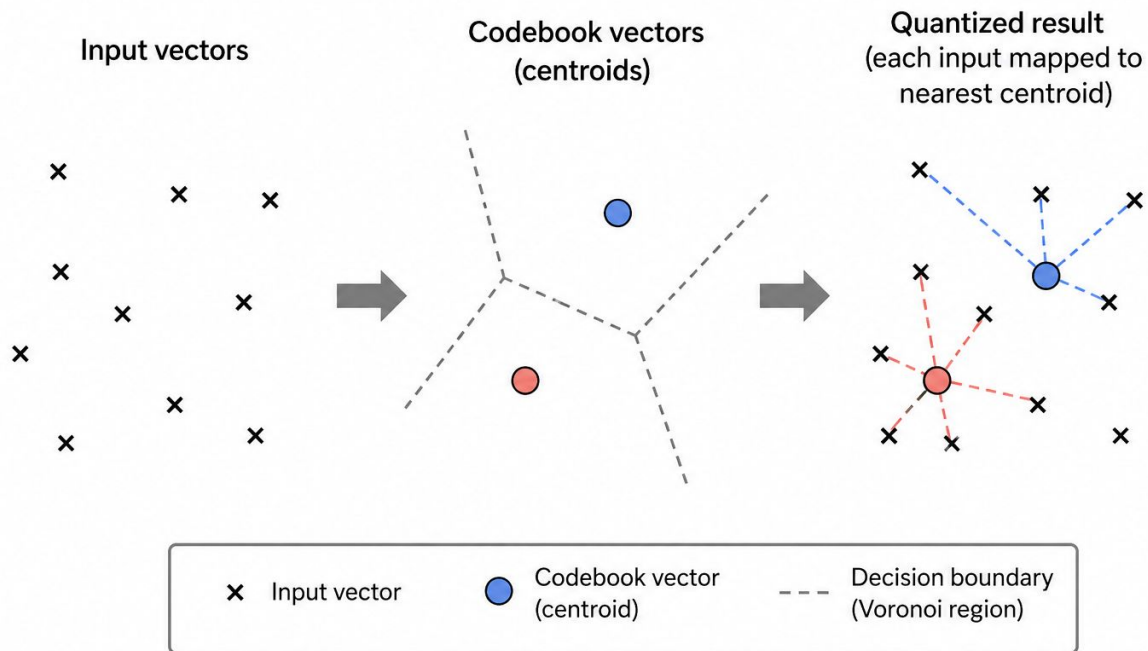
# Vector quantization

Nearest Neighbors Search (Quantized)



