

1 Theory

Мы знаем, что $y = f(x) + \varepsilon$, причём $\mathbf{E}(\varepsilon) = 0$, а $\mathbf{Var}(\varepsilon) = \sigma^2$.

X, Y — выборка и x, y — точки из распределения $P(x, y)$.

a — алгоритм, построенный по X и Y .

$$\mathbf{E}_{X,Y,x,y} (a(x) - y)^2 = \mathbf{E}_{X,Y,x,y} (\mathbf{E}_{X,Y,x,y} ((a(x) - y)^2 | x))$$

$$\begin{aligned} \mathbf{E}_{X,Y,x,y} ((a(x) - y)^2 | x) &= \mathbf{E}_{X,Y,x,y} ((a(x) - f(x) - \varepsilon)^2 | x) = \\ &= \mathbf{E}_{X,Y,x,y} ((a(x) - f(x))^2 - 2\varepsilon(a(x) - f(x)) + \varepsilon^2 | x) = \\ &= \mathbf{E}_{X,Y,x,y} ((a(x) - f(x))^2 | x) - 2\mathbf{E}_{X,Y,x,y} (\varepsilon | x) \mathbf{E}_{X,Y,x,y} (a(x) - f(x) | x) + \mathbf{E}_{X,Y,x,y} (\varepsilon^2 | x) = \\ &= \mathbf{E}_{X,Y,x,y} ((a(x) - f(x))^2 | x) - 2\mathbf{E}_{X,Y,x,y} (\varepsilon) \mathbf{E}_{X,Y,x,y} (a(x) - f(x) | x) + \mathbf{E}_{X,Y,x,y} (\varepsilon^2) = \\ &= \mathbf{E}_{X,Y,x,y} ((a(x) - f(x))^2 | x) - 2 \cdot 0 \cdot \mathbf{E}_{X,Y,x,y} (a(x) - f(x) | x) + \sigma^2 = \\ &= \mathbf{E}_{X,Y,x,y} ((a(x) - f(x))^2 | x) + \sigma^2 = \\ &= \mathbf{E}_{X,Y,x,y} ((a(x) - \mathbf{E}_{X,Y,x,y}(a(x)) + \mathbf{E}_{X,Y,x,y}(a(x)) - f(x))^2 | x) + \sigma^2 = \end{aligned}$$

Обозначим $\mathbf{E}_{X,Y,x,y}(a(x))$ за $\bar{a}(x)$

$$\begin{aligned} &= \mathbf{E}_{X,Y,x,y} ((a(x) - \bar{a}(x) + \bar{a}(x) - f(x))^2 | x) + \sigma^2 = \\ &= \mathbf{E}_{X,Y,x,y} ((a(x) - \bar{a}(x)))^2 | x) + \mathbf{E}_{X,Y,x,y} ((\bar{a}(x) - f(x))^2 | x) - \\ &\quad - 2\mathbf{E}_{X,Y,x,y} ((a(x) - \bar{a}(x))) (\bar{a}(x) - f(x) | x) + \sigma^2 = \\ &= \mathbf{E}_{X,Y,x,y} ((a(x) - \bar{a}(x)))^2 | x) + \mathbf{E}_{X,Y,x,y} ((\bar{a}(x) - f(x))^2 | x) - \\ &\quad - 2(\bar{a}(x) - f(x)) \mathbf{E}_{X,Y,x,y} ((a(x) - \bar{a}(x))) | x) + \sigma^2 = \\ &= \mathbf{E}_{X,Y,x,y} ((a(x) - \bar{a}(x)))^2 | x) + \mathbf{E}_{X,Y,x,y} ((\bar{a}(x) - f(x))^2 | x) - 2(\bar{a}(x) - f(x)) (\bar{a}(x) - \bar{a}(x)) + \sigma^2 = \\ &= \mathbf{E}_{X,Y,x,y} ((a(x) - \bar{a}(x)))^2 | x) + (\bar{a}(x) - f(x))^2 + \sigma^2 \\ &= \text{Variance}(x) + \text{Bias}(x)^2 + \sigma^2, \end{aligned}$$

где $\text{Variance}(x) = \mathbf{E}_{X,Y,x,y} ((a(x) - \bar{a}(x)))^2 | x) = \mathbf{E}_{X,Y,x,y} ((a(x) - \bar{a}(x)))^2$,

а $\text{Bias}(x) = \bar{a}(x) - f(x) = \mathbf{E}_{X,Y,x,y} (a(x) - f(x))$

$$\mathbf{E}_{X,Y,x,y} (a(x) - y)^2 = \mathbf{E}_x (\text{Variance}(x) + \text{Bias}(x)^2) + \sigma^2$$

