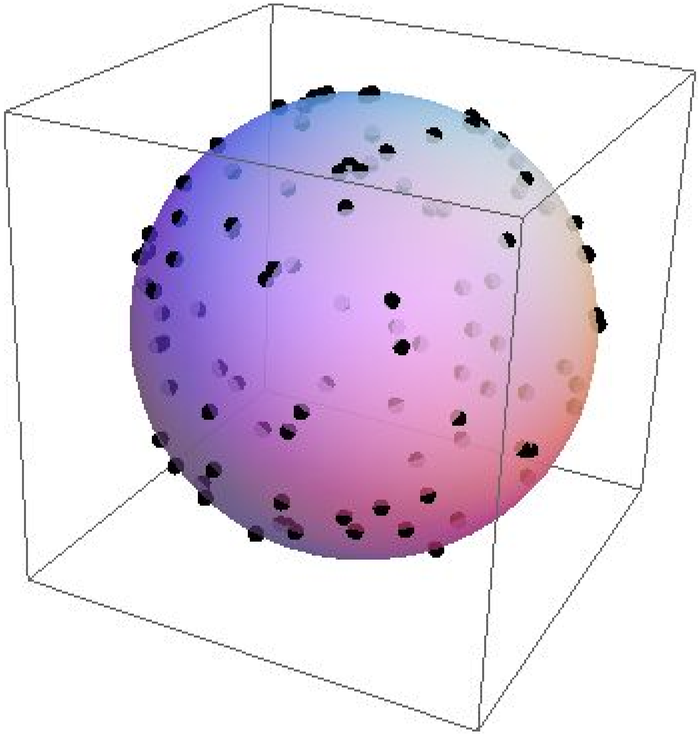
Сколько точек попадает в окрестность на гиперсфере?



Ольховская Юлия

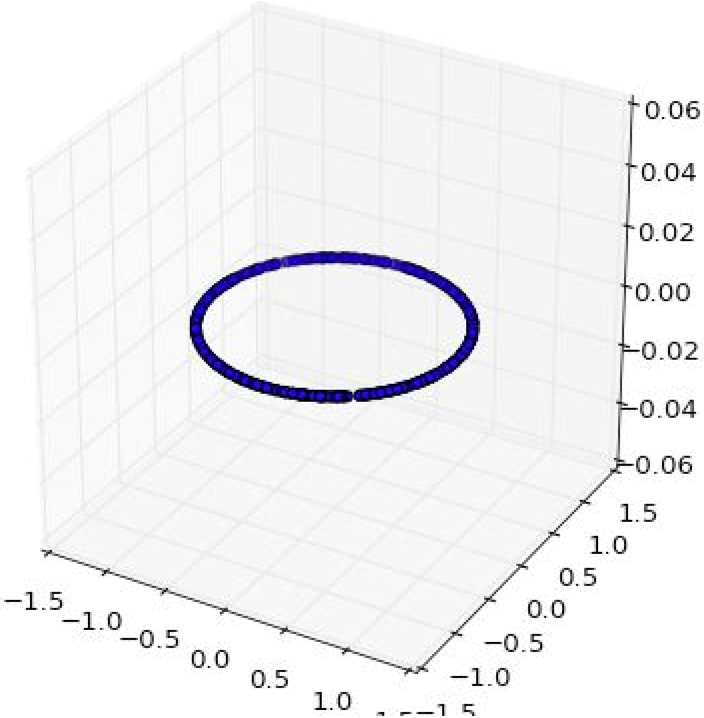
Равномерная выборка на гиперсфере в d-мерном пространстве