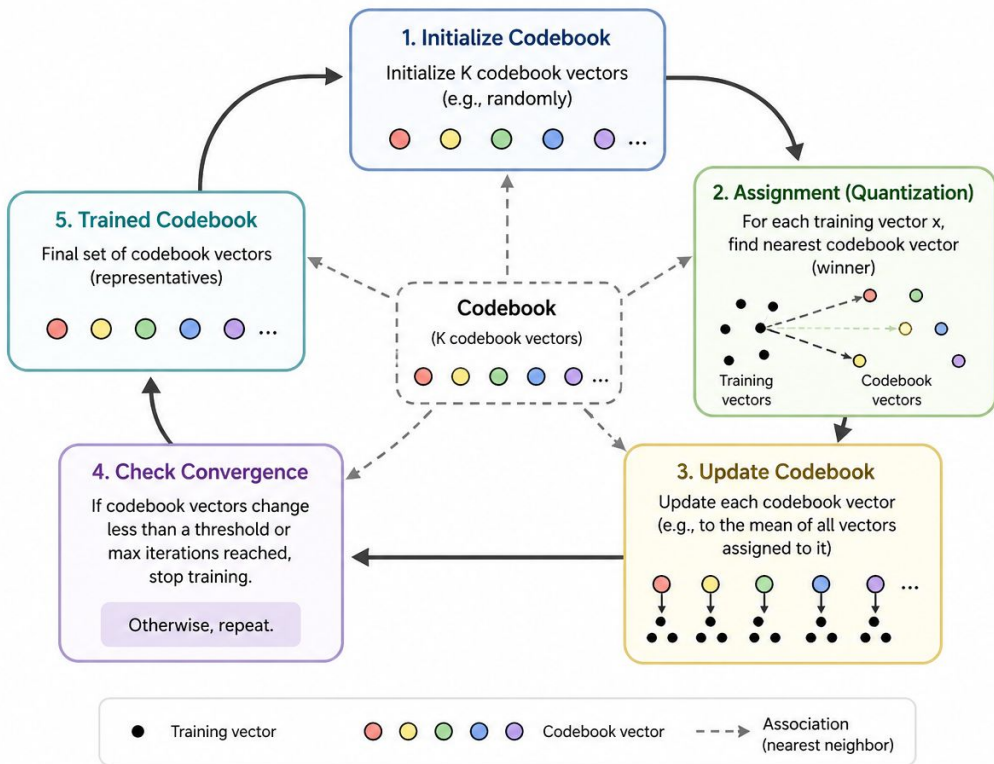
# Vector quantization

## Vector Quantization

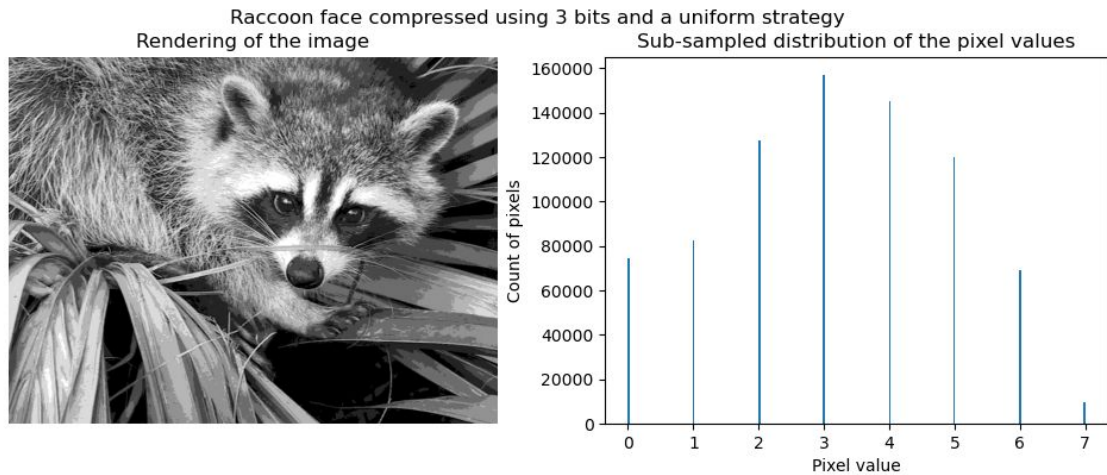
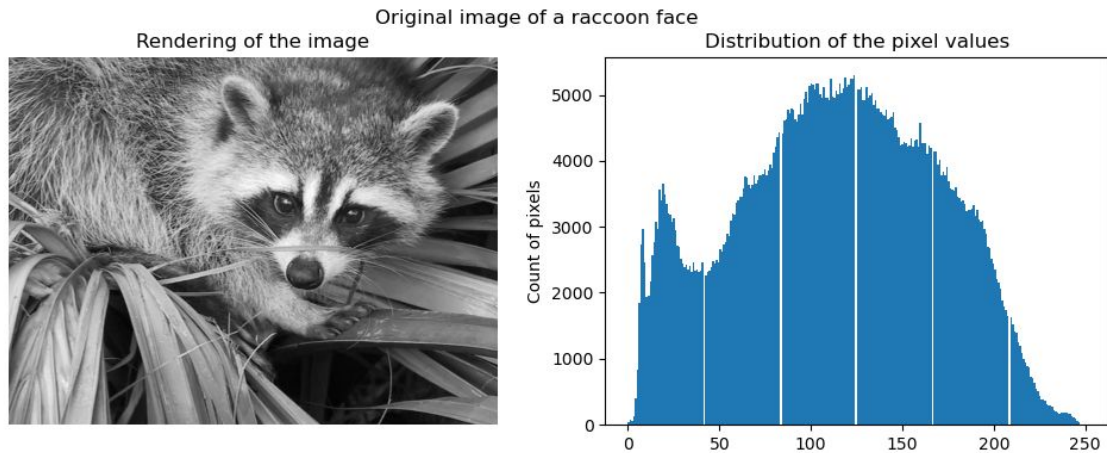


# Vector quantization

## Training Cycle (Vector Quantization)

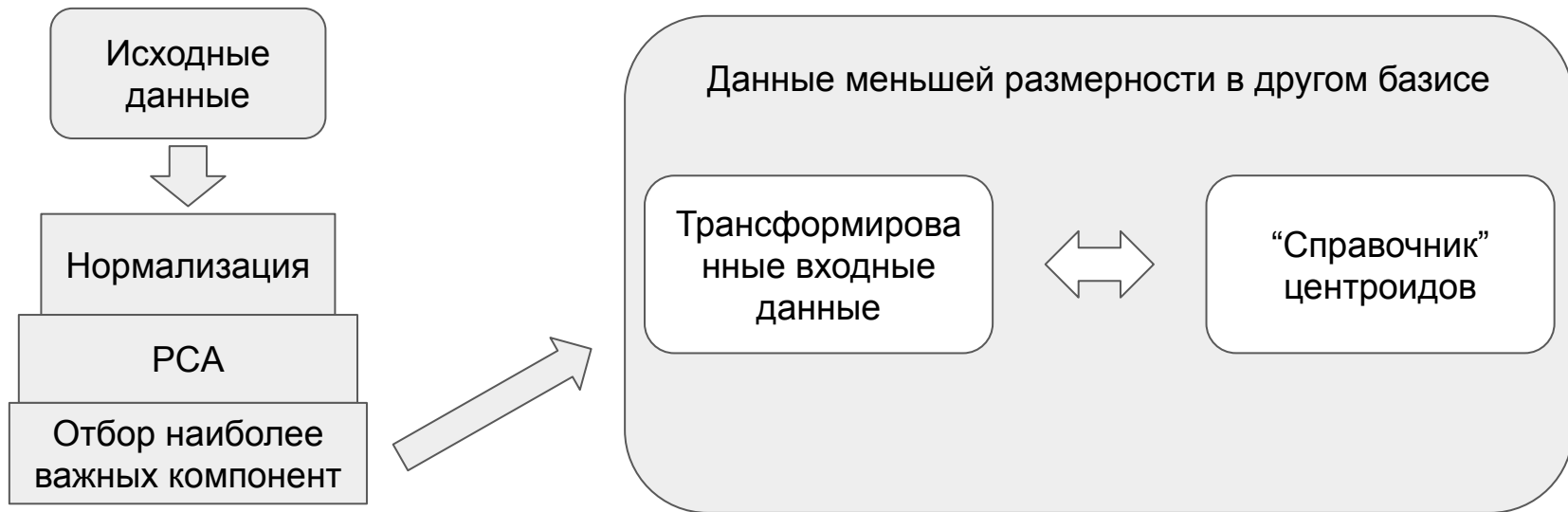


# Vector quantization





# Итого



Спасибо за внимание!