4.1. Т.к. "наивность" наивного байесовского классификатора заключается в том, что распределение признаков от класса привлекаемым, готовое распределение будет представлять собой произведение:

$$P(C \mid F_1 \ldots F_n) = \frac{1}{A} P(c) \prod_{i=1}^{n} P(F_i \mid C) = \frac{1}{A} P(c) \frac{1}{\sqrt{2\pi\sigma^2}} \prod_{k=1}^{n} e^{-\frac{(x^{(k)} - \mu_{y_k})^2}{2\sigma^2}}$$

норм.

$$= \frac{1}{A} P(c) \, e^{-\sum_{k=1}^{n} \frac{(x^{(k)} - \mu_{y_k})^2}{2\sigma^2}}$$

множитель

Т.к. мы максимизируем эту вероятность, значит, минимизируем

$$\sum_{k=1}^{n} (x_1^{(k)} - \mu_{y_k})^2$$

(т.к. все остальные параметры не зависят от тестовой выборки).

Но $\sum_{k=1}^{n} (x^{(k)} - \mu_{y_k})^2$ и есть функция расстояния (с точностью до норм, но корень - монотонная функция, и число не меняет). ⇒ мы будем минимизировать расстояние ⇒ возьмём ближайшую μ_{y_k} ⇒ возьмём класс, чей центр ближе всего к X.