

Машинное обучение. Задание 1

Сопильняк Ольга

27 февраля 2017 г.

4.1 Наивный байес и центроидный классификатор

Наивный байесовский классификатор:

$$a(x) = \operatorname{argmax}_y P(x|y)P(y) = \operatorname{argmax}_y P(y) \prod_{k=1}^n P(x^{(k)}|y)$$

$P(y)$ одинаковы для всех классов, поэтому достаточно найти argmax плотности в точке x .

Так как $P(x^{(k)}|y) = N(\mu_{nk}, \sigma)$, то максимум $\frac{1}{\sqrt{2\pi\sigma^2}}$ может достигаться в точке $x^{(k)} - \mu_{yk} = 0$. Поэтому нужно минимизировать $x^{(k)} - \mu_{yk}$, то есть, рассматривая векторы, получаем задачу

$$\|x - \mu_y\| \rightarrow \min,$$

то есть задача свелась к соотношению объекта x к классу y , центр μ_y которого ближе всего к x .

4.2 ROC-AUC случайных ответов

Заметим, что в данном случае (т.е. в случае случайных ответов с вероятностями p и $1 - p$) ROC будет прямой от $(0, 0)$ до $(1, 1)$, так как если $TPR = a$, то также и $FPR = a$.

Площадь под такой прямой равна 0.5, поэтому $ROC-AUC = 0.5$.