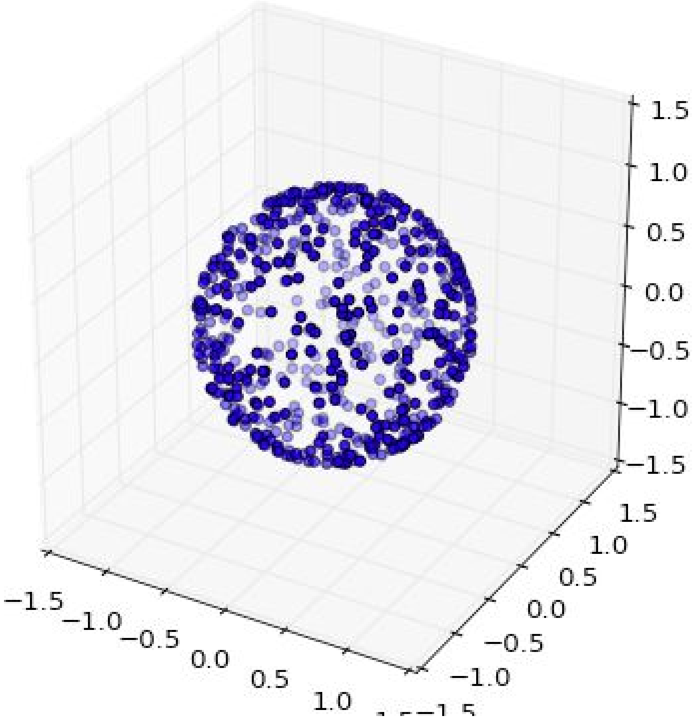
Матрица вращения U является ортогональной.

Пусть вектор X имеет распределение N(0,I). Тогда Y=UX тоже имеет распределение N(0,I).

$$\frac{1}{(\sqrt{2\pi})^d}e^{\left(-\frac{1}{2}<Ux,Ux>\right)}=\frac{1}{(\sqrt{2\pi})^d}e^{\left(-\frac{1}{2}<UU^Tx,x>\right)}=\frac{1}{(\sqrt{2\pi})^d}e^{\left(-\frac{1}{2}<x,x>\right)}$$

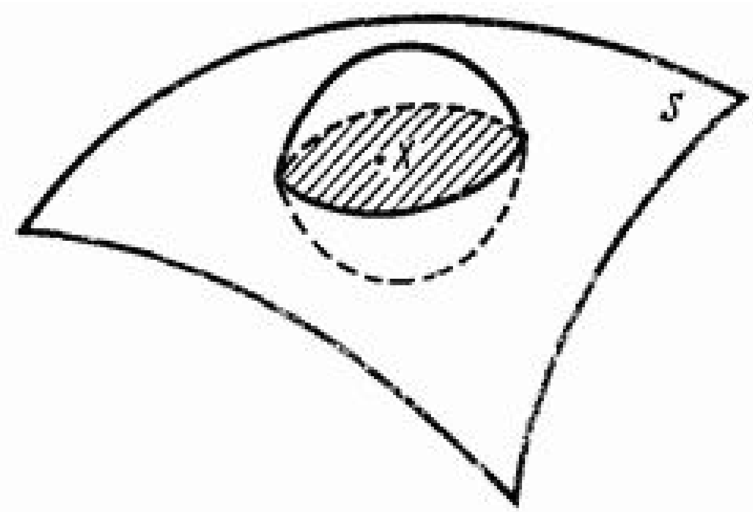
Равномерная выборка на гиперсфере в d-мерном пространстве





Размер окрестности: оценить допустимые C

Окрестность точки определяется как ε-шар в многомерном p-мерном пространстве.



По условию задачи: $\varepsilon(N) = C \cdot N^{-\frac{1}{p+1}}$

C — такая, что ожидание числа случайных точек из выборки в окрестности заданной для наименьшей рассматриваемой в эксперименте выборки больше 1

$$EI\{X_i \in S(\epsilon), i \in \{1, N\}\} = NP\{X_i \in S(\epsilon)\} = N\frac{S(\epsilon)}{S_p(R)}$$

Размер окрестности: оценить допустимые C

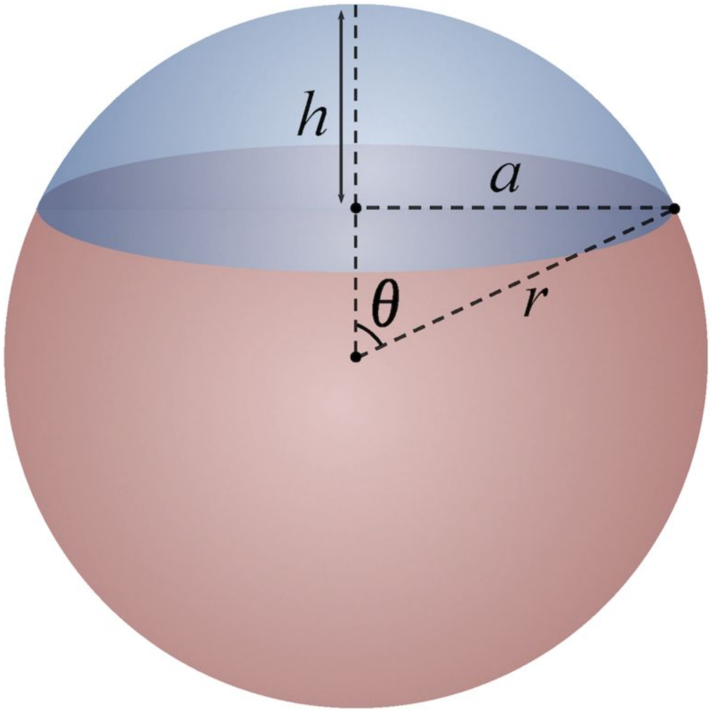
Площадь поверхности гиперсферы:

