Юлия Ольховская

Сколько точек попадает в окрестность на гиперсфере

$X_1, ..., X_N$ генерируется как равномерная выборка с гиперсферы в р-мерном пространстве.

Для того, чтобы сгенерировать выборку используем свойство: нормальное распределение инвариантно относительно поворота.

Доказательство:
Матрица вращения U является ортогональной. Пусть вектор X имеет распределение $N(0, I)$. Тогда $Y = UX$ тоже имеет распределение $N(0, I)$:

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

Размер окрестности:

Площадь части сферы, которая будет лежать внутри ϵ-шара можно оценить снизу как площадь поверхности $\epsilon/2$-шара размерности $d-1$.
По условию задачи : $\epsilon(N) = CN^{-\frac{1}{p+1}}$

Площадь поверхности n-мерной сферы:
$S_n(R) = \frac{2\pi^{(n+1)/2}}{\Gamma(\frac{n+1}{2}) R^n}$
По условию, С — константа, выбираемая так, что ожидание числа случайных точек из выборки в окрестности заданной для наименьшей рассматриваемой в эксперименте выборки больше 1.
Обозначим площадь части сферы, которая будет лежать внутри ϵ-шара как $S(\epsilon)$

$EI\{X_i \in S(\epsilon), i \in \{1, N\}\} = NP\{X_i \in S(\epsilon)\} = N\frac{S(\epsilon)}{S_p(R)}$
Оценка радиуса шара размерности $d-1$. Предположим, что у нас достаточно точек и $\epsilon \leq 1$
$r = \sqrt{\epsilon^2 - \epsilon^4/4R^2} = \sqrt{(\epsilon - \epsilon^2/2)(\epsilon + \epsilon^2/2)} \geq \epsilon - \epsilon^2/2 \geq \epsilon - \epsilon/2 = \epsilon/2$
Нам нужно взять $\frac{S(\epsilon)}{S_n(R)} N \geq \frac{S_{n-1}(\epsilon/2)}{S_n(R)} N = 1$
Пусть $R = 1$
Получаем: $\epsilon^{n-1} = \frac{\Gamma((n+1)/2)}{N\Gamma(n/2)\sqrt{\pi}} 2^{n-1}$
Подставляя условие для $\epsilon(N)$, полагая $K = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\sqrt{\pi}}$:

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

Линейная регрессия.
Можно строить линейную регрессию по средним значениям, так как среднее минимизует сумму квадратов отклонения.

Оценка регрессии:

$\log \epsilon(N) = 0.44 \log N$
$MSE : 2.77e - 32$