ML_homework № 2.

17 Дана выборка:

$$X = \begin{array}{ccc} x_1 & x_2 & x_3 \\ \end{array}$$
$$X = \begin{pmatrix} 4 & 2 & 3 \\ 0 & -3 & 2 \\ -1 & -2 & 2 \\ 3 & 1 & 1 \\ 4 & 2 & -3 \end{pmatrix}$$

Найдём $\overline{x_1} = \frac{4+0+(-1)+3+4}{5} = 2$

$\overline{x_2} = \frac{2-3-2+1+2}{5} = 0$

$\overline{x_3} = \frac{3+2+2+1-3}{5} = 1$

$\overline{x}(2;0;1)$

1 шаг. Вычитаем компоненты $\overline{x}$ из 1,2,3 столбца матрицы X соответственно, получаем:

$$X_c = \begin{array}{ccc} a_1 & a_2 & a_3 \\ \end{array}$$
$$X_c = \begin{pmatrix} 2 & 2 & 2 \\ -2 & -3 & 1 \\ -3 & -2 & 1 \\ 1 & 1 & 0 \\ 2 & 2 & -4 \end{pmatrix} =$$

2 шаг. Находим

$$C = X_c^T \cdot X_c = \begin{pmatrix} 2 & 2 & 2 \\ -2 & -3 & 1 \\ -3 & -2 & 1 \\ 1 & 1 & 0 \\ 2 & 2 & -4 \end{pmatrix}^{\cancel{T}} \cdot \begin{pmatrix} 2 & 2 & 2 \\ -2 & -3 & 1 \\ -3 & -2 & 1 \\ 1 & 1 & 0 \\ 2 & 2 & -4 \end{pmatrix}^T =$$

$$= \begin{pmatrix} 12 & -8 & -8 & 4 & 0 \\ -8 & 14 & 13 & -5 & -14 \\ -8 & 13 & 14 & -5 & -14 \\ 4 & -5 & -5 & 2 & 4 \\ 0 & -14 & -14 & 4 & 24 \end{pmatrix} \text{ или } \begin{pmatrix} 22 & 21 & -9 \\ 21 & 22 & -9 \\ -9 & -9 & 22 \end{pmatrix}$$

Матрица $\frac{1}{N-1} C = \begin{pmatrix} \frac{22}{4} & \frac{21}{4} & -\frac{9}{4} \\ \frac{21}{4} & \frac{22}{4} & -\frac{9}{4} \\ -\frac{9}{4} & -\frac{9}{4} & \frac{22}{4} \end{pmatrix} =$

– это выборочная матрица ковариации.

3 шаг. Находим собственные числа и вектора матрицы C:

$$\det(C-\lambda I) = \begin{pmatrix} 22-\lambda & 21 & -9 \\ 21 & 22-\lambda & -9 \\ -9 & -9 & 22-\lambda \end{pmatrix}$$

$$\lambda_1 = 1 \longrightarrow v_1 = \frac{\sqrt{2}}{2}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}$$

$$\lambda_2 = 16 \longrightarrow \frac{\sqrt{11}}{11}\begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} = v_2$$

$$\lambda_3 = 49 \longrightarrow \begin{pmatrix} 3 \\ 3 \\ -2 \end{pmatrix} = v_3$$

Вектора v_1, v_2, v_3 – главные компонента.

Значения

$$\frac{1}{N-1}\lambda_1 = \frac{1}{N-1}\sigma_1^2 = \frac{1}{4}$$

$$\frac{1}{N-1}\lambda_2 = \frac{1}{N-1}\sigma_2^2 = \frac{16}{4} = 4$$

$$\frac{1}{N-1}\lambda_3 = \frac{1}{N-1}\sigma_3^2 = \frac{49}{4} = 12\frac{1}{4}$$

– дисперсии по главным компонентам.

(36). Рас. задачу восст. регрессии с ф-цией потерь $L(y', y) = |y' - y|$. Д-ть, что минимум среднему риску доставляет при этом условная медиана $f(x) = median(Y|X = x)$.

Д-тво: (от противного)

Если значение среднего риска не будет совпадать с медианным значением функции $f(x)$, что и является медианой распределения, тогда функция $L(y', y) = |y' - y|$ не будет симметрична относительно цента системы координ.

А мы знаем, что $f(y', y) = |y' - y|^2$, симметрична и отражается от оси абцисс в т. $(0, 0)$.

Упражнение 4.1. Д-ть, что можно положить $v_0 = \overline{x}$. Описать мн-во всех возможных v_0, доставляющих минимум сумме квадратов расстояний до иск. многообр.

Т.О. можно считать, что данные центрированы:

$x^{(i)} \leftarrow x^{(i)} - \overline{x} \ (i = \overline{1, N})$

$$W_k = v_0 + L(v_1, \ldots, v_k), \ v_j \in R^d, \|v_j\| = 1, v_j \perp v_{j'} \ (j \neq j')$$

$$\sum_{i=1}^{N} dist^2(x^{(i)}, W_k) = \sum_{i=1}^{N} \|ort_{W_k} x^{(i)}\|^2 \longrightarrow min$$

$$v_0 = \overline{x} = \frac{1}{N} \sum_{i=1}^{N} x^{(i)} = min \left(\sum_{i=1}^{N} \| x^{(i)} - w_0 \|^2 \right)$$