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.

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

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

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

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

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

$$= \mathbf{E}_{X,Y,x,y}\left(\left(a(x) - f(x)\right)^2 | x\right) - 2 \cdot 0 \cdot \mathbf{E}_{X,Y,x,y}\left(a(x) - f(x) | x\right) + \sigma^2 =$$

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

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

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

$$= \mathbf{E}_{X,Y,x,y}\left(\left(a(x) - \overline{a}(x) + \overline{a}(x) - f(x)\right)^2 | x\right) + \sigma^2 =$$

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

$$-2\mathbf{E}_{X,Y,x,y}\left(\left(a(x) - \overline{a}(x)\right)\right)\left(\overline{a}(x) - f(x)\right) | x\right) + \sigma^2 =$$

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

$$-2\left(\overline{a}(x) - f(x)\right)\mathbf{E}_{X,Y,x,y}\left(\left(a(x) - \overline{a}(x)\right)\right) | x\right) + \sigma^2 =$$

$$= \mathbf{E}_{X,Y,x,y}\left(\left(a(x) - \overline{a}(x)\right)\right)^2 | x\right) + \mathbf{E}_{X,Y,x,y}\left(\left(\overline{a}(x) - f(x)\right)^2 | x\right) - 2\left(\overline{a}(x) - f(x)\right)\left(\overline{a}(x) - \overline{a}(x)\right)\right) + \sigma^2 =$$

$$= \mathbf{E}_{X,Y,x,y}\left(\left(a(x) - \overline{a}(x)\right)\right)^2 | x\right) + \left(\overline{a}(x) - f(x)\right)^2 + \sigma^2$$

$$= Variance(x) + Bias(x)^2 + \sigma^2,$$

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

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

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

2 kNN

Для kNN, предсказание алгоритма выглядит так:

$$a(x) = \frac{1}{k}\sum_{i=1}^{k} y(N_i(x))$$

$$\mathbf{E}_{X,Y,x,y}\left(a(x)|x,X,Y\right) = \mathbf{E}_{X,Y,x,y}\left(\frac{1}{k}\sum_{i=1}^{k} y(N_i(x))|x,X,Y\right) =$$

$$= \mathbf{E}_{X,Y,x,y}\left(\left(\frac{1}{k}\sum_{i=1}^{k} f(N_i(x))\right) + \varepsilon_{n_i}\right)|x,X,Y\right) = \frac{1}{k}\sum_{i=1}^{k} f(N_i(x))$$

$$\mathbf{E}_{X,Y,x,y}\left(a(x)^2|x,X,Y\right) = \mathbf{E}_{X,Y,x,y}\left(\frac{1}{k^2}\sum_{i,j} y(N_i(x))y(N_j(x))\middle|x,X,Y\right) =$$

$$= \mathbf{E}_{X,Y,x,y}\left(\frac{1}{k^2}\sum_{i,j}\left(f(N_i(x))f(N_j(x)) + \varepsilon_{n_i}f(N_j(x)) + \varepsilon_{n_j}f(N_i(x)) + \varepsilon_{n_i}\varepsilon_{n_j}\right)\middle|x,X,Y\right) =$$

$$= \frac{1}{k^2}\sum_{i,j}\Bigg(\mathbf{E}_{X,Y,x,y}\left(f(N_i(x))f(N_j(x))|x,X,Y\right) + \mathbf{E}_{X,Y,x,y}\left(\varepsilon_{n_i}f(N_j(x))|x,X,Y\right) +$$

$$+\mathbf{E}_{X,Y,x,y}\left(\varepsilon_{n_j}f(N_i(x))|x,X,Y\right) + \mathbf{E}_{X,Y,x,y}\left(\varepsilon_{n_i}\varepsilon_{n_j}|x,X,Y\right)\Bigg) = \frac{1}{k^2}\sum_{i,j}\Bigg(f(N_i(x))f(N_j(x)) + \sigma^2\{i=j\}\Bigg) =$$

$$= \left(\frac{1}{k}\sum_{i} f(N_i(x))\right)^2 + \frac{\sigma^2}{k}$$

Отсюда легко получить, что

$$Bias = \mathbf{E}_{X,Y,x}\left(\frac{1}{k}\sum_{i=1}^{k} f(N_i(x)) - f(x)\right) = \frac{1}{k}\sum_{i=1}^{k}\mathbf{E}_{X,Y,x}f(N_i(x)) - \mathbf{E}_x f(x)$$

$$Variance = \frac{\sigma^2}{k}$$

3 Композиции

Пусть a_i одинаково распределены.

$$a(x) = \frac{1}{k}\sum_{i=1}^{k} a_i(x)$$

$$\mathbf{E}_{X,Y,x,y}\left(a(x)\right) = \frac{1}{k}\mathbf{E}_{X,Y,x,y}\left(a_i(x)\right) = \mathbf{E}_{X,Y,x,y}\left(a_1(x)\right)$$

Поэтому bias не изменится.

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

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

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

$$= \left(\frac{1}{k} + \frac{r(k-1)}{k}\right)\mathbf{Var}_{X,Y,x}a_1(x)$$

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