

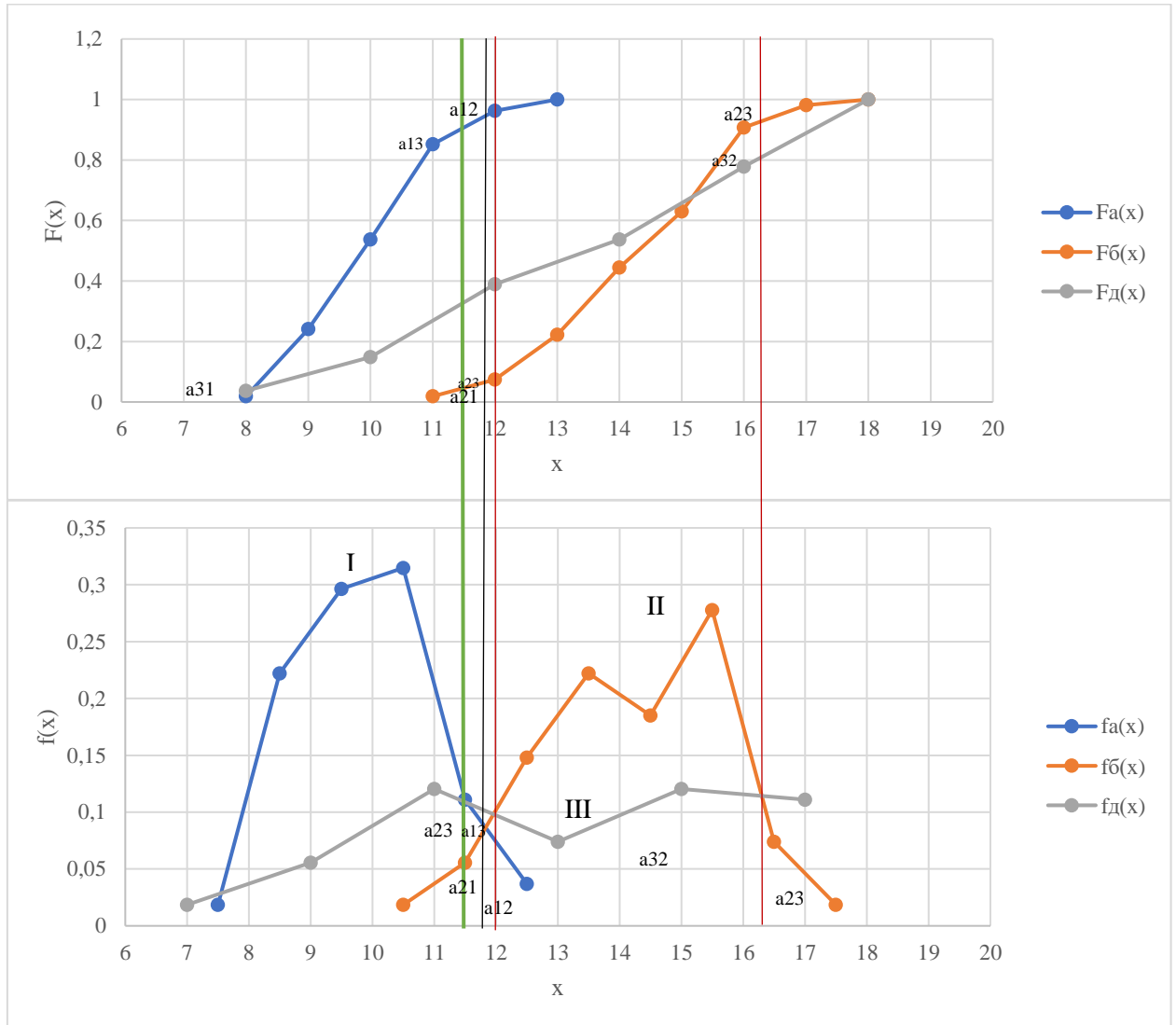
Лабораторная работа №3.

Элементы теории распознавания образов.

Наша задача состоит в том, чтобы вычислить достоверность модели классификации горных пород.

Даны: $p(H1)=0,2$; $p(H2)=0,5$; $p(H3)=0,3$

Построим интегральную и дифференциальную функции распределения по изучаемому признаку:



Заполним табл.1:

Таблица 1 – Парные ошибки классификации

i/j	1	2	3
1	$\gamma_1 = 0,86$	$a_{12}=b_{12}=0,06$	$a_{13}=b_{13}=0,1$
2	$a_{21}=b_{21}=0,06$	$\gamma_2 = 0,81$	$a_{23}=b_{23}=0,14$
3	$a_{31}=b_{31}=0,32$	$a_{32}=b_{32}=0,41$	$\gamma_3 = 0,80$

$$\gamma_1 = 1 - (P(H_i) * \left(\frac{\sum_{j=1}^{m-1} a_{i,j} * P(H_j)}{\sum_{j=1}^{m-1} P(H_j)} \right) + \left(\sum_{j=1}^{m-1} a_{j,i} * P(H_j) \right)); j \neq i$$

$$\gamma_1 = 1 - 0,2 * \left(\frac{0,06 * 0,5 + 0,1 * 0,3}{0,5 + 0,3} \right) + (0,06 * 0,5 + 0,32 * 0,3) = 0,86; j \neq i$$

$$\gamma_2 = 1 - 0,5 * \left(\frac{0,06 * 0,2 + 0,14 * 0,3}{0,2 + 0,3} \right) + (0,06 * 0,2 + 0,41 * 0,3) = 0,81; j \neq i$$

$$\gamma_3 = 1 - 0,3 * \left(\frac{0,32 * 0,2 + 0,41 * 0,5}{0,2 + 0,5} \right) + (0,1 * 0,2 + 0,614 * 0,5) = 0,80; j \neq i$$

Рассчитаем достоверность классификации горных пород по признаку X:

$$G(X) = \sum_{i=1}^m P(H_i) * \gamma_i(x) = 0,2 * 0,86 + 0,5 * 0,81 + 0,3 * 0,80 = 0,82$$

2. Вычисление информативности модели документации горных пород.