$$S_n(R) = \frac{2\pi^{(n+1)/2}}{\Gamma(\frac{n+1}{2})R^n}$$

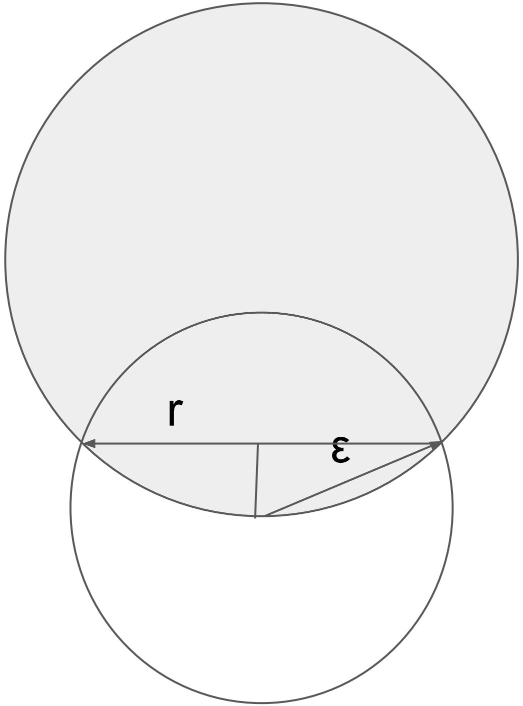
По условию задачи $\varepsilon(N) = C \cdot N^{-\frac{1}{p+1}}$

Площадь сферического сегмента:

$$A = \frac{1}{2} A_n \, r^{n-1} I_{(2rh-h^2)/r^2}\left(\frac{n-1}{2}, \frac{1}{2}\right)$$



Размер окрестности: оценить допустимые C



d = 2

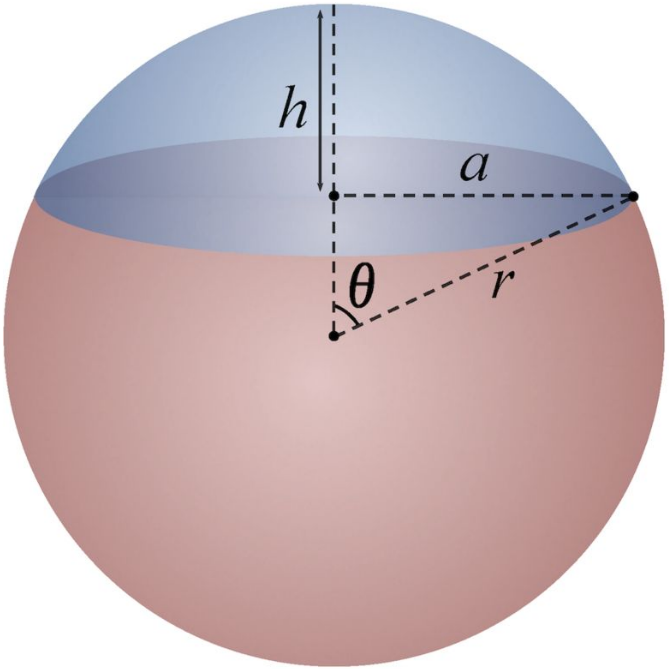
Утверждение:

Площадь сферического сегмента внутри ε-шара можно оценить снизу как площадь поверхности ε/2-шара размерности d-1.

Получаем: $$C = 2\sqrt[n-1]{N^{\frac{2n}{n+1}}K}$$

где $$K = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\sqrt{\pi}}$$

Размер окрестности: оценить допустимые C



d = 3

Утверждение:

Площадь сферического сегмента внутри ε-шара можно оценить снизу как площадь поверхности ε/2-шара размерности d-1.

