В наивном байессовском классификаторе мы выбираем для объекта с признаками x такой класс y, у которого $P(y|x)$ наибольшая. По формуле Байеса:

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}.$$

Максимизация идет по y, поэтому знаменатель вообще не важен, а $P(y)$ по условию равны для любого y. Тогда выбран будет класс с наибольшей $P(x|y)$.

Давайте теперь посмотрим что же из себя представляет $P(x|y)$. Во-первых, мы знаем, что

$$P(x^{(k)}|y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x^{(k)}-\mu_{yk})^2}{2\sigma^2}},$$

а во-вторых, в наивном байессовском классификаторе мы предполагаем, что все признаки независимы. Тогда имеем:

$$P(x|y) = \prod_{k=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x^{(k)}-\mu_{yk})^2}{2\sigma^2}} = \frac{1}{(\sqrt{2\pi\sigma^2})^n} e^{-\frac{\sum\limits_{k=1}^{n}(x^{(k)}-\mu_{yk})^2}{2\sigma^2}}.$$

То есть, максимизируя $P(x|y)$, мы на самом деле минимизируем $\sum\limits_{k=1}^{n}(x^{(k)} - \mu_{yk})^2$, а это расстояние до μ_y. Что и требовалось показать.