Машинное обучение. Теоретическое задание 1.

Алёна Егорова, 494 группа

Задание 1.

$\hat{y} = \arg\max\limits_{y} \left(\prod\limits_{i=1}^{n} P(x^{(i)}|y)P(y) \right) = \arg\max\limits_{y} \left(\prod\limits_{i=1}^{n} P(x^{(i)}|y) \right)$

Можно искать максимум прологарифмированной функции:

$\hat{y} = \arg\max\limits_{y} \left(\prod\limits_{i=1}^{n} \ln(P(x^{(i)}|y)) \right)$

Подставим плотность:

$\hat{y} = \arg\max\limits_{y} \left(-\frac{1}{2}\ln(2\pi\sigma^2) - \sum\limits_{i=1}^{n} \frac{(x^{(i)} - \mu_{yi})^2}{2\sigma^2} \right) = \arg\min\limits_{y} \left(\sum\limits_{i=1}^{n} (x^{(i)} - \mu_{yi})^2 \right)$

Заметим, что под суммой стоит расстояние между $x^{(i)}$ и μ_{yi}

Тогда получаем, что $\hat{y} = \arg\min\limits_{y} (\rho(x, \mu_y))$

Задание 2.

Площадь под кривой равна сумме площади треугольника над функцией $y = x$ и большого треугольника ниже $y = x$, где $0 \le x \le 1$. Площадь большого треугольника равняется 0.5 (простые соображения геометрии). Тогда посчитаем площадь меньшего треугольника.

2 вершины меньшего треугольника фиксированы: $(0,0)$ и $(1,1)$. Третья точка выбирается следующим образом: по координате x: $FPR = \frac{\sum\limits_{x \in 0} a(x)}{|\{0\}|}$ (отнесли к классу $\{0\}$). По координате y: $TPR = \frac{\sum\limits_{x \in 1} a(x)}{|\{1\}|}$ (отнесли к классу $\{1\}$).

Через векторное произведение площадь S будет равна $\frac{|\vec{a} - \vec{b}|}{2}$, где $\vec{a} = (FPR - 0, TPR - 0)$ и $\vec{b} = (1 - 0, 1 - 0)$. Тогда $S = \frac{FPR - TPR}{2}$ (детерминант матрицы). Найдем математическое ожидание площади: $ES = \frac{E(FPR) - E(TPR)}{2} = \frac{Ea(x) - Ea(x)}{2} = 0$.

Таким образом, итоговая площадь, как было сказано раньше, сумма двух площадей, а точнее $0 + 0.5 = 0.5$.

Задание 3.

$X-$ объект. $E_b = P($ошибиться в байесовском подходе$)$

Тогда $E_b = 1 - max\{P(0|X), P(1|X)\} = min\{P(0|X), P(1|X)\}$

$E_{1NN} = P($ошибиться в 1NN$)$.

Тогда $E_{1NN} = P(Y_{1NN} = 1|X_{1NN})P(0|X) + P(Y_{1NN} = 0|X_{1NN})P(1|X_{1NN}) \simeq 2P(1|X)P(0|X)$,

т.к. $P(Y|X)$ непрерывен по X $(P(Y_{1NN}|X_{1NN}) \simeq P(Y_{1NN}|X_{1NN}))$

Далее имеем, $2P(1|X)P(0|X) \le 2min\{P(1|X), P(0|X)\} \Rightarrow E_{1NN} \le 2E_b$

Т.к. $max\{P(1|X), P(0|X)\} \ge \frac{1}{2} \Rightarrow E_b \le E_{1NN}$

Получаем, что метод одного ближайшего соседа асимптотически имеет матожидание ошибки не более чем вдвое больше по сравнению с оптимальным байесовским классификатором.