Оценка энтропии многоразмерного датасета

Васин Александр
Хомутова Ксения

Актуальность исследования

Оценка энтропии является ключевой проблемой во многих прикладных задачах.

Большое число параметров

↓

"Проклятие размерности"

Литература

Kirthevasan Kandasamy, Akshay Krishnamurthy, Barnabas Poczos, Larry Wasserman, James M. Robins, Influence Functions for Machine Learning: Nonparametric Estimators for Entropies, Divergences and Mutual Informations //arXiv:1411.4342

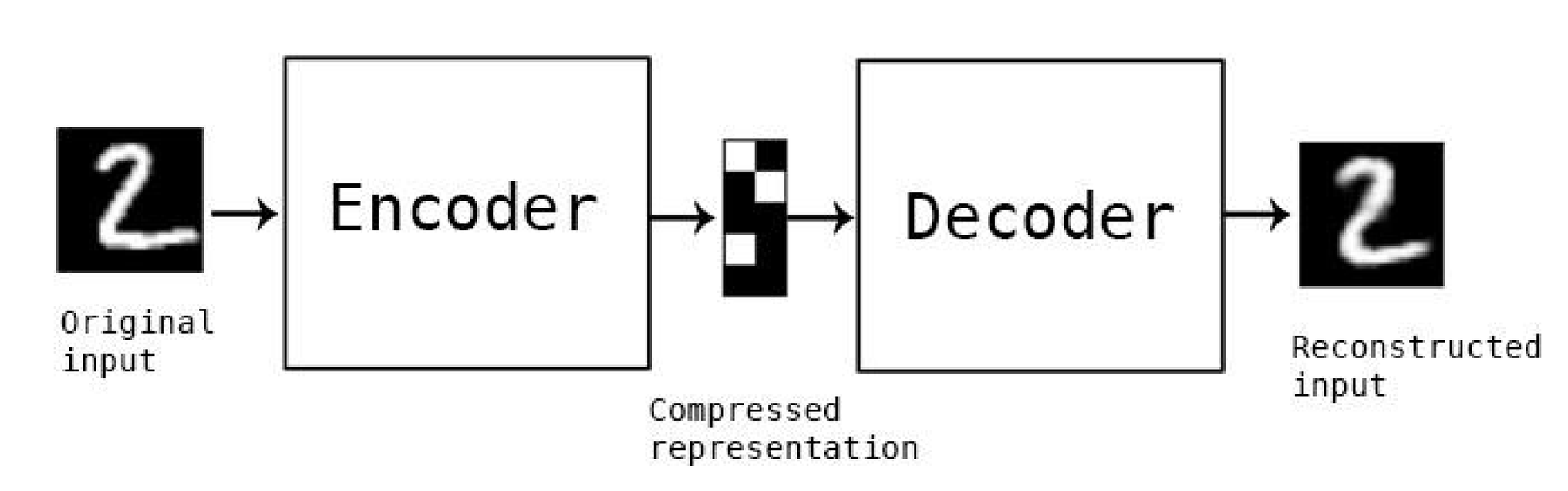
Kirthevasan Kandasamy, Akshay Krishnamurthy, Barnabas Poczos, Larry Wasserman, James M. Robins, Nonparametric von Mises Estimators for Entropies, Divergences and Mutual Informations //Advances in Neural Information Processing Systems 28 (NIPS 2015)

