Сравнение методов оценок плотностей на конечных размерах выборок.

Ярошенко Ангелина, Ильина Дарья

Постановка задачи

Зачем оценивать энтропию?

- Оценка пропускной способности канала

Постановка задачи

Зачем оценивать энтропию?

- Оценка пропускной способности канала
- Построение стратегий в сэмплировании

Постановка задачи

Зачем оценивать энтропию?

- Оценка пропускной способности канала
- Построение стратегий для методом семплирования
- Теоретическая интерпретация работы нейронных сетей

Постановка задачи

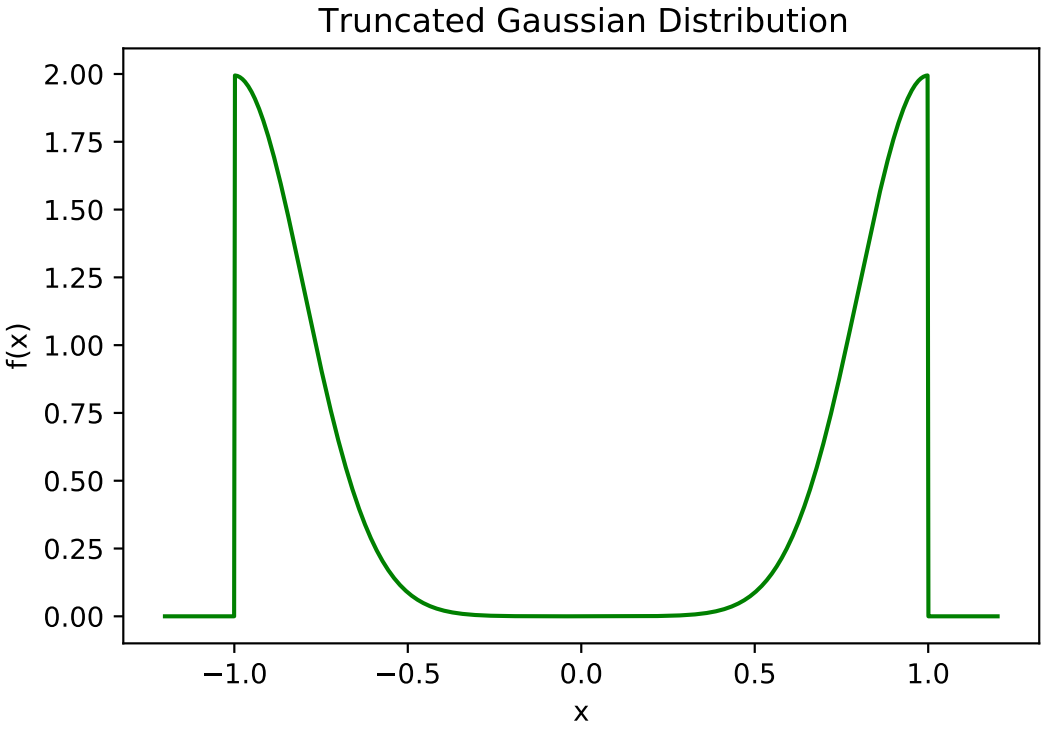
Оцениваемый интеграл:

$$H(x) = -\int\limits_{\infty} f(x) \ln f(x) dx$$

Для распределения:

$$f(x) = \sum_{i=1}^{K} w_i N(\mu_i, \sigma_i^2) \quad \text{,где} \quad \sum_{i=1}^{K} = 1$$

