Задача 1

$$P(c|d) = \frac{P(d|c)\cdot P(c)}{P(d)}$$

$$c_{map} = \operatorname*{argmax}_{c\in C} \frac{P(d|c)\, P(c)}{P(d)} = \operatorname*{argmax}_{c\in C} P(d|c)\cdot P(c) =$$

$$\left| P(d|c) \approx P(w_1|c)\cdot P(w_2|c)\cdot\ldots\cdot P(w_n|c) = \prod_{i=1}^{n} P(w_i|c) \right|$$

$$= \operatorname*{argmax}_{c\in C} \left(P(c) \prod_{i=1}^{n} P(w_i|c)\right) = \left| \underbrace{P(c)}_{\text{const}} \ \forall c \in C \right|$$

$$= \operatorname*{argmax}_{c\in C} \prod_{i=1}^{n} P(w_i|c) = \operatorname*{argmax}_{c\in C} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(w_i-\mu_c)^2}{2\sigma^2}\right) =$$

$$= \operatorname*{argmax}_{c\in C} \prod_{i=1}^{n} \exp\left(-\frac{(w_i-\mu_c)^2}{2\sigma^2}\right) = \operatorname*{argmax}_{c\in C} \exp\left(-\frac{\sum_{i=1}^{n}(w_i-\mu_c)^2}{2\sigma^2}\right) =$$

$$= \operatorname*{argmin}_{c\in C} \exp\left(\sum_{i=1}^{n} (w_i-\mu_c)^2\right) = \operatorname*{argmin}_{c\in C} \left(\sum_{i=1}^{n}(w_i-\mu_c)^2\right)$$

Из последнего выражения видно, что классификация сводится к отнесению объекта w к классу c, центр которого μ_c ближе всего к w.

сводится к отнесению объекта w к классу c, центр которого μ_c ближе всего к w.

Задача 2

Исходя из геометрического смысла ROC-AUC кривой (площадь под кривой эквивалентна вер-ти, что классификатор присвоит больший вес случайно выбранной величине из класса "1", чем случайно выбранной величине из класса "0") и случайного выбора элементов, следует что $\forall$ 2 эл-та с равной вероятностью могут входить в разные классы, т.е. с вероятностью $\frac{1}{2}$, а это и означает, что площадь в среднем равна $\frac{1}{2}$.

Задача 3

E_N - усреднение (по $\forall$ обучающим выборкам $T = ((x_1, y_1), \dots, (x_N, y_N))$) ошибки 2-ух классового классификатора, "обученного" методом ближайшего соседа.

Задача: сравнить E_N и E_B. (E_B - вер-ть ошибки байесовского классификатора)

$$E_B = \min(P(0|x), P(1|x))$$

Пусть $(x_n, y_n) \in T$, $(x, y) \perp\!\!\!\perp T$

$$P(Y \neq Y_n) = P(Y = 0, Y_n = 1) + P(Y = 1, Y_n = 0) =$$

$$= P(0|x) \cdot P(1|x_n) + P(1|x) P(0|x_n) \underset{\substack{x_n \to x \\ n \to \infty}}{\approx} 2P(0|x) \cdot P(1|x)$$

$$E_N = P(y \neq y_n) = 2P(0|x) \cdot P(1|x) = 2E_B(1 - E_B) \leq 2E_B \quad \blacksquare$$