Цель и выполненные задачи исследования

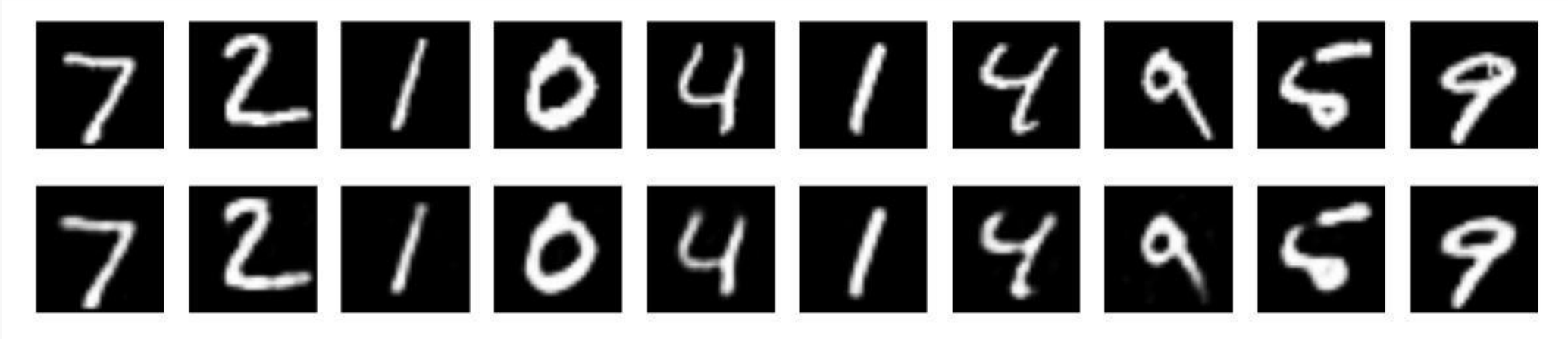
Оценка энтропии многоразмерного датасета.

- Построение и обучение Convolutional Neural Network с автоэнкодером на датасете MNIST;
- Изменение количества нейронов в скрытом слое автоэнкодера;
- Выявление зависимости между количеством нейронов в скрытом слое и качеством восстанавливаемого изображения;
- Применение Kernel Density Estimation для оценки ядерной плотности найденного представления;
- Подсчет энтропии для представленных сжатий.