Усеченное Гауссовское распределение



Задача рассматривается в пространствах размерностей 5, 10, 15

Рассмотренные методы

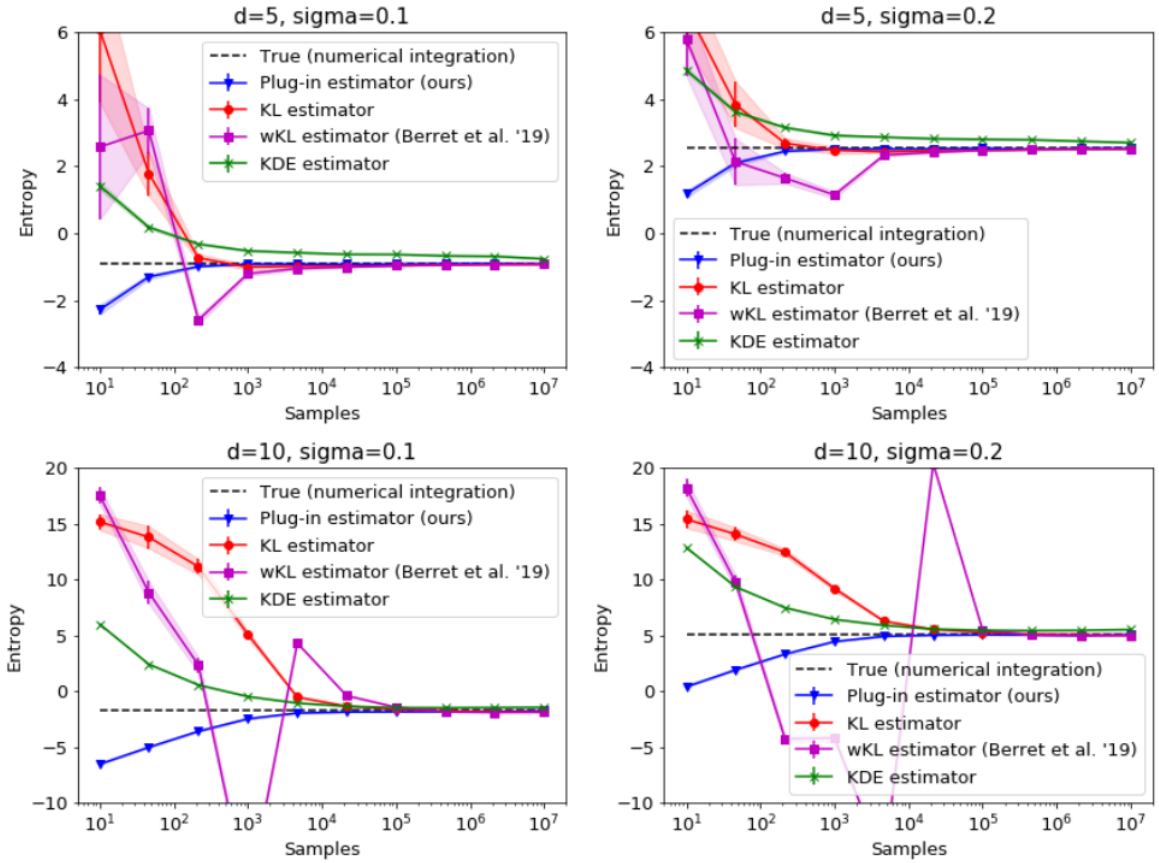
- **KL** Kozachenko L. F., Leonenko N. N. Sample estimate of the entropy of a random vector //Problemy Peredachi Informatsii. – 1987. – T. 23. – №. 2. – C. 9-16.

