hw1 - task 4

Ирина Гавенко

Задача №4.1

Определение наивного байесовского классификатора:
$a(x) = argmax_y P(y) \prod_{k=1}^{n} P(x^{(k)}|y) \Longrightarrow$
из условия классы имеют одинаковые априорные вероятности:
$\Longrightarrow argmax_y \prod_{k=1}^{n} P(x^{(k)}|y)$

Из условия: $P(x^{(k)}|y) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{\frac{z^{(k)}-\mu_{yk}}{2\sigma^2}}$. То есть $P(x^{(k)}|y) \sim N(\mu_{yk}, \sigma)$.

Тогда рассмотрим argmax: максимум функции условной вероятности будет достигаться при минимуме суммы квадратов отклонений: $\rho(x, \mu_y)$.

То есть x отнесем к тому классу, чей центр μ_y будет ближе к нему находиться.

Задача №4.2

В треугольнике ROC-AUC нам заданы две точки: $(0,0), (1,1)$. Рассмотрим, где будет находиться третья точка.

Мы получаем случайные ответы: с вероятностью p – единицу, с вероятностью $(p-1)$ – ноль. То есть биномиальное распределение.

Тогда координата по x третьей точки - это количество единиц (S_1), а по y - количество нулей (S_0).

$E(S_1) = n * p$, где n - общее количество исходов.

Тогда $E(TPR) = \frac{E(S_1) \cdot \frac{S_1}{n}}{S_1}) = \frac{np}{n} = p$

$E(FPR) = \frac{E(S_1) \cdot \frac{S_2}{n}}{S_2} = \frac{np}{n} = p$

Получаем, что третья точка будет лежать на диагоняли, соединяющей $(0,0)$ и $(1,1)$.

Площадь под этой диагональю и, соответсвенно, ROC-AUC равны 0.5.