Дата //

Задача №1

Пусть X - случайная величина = расстоянию до центра шара.

$$F_X(r) = P(x<r) = \frac{V_r}{V_1} = \frac{cr^d}{c1^d} = \boxed{r^d} \quad (*)$$

$X_{(1)}$ - первая порядковая статистика т.е. с.в.= расстоянию до ближ точки.

Воспользуемся формулой распределения k-ой порядковой статистики для k = 1

$$F_{X_{(1)}}(x) = \sum_{i=1}^{n} C_n^i F^i(x)(1-F(x))^{n-i} =$$

$$= \sum_{i=0}^{n} C_n^i F^i(x)(1-F(x))^{n-i} - F^0(x)(1-F(x))^{n-0} =$$

$$\overset{\text{бином}}{=} (F(x)+(1-F(x)))^n - (1-F(x))^n =$$

$$= 1-(1-F(x))^n$$

Решим $F_{X_{(1)}}(x) = \frac{1}{2}$;

$$1-(1-F(x))^n = \frac{1}{2};$$

$$2^{-1} = (1-F(x))^n;$$

$$2^{-\frac{1}{n}} = 1-F(x);$$

Дата //

Подставим $^{(*)}$ $F(x) = x^d$

$$2^{-\frac{1}{n}} = 1-x^d;$$

$$\boxed{\text{median} = x = \left(1-\frac{1}{\sqrt[n]{2}}\right)^{\frac{1}{d}}}$$

① Если d растёт, то $x \to 1$ при $d \to \infty$, например $d=1000$ $n=10^9$ $\Rightarrow x \sim 0,98$ т.о. все объекты оказываются ~ на одном расстоянии и метод ближайшего соседа становится тяжело применим,

② При $d=10$ $n=1000$ $X \approx 0,48$

при $d=100$, $n=10^{12}$ $X \approx 0,74$ т.е. даже такого объёма выборки не хватило, чтобы сместить медиану в 0.5

Дата //

Задача №2 $\quad a, b \in U[-1,1]$

Пусть $Z = (\bar{z}, a)$, $Y = (\bar{y}, b)$, $X = (\bar{x}, 0)$

$$\rho_x^2 = |\bar{z} - \bar{x}|^2 = \sum_{i=1}^{n} (z_i - x_i)^2$$

$$\rho_y^2 = |\bar{z} - \bar{y}|^2 = \sum_{i=1}^{n} (z_i - y_i)^2$$

$$\tilde{\rho}_x^2 = |\bar{z} - \bar{x}|^2 + a^2 = \rho_x^2 + a^2$$

$$\tilde{\rho}_y^2 = |\bar{z} - \bar{y}|^2 + (a-b)^2 = \rho_y^2 + (a-b)^2$$

$$\mathbb{P}(y\text{-ближайший}) = \mathbb{P}(\tilde{\rho}_y^2 < \tilde{\rho}_x^2) =$$

$$= \mathbb{P}(\underbrace{\tilde{\rho}_y^2 - \rho_x^2}_{V} < 0) = \mathbb{P}(V < 0)$$

$$\boxed{V = \rho_y^2 - \rho_x^2 + b^2 - 2ab}$$

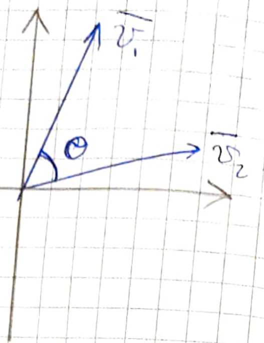
$$\mathbb{P}(V<0) = \mathbb{P}(\rho_y^2 - \rho_x^2 + b^2 - 2ab < 0) =$$

$$= \boxed{\mathbb{P}(b^2 - 2ab < \rho_x^2 - \rho_y^2)}$$

$$b^2 - 2ab = b(b - 2a)$$

Задача №3

Пусть P — вероятность того, что $\bar{v}_1, \bar{v}_2$ в разных полупространствах.



Угол прямой относительно $\bar{v}_2$ против часовой стрелки, принимает значения $[0, \pi]$ и только при $\alpha \in [0, \theta]$, прямая будет разделять векторы.