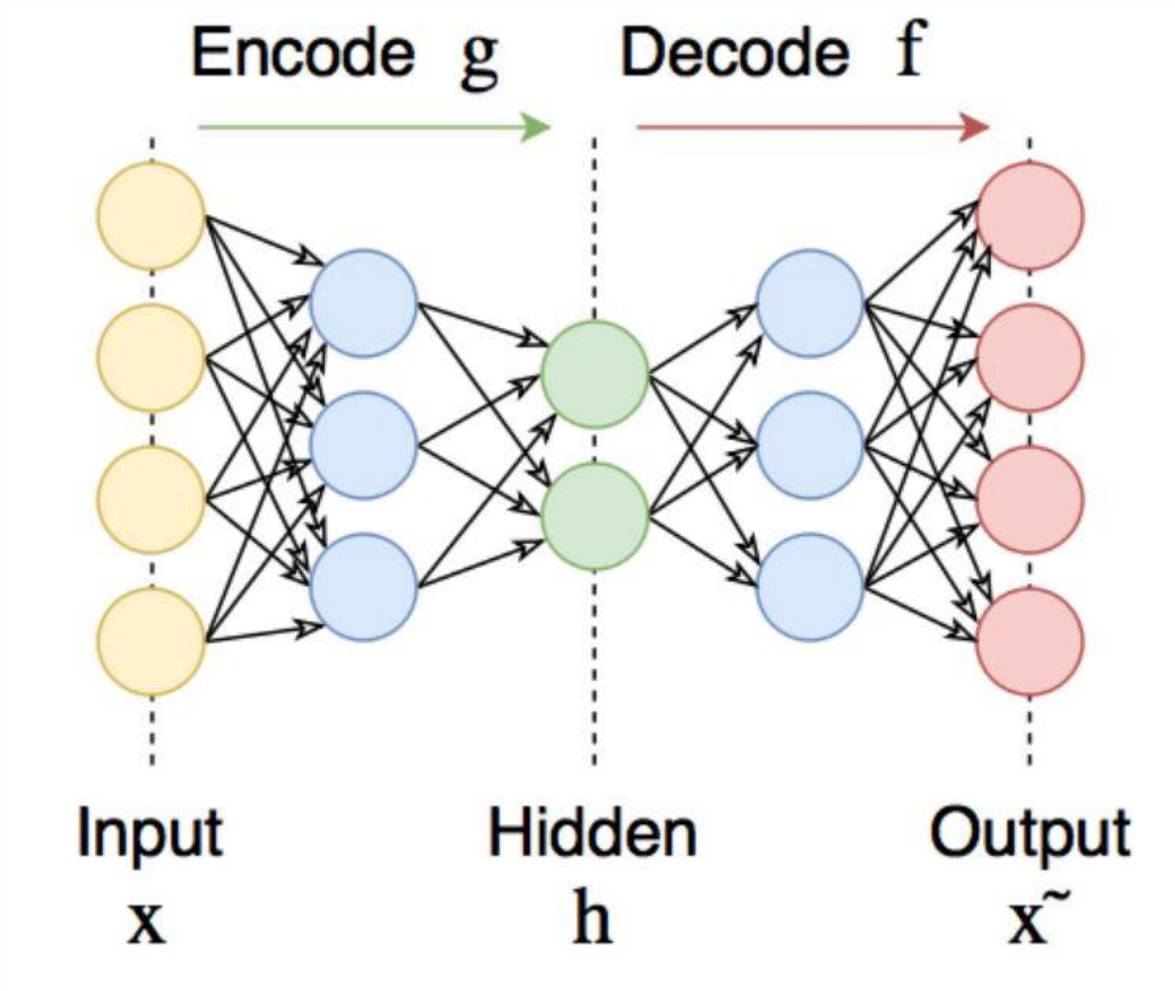
1 этап


Encoder
Decoder
Original input
Compressed representation
Reconstructed input

1 этап



2 этап



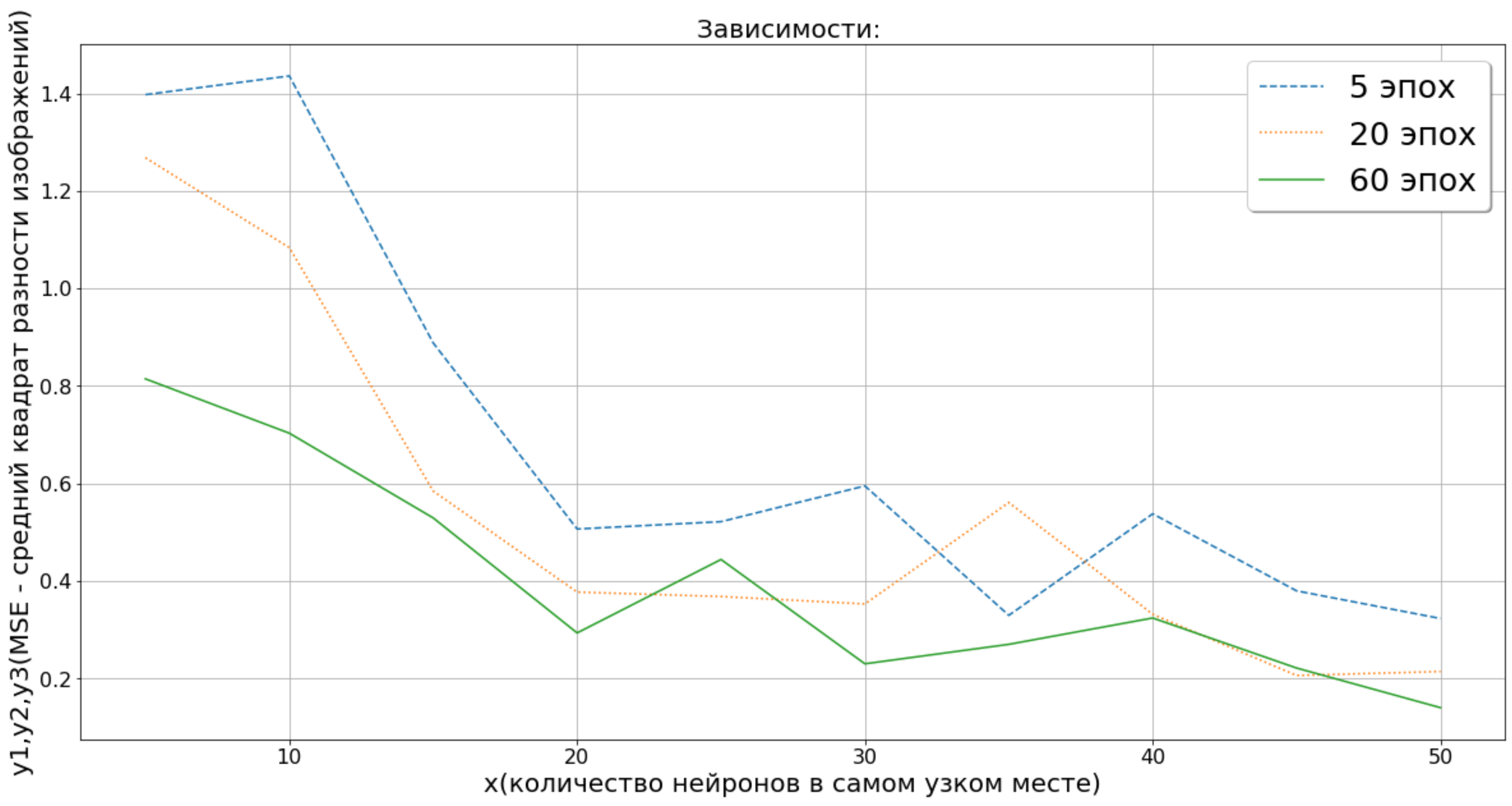
2 этап