Получаем:

$$\epsilon(N) = 2\sqrt[n-1]{K}N^{-1/(n-1)}$$

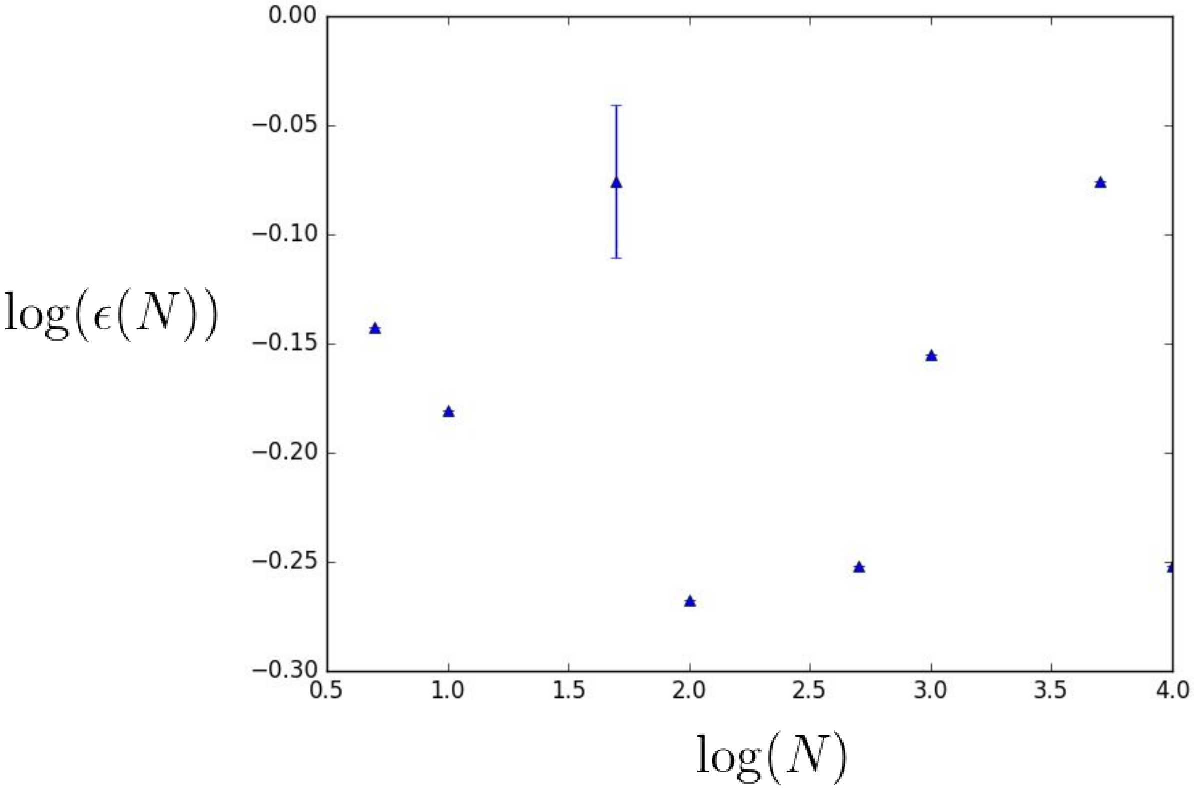
где

$$K = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\sqrt{\pi}}$$

Зависимость числа точек в окрестности от N

d = 3

50 экспериментов