Т.о. $\boxed{P = \frac{\theta}{\pi}}$ — вероятность в разных плоскостях.

$\boxed{1-P = 1-\frac{\theta}{\pi}}$ — ~~вероятность в разных полуплоскостях~~ — вероятность того, что лежат в одной плоскости.

Интерпретация:

Метод LSH — хорошо работает, если данные более менее равномерно распределены.

Если же например θ — мал, и векторы кучкуются в одном угле, то все $\bar{\omega}$ будут говорить, что $\bar{v}_1$ и $\bar{v}_2$ это один и тот же объект т.к. при малом угле θ

Дата ______ // ______ // ______

$P \to 0 \Rightarrow sign\langle \omega, v_1 \rangle = sign\langle \omega, v_2 \rangle$ для почти всех ω.

Чтобы разделить объекты, придётся брать очень много тестовых векторов ω.

Итого: При малых θ, таких что

$$P = \frac{\theta}{\pi} \to 0$$

метод LSH с семейством $f_\omega(x)$ ПЛОХО ПРИМЕНИМ.

Задача №5

Дано: $x_1, x_2, .. x_n$ – выручка в последовательные дни.

Допустим x_i это реализация случайной величины X_i

$X_i \to$ с.в равная выручке в i-ый день.

Гипотеза 1. Считаем $X_1, .. X_n$ — i i d

По сути нужно найти $E X_{n+1}$ – мат. ожидание выручки в n+1 день.

Гипотеза 2.

$X_i \sim N(\mu, \sigma^2)$

В качестве некоторого обоснования гипотезы №2, можно считать, что:

$$X_i = Y_{i1} + Y_{i2} + \dots + \dots Y_{in}$$

где Y_{ik} – вклад k-го человека в выручку в день i

А также, считать выполненными условия обобщённой ЦПТ Линдеберга.

Обобщённая ЦПТ, говорит о том, что результаты классической ЦПТ верны в более общих ситуациях.

Y_{ik} – не обязательно не зависимые
не обязательно i i d.

Т.о пусть $X_i \sim N(\mu, \sigma^2) \sim \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$

① Метод макс. правдоподобия:

$$L(\bar{x}, \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\varphi = \ln L(\bar{x}, \mu, \sigma^2) = -\ln(2\pi)^{\frac{n}{2}} - \frac{n}{2}\ln\sigma^2 - \frac{\sum (x_i-\mu)^2}{2\sigma^2}$$

$$\varphi'_{\mu} = \frac{2\sum_{i=1}^{n}(x_i-\mu)}{2\sigma^2} = 0;$$ т.к точка экстремума

$$\Rightarrow \boxed{\hat{\mu} = \frac{1}{n}\sum x_i = \bar{X}}$$ – оценка мат ожидания выручки.

$$\varphi'_{\sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{2(\sigma^2)^2} = 0;$$

$$\Downarrow$$

$$\boxed{\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

Оценка дисперсии выручки.

Дата

Задача №4

A - произв. матрица, $A \neq 0$

Дано: $A^T A v = \sigma^2 v$ (*)

Докажем, что если $h = \frac{1}{\sigma} A v$, то $AA^T h = \lambda h$, $\lambda \neq 0$

$$AA^T h = AA^T \cdot \left(\frac{1}{\sigma} A v\right) = \frac{1}{\sigma}(AA^T)\cdot(Av) =$$

$$\overset{\text{ассоциативность}}{=} \frac{1}{\sigma} A(A^T A v) \overset{(*)}{=} \frac{1}{\sigma} A(\sigma^2 v) =$$

$$= \sigma A v = \sigma^2 \underbrace{\left(\frac{1}{\sigma} A v\right)}_{h} = \boxed{\sigma^2 h}$$

Т.о. доказали, что $AA^T h = \sigma^2 h$;

$\Rightarrow h = \frac{1}{\sigma} A v$ - собственный вектор для матрицы AA^T, с собст зн $\lambda = \sigma^2$

Ч.Т.Д.