5 нейронов в самом узком месте автоэнкодера - 5 эпох

20 нейронов в самом узком месте автоэнкодера - 20 эпох

50 нейронов в самом узком месте автоэнкодера - 60 эпох

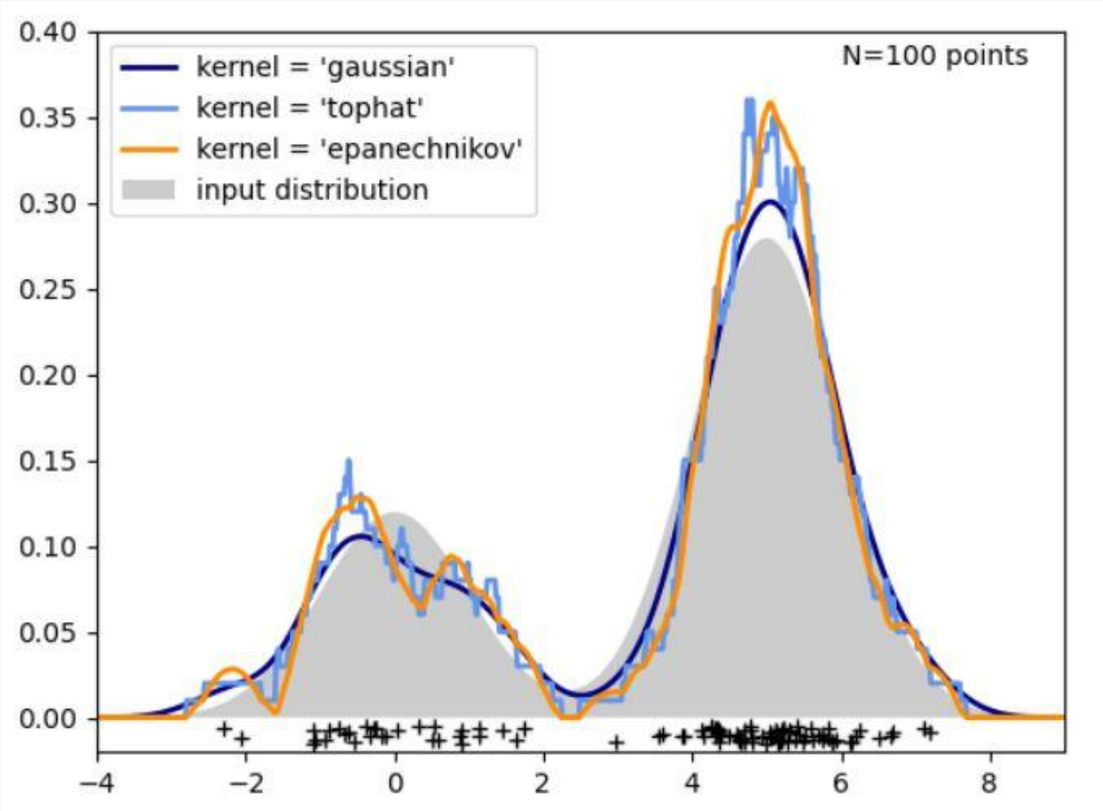
3 этап