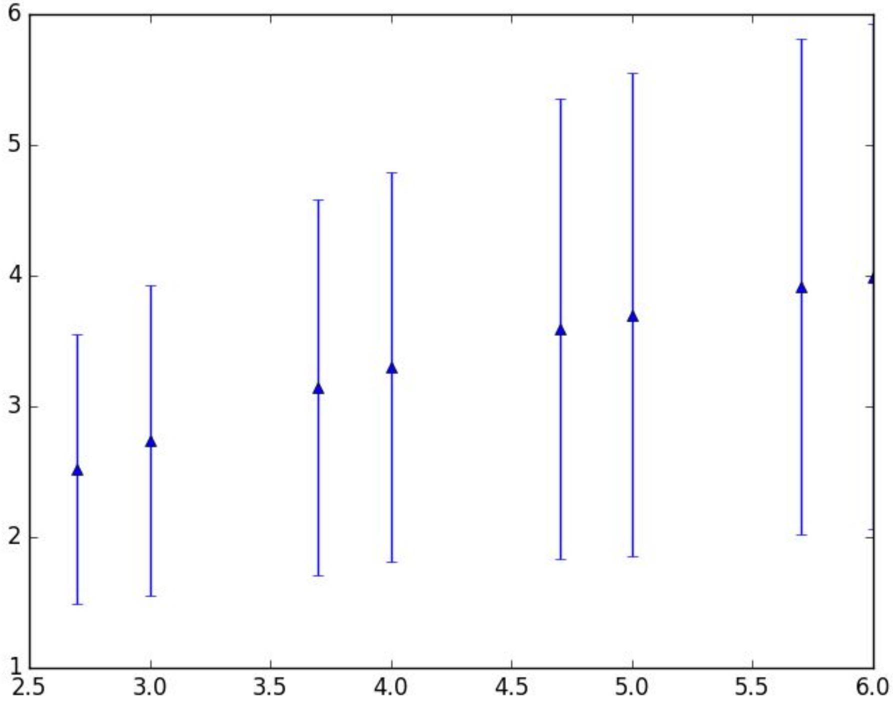


Зависимость числа точек в окрестности от N

d = 20

50 экспериментов

$\log(\epsilon(N))$



$\log(N)$

Зависимость числа точек в окрестности от N

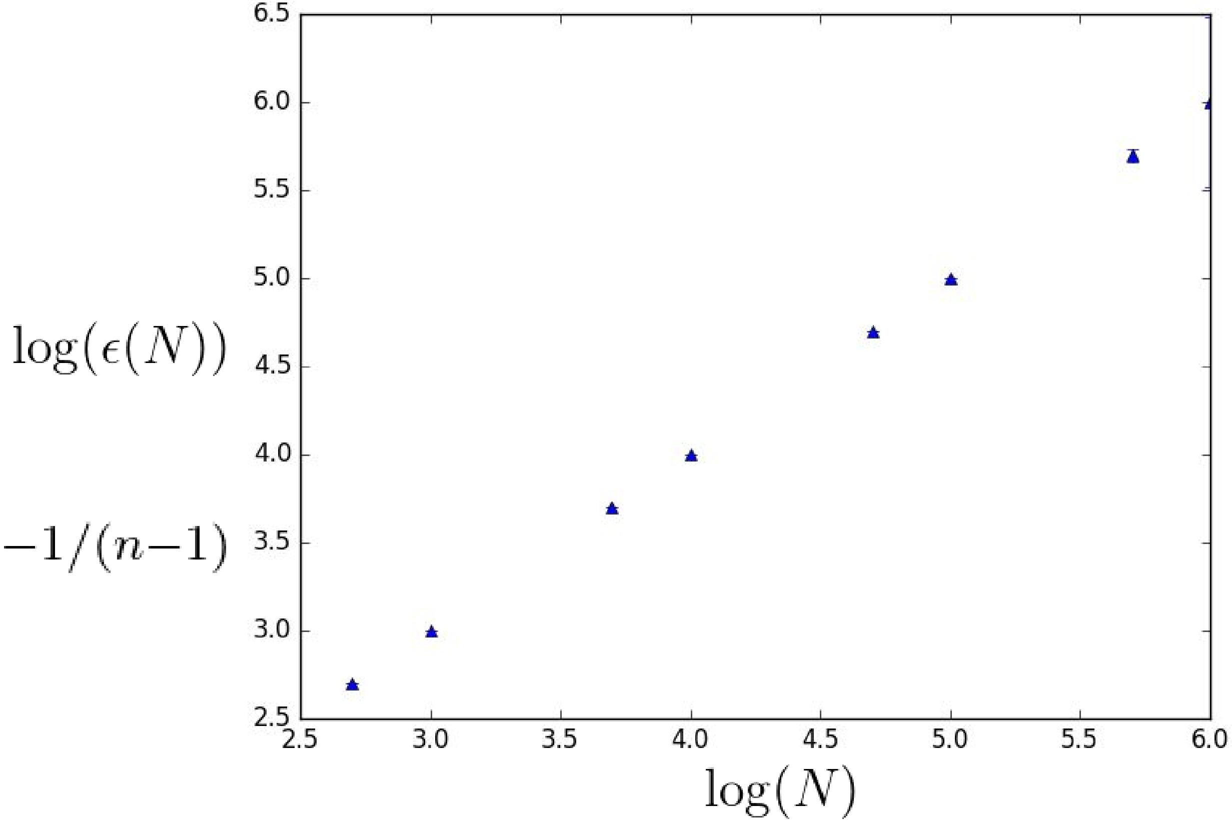
d = 80

50 экспериментов

