Задание №17

X_1	4	0	-1	3	4
X_2	2	-3	-2	1	2
X_3	3	2	2	1	-3

Найти главное направление и дисп. по главной компоненте.

Изобразите точки и гл. направление

Итак, имеем

$$X = \begin{pmatrix} 4 & 2 & 3 \\ 0 & -3 & 2 \\ -1 & -2 & 2 \\ 3 & 1 & 1 \\ 4 & 2 & -3 \end{pmatrix} \to \bar{X} = (2, 0, 1)$$

$$X_c = \begin{pmatrix} 2 & 2 & 2 \\ -2 & -3 & 1 \\ -3 & -2 & 1 \\ 1 & 1 & 0 \\ 2 & 2 & -4 \end{pmatrix} \text{ — центрированные данные}$$

$$C = X_c^T X_c = \begin{pmatrix} 2 & -2 & -3 & 1 & 2 \\ 2 & -3 & -2 & 1 & 2 \\ 2 & 1 & 1 & 0 & -4 \end{pmatrix} \begin{pmatrix} 2 & 2 & 2 \\ -2 & -3 & 1 \\ -3 & -2 & 1 \\ 1 & 1 & 0 \\ 2 & 2 & -4 \end{pmatrix} = \begin{pmatrix} 21 & 21 & -9 \\ 21 & 22 & -9 \\ -9 & -9 & 22 \end{pmatrix}$$

$$\frac{1}{N-1} C = \begin{pmatrix} 5.5 & 5.25 & -2.25 \\ 5.25 & 5.5 & -2.25 \\ -2.25 & -2.25 & 5.5 \end{pmatrix} \text{ — выборочная матрица ковариации}$$

Находим с.ч. и собств. вектора матрицы C

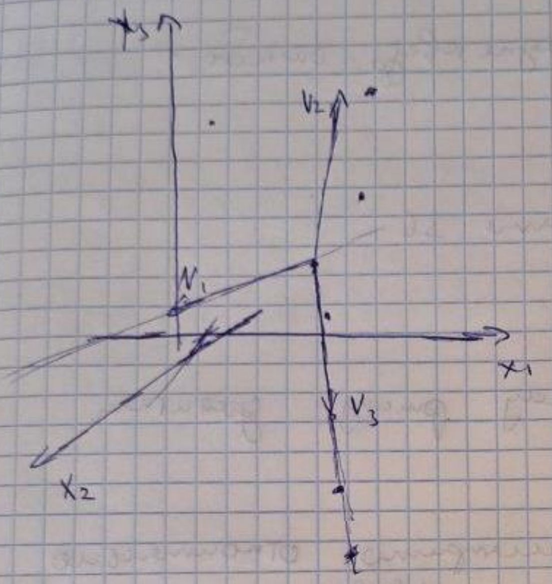
$$\begin{pmatrix} 22-\lambda & 21 & -9 \\ 21 & 22-\lambda & -9 \\ -9 & -9 & 22-\lambda \end{pmatrix} \to \lambda_1 = 1 \quad \lambda_2 = 16 \quad \lambda_3 = 49$$

$$V_1 = \frac{\sqrt{2}}{2}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \quad V_2 = \frac{\sqrt{11}}{11}\begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} \quad V_3 = \frac{\sqrt{22}}{22}\begin{pmatrix} 3 \\ 3 \\ -2 \end{pmatrix}$$

главные компоненты

Дисперсия по главным компонентам

$\frac{1}{N-1}\lambda_1 = 0.25$ $\frac{1}{N-1}\lambda_2 = 4$ $\frac{1}{N-1}\lambda_3 = 12.25$



Задание 35

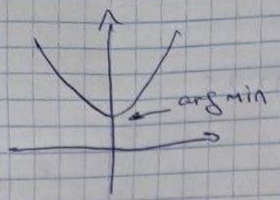
$L(y', y) = (y' - y)^2$

Док-ть, что если $f^*(x) = \arg\min_c M((Y-c)^2 | X = x)$, то $f^*(x) = M(Y | X = x)$

$R^*(f)$ (Средний ~~риск~~ риск) — ?

$\arg\min_c M((Y-c)^2 | X = x) \in \{ c: \forall k\ M((Y-k)^2 | X = x) \geq$
$\geq M((Y-c)^2 | X = x) \}$

Для квадр. ф-ии min достигается в вершине параболы $\longrightarrow$ при фикс. $\bar{Y}$



$$\underset{c}{\mathrm{argmin}}\ M((Y-c)^2 \mid X=x) = M(Y \mid X=x)$$

$$((Y-c)^2)' = 0 \qquad (-2(Y-c))'_c > 0$$

$$-2(Y-c) = 0 \qquad 2 > 0 \Rightarrow \min$$

$$c = Y$$

Средний риск равен среднекв. ошибке:

$$R(f^*) = M((y'-y)^2)$$

Задача 36

$$L(y',y) = |y'-y|$$

Док-ть, что min среднему риску достигает $f(x) = \mathrm{median}(Y \mid X=x)$

$|y-y'|^2 = (y-y')^2$, симметрично относительно центра [illegible]

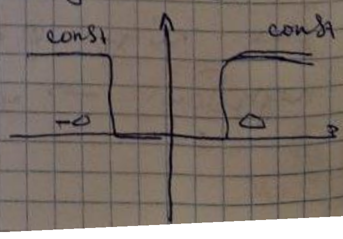
$\Leftarrow$

средний риск будет равен среднему значению ф-ции $\Rightarrow$ равен ~~медиане~~ медиане распределения

Задача № 37

Если потери от ошибок нулевые в ~~некоторой~~ окр-ти некоторого числа ~~а~~ нулевые, а для всей остальной прямой — равные, то условная мода даёт min ср. риска, т.к. мода характеризует наиболее часто встречающееся значение.

$$L(y',y) = \begin{cases} 0, & y'-y \le \Delta \\ \text{const}, & |y'-y| > \Delta \end{cases}$$



Упражнение 41

$$\overline{x} = \frac{1}{N} \sum_{i=1}^{N} x^{(i)}$$

$W_k = v_0 + L(v_1 \dots v_k)$ — k-мерное лин. многообразие

Будем искать v_0 решение задачи minimize L_0:

$$v_0 = \arg\min_{a_0 \in R^n} \left(\sum_{i=1}^{N} \mathrm{dist}^2(x_i, a_0) \right) = \arg\min_{a_0 \in R^n} \left(\sum_{i=1}^{N} \|(x^{(i)} - a_0)\|^2 \right) =$$

$$= \frac{1}{N} \sum_{i=1}^{N} x^{(i)} = \overline{x}$$

выборочное среднее