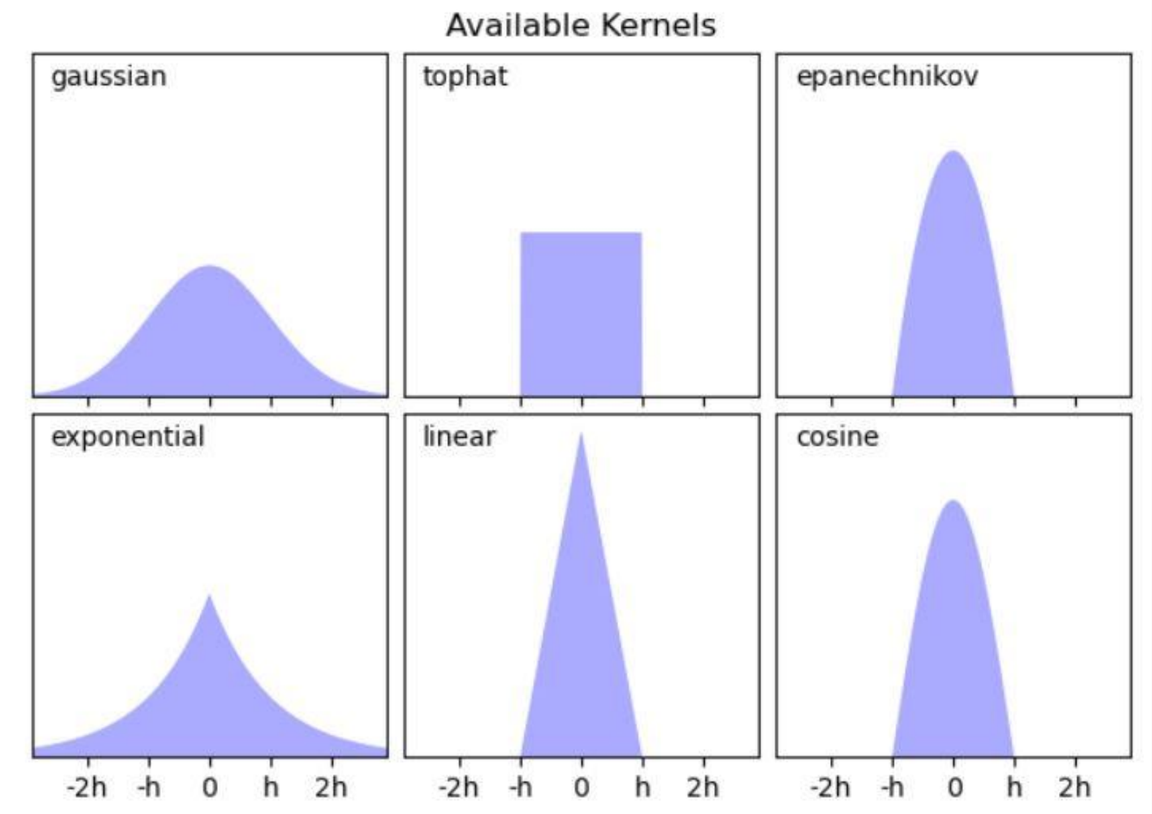

Зависимости:
5 эпох
20 эпох
60 эпох
y1,y2,y3(MSE - средний квадрат разности изображений)
1.4
1.2
1.0
0.8
0.6
0.4
0.2
10
20
30
40
50
x(количество нейронов в самом узком месте)