$$H_N = m\ln(\rho) + \ln V_m + \ln(\gamma) + \ln(N-1)$$

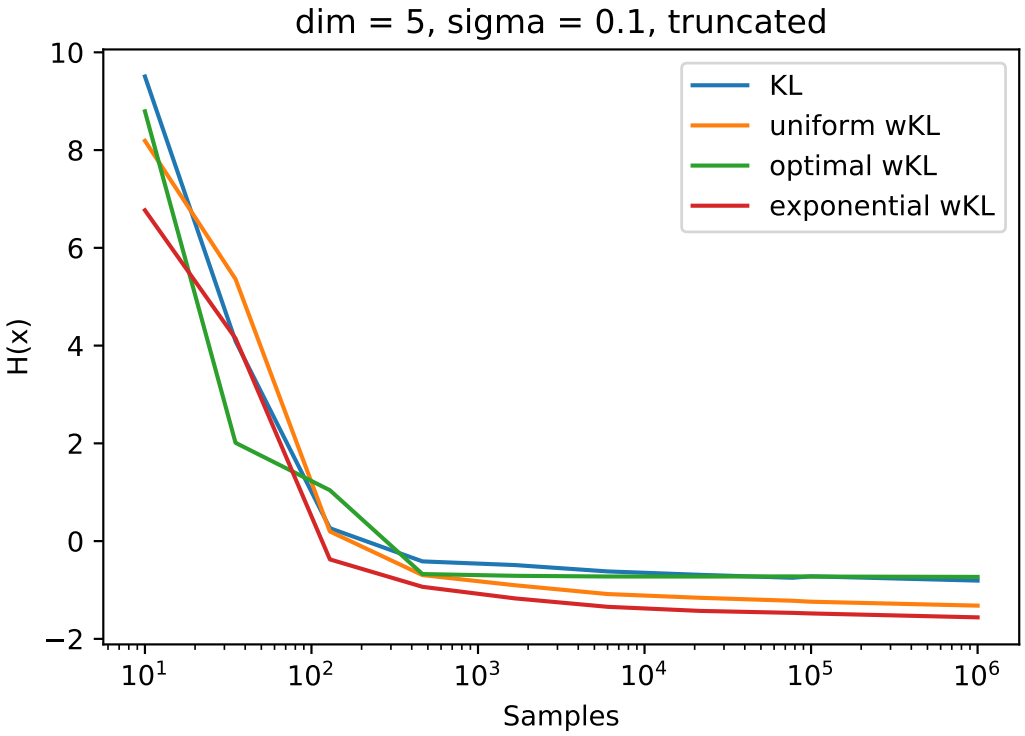
- **wKL** Berrett T. B. et al. Efficient multivariate entropy estimation via k-nearest neighbour distances //The Annals of Statistics. – 2019. – T. 47. – №. 1. – C. 288-318.

$$H_N = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k} w_j \ln(\xi_{j,i})$$

