SBT ML2017, Задание 1

Жакиянова Меруерт

April 2017

1 Наивный байес и центроидный классификатор

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

$a(x) = \arg\max_y P(y) \prod_{k=1}^n P(x^{(k)}|y) = \arg\max_y [ln[P(y) \prod_{k=1}^n P(x^{(k)}|y)]] = \arg\max_y [ln[P(y) + \sum_{k=1}^n ln[(\frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{(x^{(k)}-\mu_{yk})^2}{2\sigma^2}))]]] = \arg\max_y [ln[P(y) + \sum_{k=1}^n ln[\frac{1}{\sqrt{2\pi\sigma^2}}] + \sum_{k=1}^n (-\frac{(x^{(k)}-\mu_{yk})^2}{2\sigma^2})] = \arg\max_y [ln[P(y) + nln[\frac{1}{\sqrt{2\pi\sigma^2}}] + \sum_{k=1}^n (-\frac{(x^{(k)}-\mu_{yk})^2}{2\sigma^2})] = \arg\max_y [ln[P(y) + nln[\frac{1}{\sqrt{2\pi\sigma^2}}] - \frac{1}{2\sigma^2} \sum_{k=1}^n (x^{(k)} - \mu_{yk})^2]$

Классы имеют одинаковые априорные вероятности => P(y)=$const$ и $a(x) = \arg\min_y \frac{1}{2\sigma^2} \sum_{k=1}^n (x^{(k)} - \mu_{yk})^2 = \arg\min_y ||x - \mu_y||$

Получаем центроидный классификатор.

2 Ошибка 1NN и оптимального байесовского классификатора

Вероятность ошибки байесовского классификатора:
$E_B = min\{P(1|x); P(0|x)\} = 1 - \max_{y\in\{0,1\}} \mathrm{P(y|x)}$

Пусть $r = \max_{y\in\{0,1\}} \mathrm{P(y|x)}$

Ошибка метода ближайшего соседа $E_{N,n} = P(y \neq y_n)$. При $n \to \infty$ распределение вероятностей классов для ближайшего соседа $x - \{P(y_n|x_n)\}_{y_n=0}^1$ стремится к распределению для x: $\{P(y|x)\}_{y=0}^1$.

$E_{N,n} = |$независимость принадлежностей классов для x и $x_n| = \sum_{y \neq y_n \in\{0,1\}} P(y|x)P(y_n|x_n) \to \sum_{y\in\{0,1\} P(y|x)(1-P(y|x))} = 2r(1-r) \leq 2(1-r) = 2E_B$