4 этап

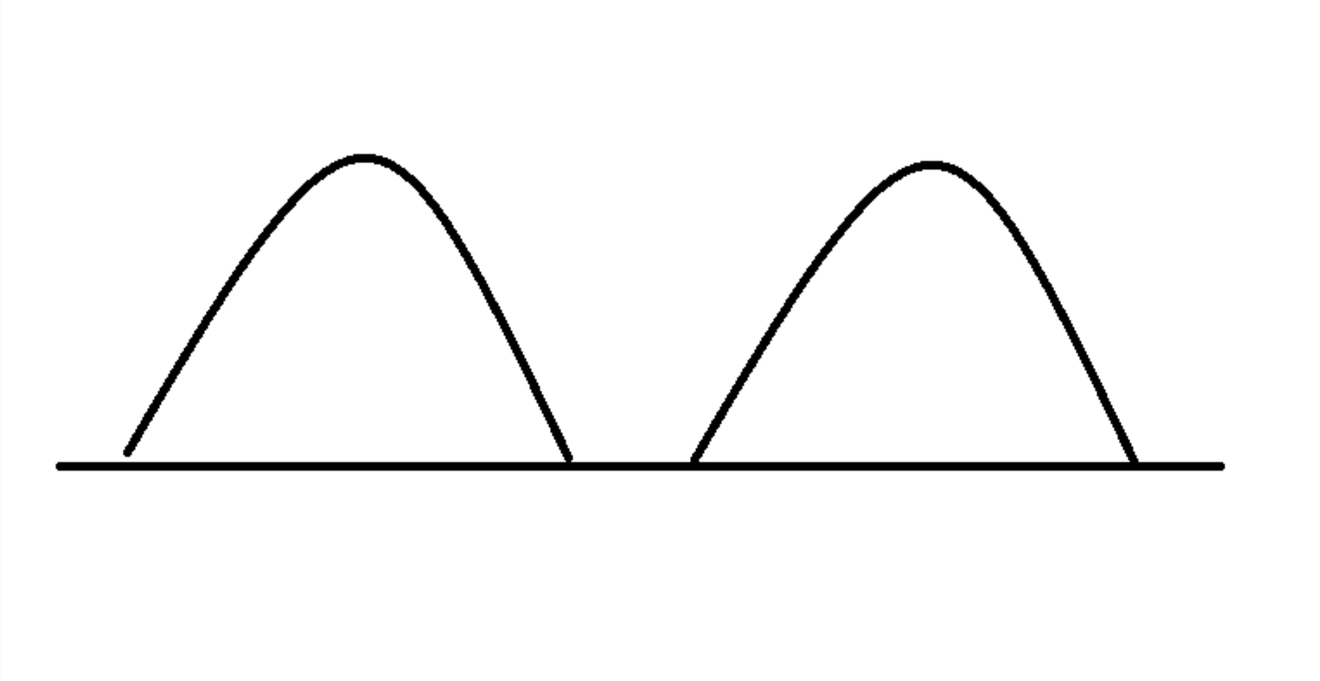
Kernel Density Estimation (KDE)

$$\hat{f}(h,x) = \frac{1}{Nh}\sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right)$$





5 этап



Размерность = 20
bandwidth = 2.335721

Размерность = 10
bandwidth = 0.31622

sklearn
entrophy= 11.002099

matlab
entrophy= 8.895774

github.com/kirthevasank/if-estimators

Вклад в проект

Александр

Анализ теоретического материала

Написание кода

Визуализация данных

Ксения

Поиск и анализ теоретичского материала

Подведение итогов

Оформление презентации

Спасибо за внимание!