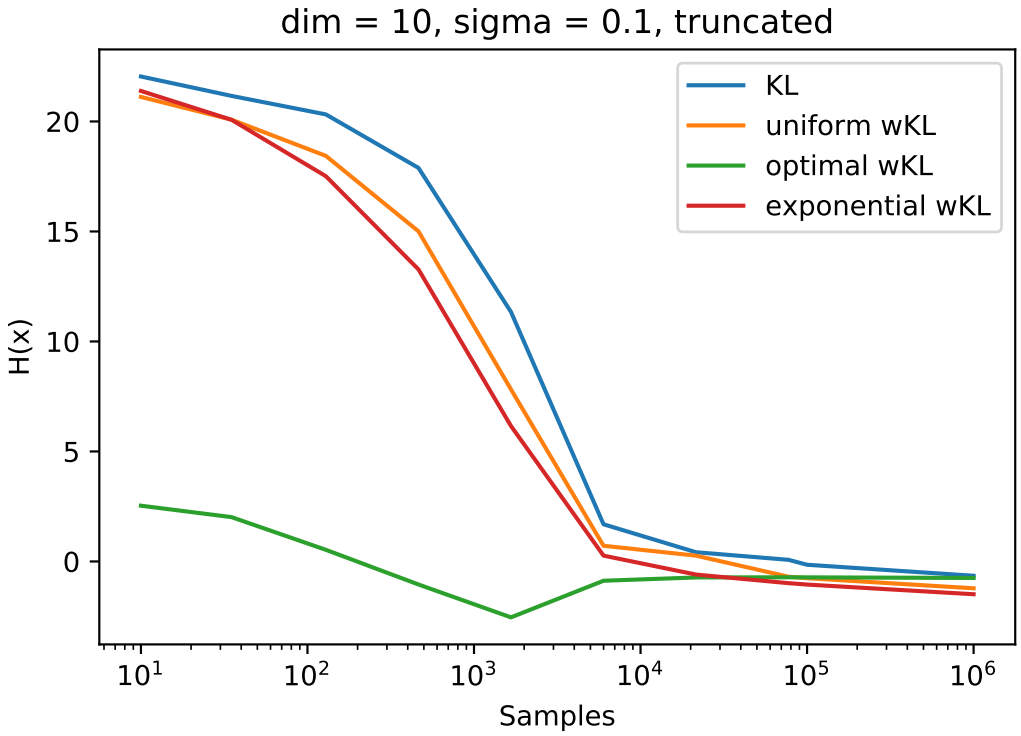
Предполагаемые результаты



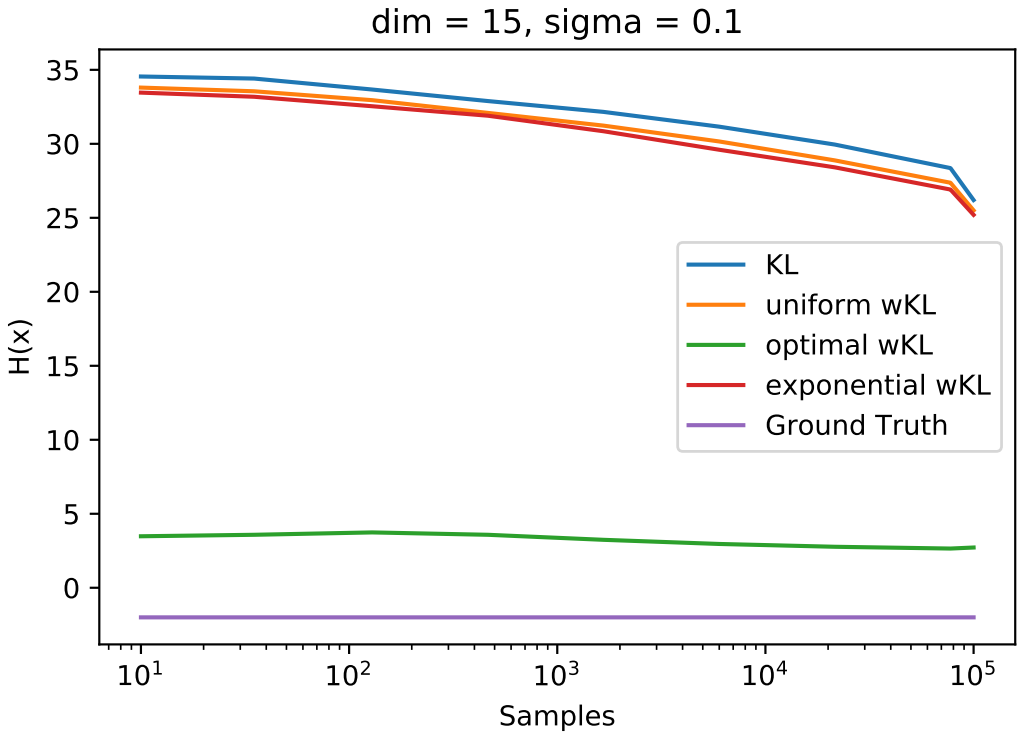
Результаты


dim = 5, sigma = 0.1, truncated
KL
uniform wKL
optimal wKL
exponential wKL
H(x)
Samples
10
8
6
4
2
0
−2
10^1
10^2
10^3
10^4
10^5
10^6

Результаты


dim = 10, sigma = 0.1, truncated
KL
uniform wKL
optimal wKL
exponential wKL
H(x)
20
15
10
5
0
10^1
10^2
10^3
10^4
10^5
10^6
Samples

Результаты


dim = 15, sigma = 0.1
35
30
25
20
15
10
5
0
H(x)
10^1
10^2
10^3
10^4
10^5
Samples
KL
uniform wKL
optimal wKL
exponential wKL
Ground Truth

Результаты

- Теоретически оптимальные веса довольно редко дают качественный результат
- Методы довольно сильно зависят от количества ближайших соседей

Вклад участников команды

Ярошенко Ангелина:

- Реализация KL и wKL, поиск
- Подготовка презентации

Ильина Дарья:

- Подготовка выборок
- Подготовка презентации
- Подбор параметров для KDE