Рассчитаем энтропию модели геологического пространства, учитывая, что

$$\eta(p) = -p * \log_2 p;$$

$$H(x) = \eta(0,2) + \eta(0,5) + \eta(0,3) = 0,46 + 0,5 + 0,52 = 1,49$$

Рассчитаем условные вероятности диагностики по формуле и сведем в таб.2:

$$\mu_{ij} = p\left(\frac{y_j}{x_i}\right) = a_{ij} * \frac{p(x_j)}{\sum_{j=1}^m p(x_j)}; j \neq i$$

$$\mu_{12} = 0,06 * \frac{0,5}{0,5 + 0,3} = 0,038;$$

$$\mu_{13} = 0,1 * \frac{0,3}{0,5 + 0,3} = 0,038;$$

$$\mu_{11} = 1 - (0,038 + 0,038) = 0,924;$$

$$\mu_{21} = 0,06 * \frac{0,2}{0,2 + 0,3} = 0,024;$$

$$\mu_{23} = 0,14 * \frac{0,3}{0,2 + 0,3} = 0,084;$$

$$\mu_{22} = 1 - (0,024 + 0,084) = 0,892;$$

$$\mu_{31} = 0,32 * \frac{0,2}{0,2 + 0,5} = 0,091;$$

$$\mu_{32} = 0,41 * \frac{0,5}{0,2 + 0,5} = 0,293;$$

$$\mu_{33} = 1 - (0,091 + 0,293) = 0,616;$$

Таблица 2 – Условные вероятности диагностики

i/j	Y_1	Y_2	Y_3
X_1	0,924	0,038	0,038
X_2	0,024	0,892	0,084
X_3	0,091	0,293	0,616

Составим таб.3 вероятностей совместной системы (ху) P(ху):

$$P_{ij} = p(x_i) * p\left(\frac{y_j}{x_i}\right)$$

$$P_{11} = 0,2 * 0,924 = 0,185$$

$$P_{12} = 0,2 * 0,038 = 0,008$$

$$P_{13} = 0,2 * 0,038 = 0,008$$

$$P_{21} = 0,5 * 0,024 = 0,012$$

$$P_{22} = 0,5 * 0,892 = 0,446$$

$$P_{23} = 0,5 * 0,084 = 0,042$$

$$P_{31} = 0,3 * 0,091 = 0,027$$

$$P_{32} = 0,3 * 0,293 = 0,088$$

$$P_{33} = 0,3 * 0,616 = 0,185$$

Таблица 3 – Двумерного закона распределения

i/j	Y_1	Y_2	Y_3	p(x)
X_1	0,185	0,008	0,008	0,2
X_2	0,012	0,446	0,042	0,5
X_3	0,027	0,088	0,185	0,3
p(y)	0,224	0,542	0,235	1

Рассчитаем энтропию системы геологической документации:

$$E(y) = \sum_{j=1}^m \eta(p(y_j)) = \eta(0,224) + \eta(0,542) + \eta(0,235) = 1,453$$

Пользуясь таблицей условных вероятностей, рассчитаем частные условные энтропии по формуле:

$$E\left(\frac{y}{x_i}\right) = - \sum_{j=1}^m p\left(\frac{y_j}{x_i}\right) \log_2 p\left(\frac{y_j}{x_i}\right) = \sum_{j=1}^m \mu_{ij} \log_2 \mu_{ij}$$

$$E\left(\frac{y}{x_1}\right) = \eta(0,924) + \eta(0,038) + \eta(0,038) = 0,464$$

$$E\left(\frac{y}{x_2}\right) = \eta(0,024) + \eta(0,892) + \eta(0,084) = 0,576$$

$$E\left(\frac{y}{x_3}\right) = \eta(0,091) + \eta(0,293) + \eta(0,616) = 1,264$$

Тогда полная условная энтропия на одну точку наблюдения, вычисленная как средневзвешенная, будет равна:

$$\left(\frac{Y}{X}\right) = 0,2 * 0,464 + 0,5 * 0,576 + 0,3 * 1,264 = 0,76$$

Количество информации, заключенное в легенде геологической документации, состоящей из трех типов пород на одну точку геологической карты равна:

$$J_{y \rightarrow x} = E(y) - E\left(\frac{y}{x}\right) = 1,453 - 0,76 = 0,693 \text{ дв. ед.}$$

$$\frac{J_{y \rightarrow x}}{J_x} * 100\% = \frac{J_{y \rightarrow x}}{H(x)} * 100\% = \frac{0,693}{1,49} * 100 = 46,51 \%$$

Таким образом, составленная классификация горных пород района на основе изучения признака X позволит получить более половины возможной информации о геологическом пространстве в рамках принятой модели.