Нужно больше точек,

так как большое ε

$$\epsilon(N) = 2\,\sqrt[n-1]{K}\,N^{-1/(n-1)}$$

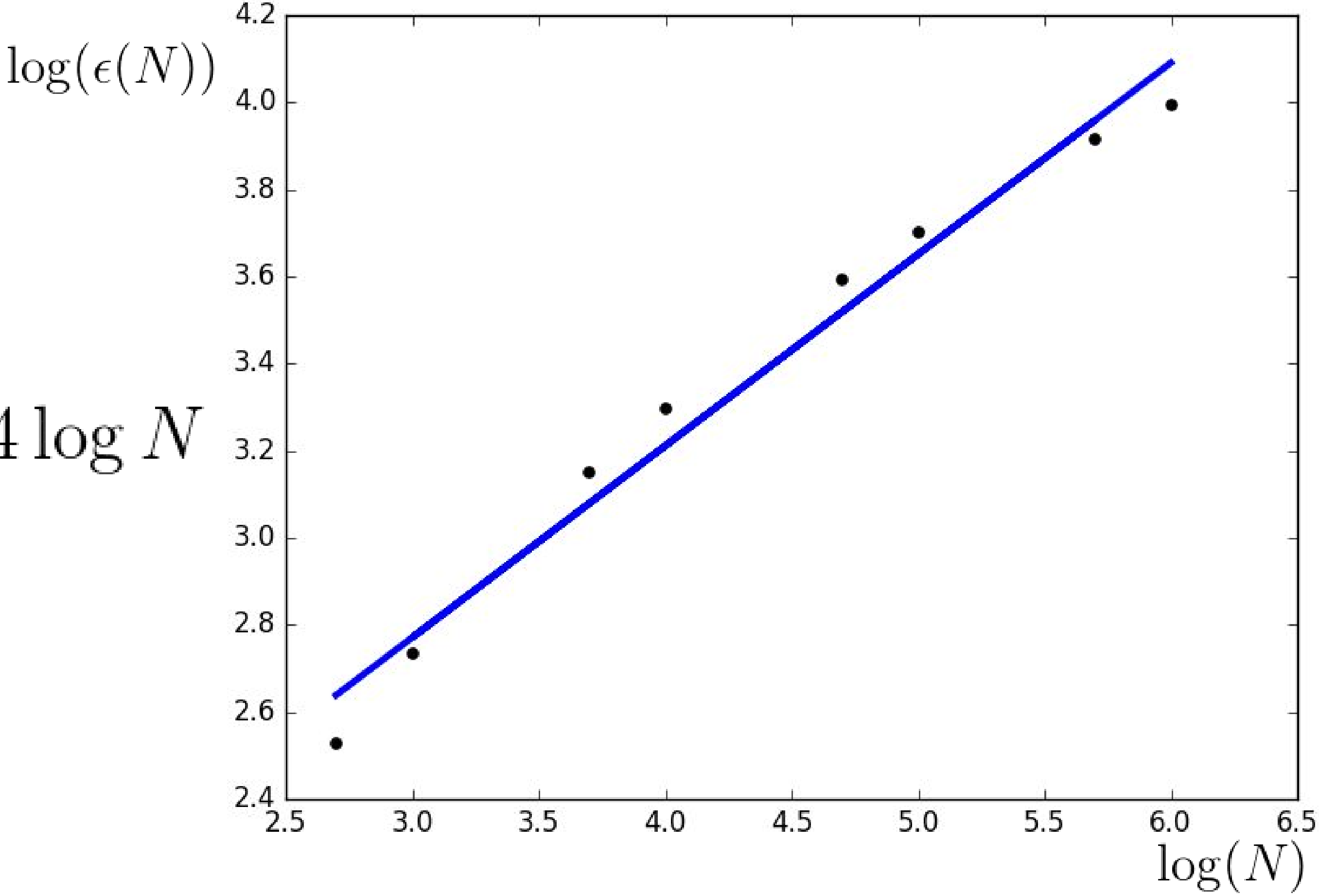


Линейная регрессия для d = 20

Оценка регрессии:

$$\log \epsilon(N) = 0.44 \log N$$

MSE : 2.77e-32



Спасибо за внимание!