Задача. *4.1 Наивный байес и центроидный классификатор*

Proof. Наивный байесовский классификатор присваивает метки классов по следующей формуле:

$$\arg\max_y p(Y = y) \prod_{i=1}^{n} p(X^{(k)} = x^{(k)} \mid Y = y)$$

В условиях задачи это можно переписать как

$$\arg\max_y \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x^{(k)}-\mu_{yk})^2}{2\sigma^2}} =$$

$$\arg\max_y \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\sum_{i=1}^{n} - \frac{(x^{(k)}-\mu_{yk})^2}{2\sigma^2}} =$$

$$\arg\max_y \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\rho(x,\mu_y)}{2\sigma^2}} =$$

$$\arg\min_y \rho\left(x, \mu_y\right)$$

□

Задача. *4.2 ROC-AUC*

Proof. ROC-кривая будет состоять из трех точек: $(0,0)$, $(1,1)$ и (p,p), то есть всегда будет просто отрезком прямой, ограничивающим площадь 0.5 □

Задача. *4.3 Ошибка 1NN и оптимального байесовского классификатора*

Proof. В условиях задачи (равномерное заполнение пространства) при $n \to \infty$ ближайший сосед приближается к объекту: $x_n \to x$. Тогда имеем

$$E_N = P(y \neq y_n) = P(1|x)P(0|x_n) + P(0|x)P(1|x_n) \to P(1|x)P(0|x) + P(0|x)P(1|x) =$$

$$2P(0|x)P(1|x) \leq 2\min\left\{P(0|x), P(1|x)\right\} = 2E_B$$

□