2 kNN

Для kNN, предсказание алгоритма выглядит так:

$$\begin{aligned}
 a(x) &= \frac{1}{k} \sum_{i=1}^k y(N_i(x)) \\
 \mathbf{E}_{X,Y,x,y}(a(x)|x, X, Y) &= \mathbf{E}_{X,Y,x,y} \left(\frac{1}{k} \sum_{i=1}^k y(N_i(x)) | x, X, Y \right) = \\
 &= \mathbf{E}_{X,Y,x,y} \left(\left(\frac{1}{k} \sum_{i=1}^k f(N_i(x)) \right) + \varepsilon_{n_i} \right) | x, X, Y = \frac{1}{k} \sum_{i=1}^k f(N_i(x)) \\
 \mathbf{E}_{X,Y,x,y}(a(x)^2 | x, X, Y) &= \mathbf{E}_{X,Y,x,y} \left(\frac{1}{k^2} \sum_{i,j} y(N_i(x)) y(N_j(x)) \right) | x, X, Y = \\
 &= \mathbf{E}_{X,Y,x,y} \left(\frac{1}{k^2} \sum_{i,j} (f(N_i(x)) f(N_j(x)) + \varepsilon_{n_i} f(N_j(x)) + \varepsilon_{n_j} f(N_i(x)) + \varepsilon_{n_i} \varepsilon_{n_j}) \right) | x, X, Y = \\
 &= \frac{1}{k^2} \sum_{i,j} \left(\mathbf{E}_{X,Y,x,y}(f(N_i(x)) f(N_j(x)) | x, X, Y) + \mathbf{E}_{X,Y,x,y}(\varepsilon_{n_i} f(N_j(x)) | x, X, Y) + \right. \\
 &\quad \left. + \mathbf{E}_{X,Y,x,y}(\varepsilon_{n_j} f(N_i(x)) | x, X, Y) + \mathbf{E}_{X,Y,x,y}(\varepsilon_{n_i} \varepsilon_{n_j} | x, X, Y) \right) = \frac{1}{k^2} \sum_{i,j} \left(f(N_i(x)) f(N_j(x)) + \sigma^2 \{i = j\} \right) = \\
 &= \left(\frac{1}{k} \sum_i f(N_i(x)) \right)^2 + \frac{\sigma^2}{k}
 \end{aligned}$$

Отсюда легко получить, что

$$\begin{aligned}
 Bias &= \mathbf{E}_{X,Y,x} \left(\frac{1}{k} \sum_{i=1}^k f(N_i(x)) - f(x) \right) = \frac{1}{k} \sum_{i=1}^k \mathbf{E}_{X,Y,x} f(N_i(x)) - \mathbf{E}_x f(x) \\
 Variance &= \frac{\sigma^2}{k}
 \end{aligned}$$

3 Композиции

Пусть a_i одинаково распределены.

$$\begin{aligned}
 a(x) &= \frac{1}{k} \sum_{i=1}^k a_i(x) \\
 \mathbf{E}_{X,Y,x,y}(a(x)) &= \frac{1}{k} \mathbf{E}_{X,Y,x,y}(a_i(x)) = \mathbf{E}_{X,Y,x,y}(a_1(x))
 \end{aligned}$$

Поэтому bias не изменится.

$$\mathbf{Var}_{X,Y,x,y}(a(x)^2) = \frac{1}{k^2} \mathbf{Var}_{X,Y,x,y} \left(\sum_{i,j} a_i(x) a_j(x) \right) = \frac{1}{k} \mathbf{Var}_{X,Y,x} a_1(x) + \frac{1}{k^2} \sum_{i \neq j} \mathbf{cov}(a_i(x), a_j(x))$$

Если коэффициент корреляции одинаковый у всех пар и равен r , то

$$\begin{aligned}
 \frac{1}{k} \mathbf{Var}_{X,Y,x} a_1(x) + \frac{1}{k^2} \sum_{i \neq j} \mathbf{cov}(a_i(x), a_j(x)) &= \frac{1}{k} \mathbf{Var}_{X,Y,x} a_1(x) + \frac{1}{k^2} \sum_{i \neq j} r \mathbf{Var}_{X,Y,x} a_1(x) = \\
 &= \left(\frac{1}{k} + \frac{r(k-1)}{k} \right) \mathbf{Var}_{X,Y,x} a_1(x)
 \end{aligned}$$

Таким образом, variance будет тем ниже, что менее коррелированы составляющие.