Теоретические задачи

2.1. Bias-varianse-noise decomposition

$$E_{x,y} E_{X^\ell}(y - a_{X^\ell}(x))^2 = E_{x,y} E_{X^\ell}\left(y - E_{X^\ell}[a_{X^\ell}(x)] + E_{X^\ell} a_{X^\ell}(x) - a_{X^\ell}(x)\right) =$$

$$= E_{x,y} E_{X^\ell}(y - E_{X^\ell}[a_{X^\ell}(x)])^2 + E_{x,y} E_{X^\ell} 2\cdot(y - E_{X^\ell}[a_{X^\ell}(x)])(E_{X^\ell}[a_{X^\ell}(x)] - a_{X^\ell}(x)) +$$
$$E_{x,y} E_{X^\ell}(E_{X^\ell}[a_{X^\ell}(x)] - a_{X^\ell}(x))^2;$$

$$E_{x,y} E_{X^\ell} 2(y - E_{X^\ell}[a_{X^\ell}(x)])(E_{X^\ell}[a_{X^\ell}(x)] - a_{X^\ell}(x)) =$$
$$= 2\left(E_{x,y} E_{X^\ell}(y\cdot E_{X^\ell}[a_{X^\ell}(x)]) - E_{x,y}(E_{X^\ell}[a_{X^\ell}(x)])^2 - E_{x,y}(y\cdot E_{X^\ell}[a_{X^\ell}(x)]) + E_{x,y}(E_{X^\ell}[a_{X^\ell}(x)])^2\right) = 0.$$

Тогда

$$E_{x,y} E_{X^\ell}(y - a_{X^\ell}(x))^2 = E_{x,y} E_{X^\ell}(y - E_{X^\ell}[a_{X^\ell}(x)])^2 + Var =$$
$$= E_{x,y} E_{X^\ell}(y - E(y|x) + E(y|x) - E_{X^\ell}[a_{X^\ell}(x)])^2 + Var =$$
$$= E_{x,y}(y - E(y|x))^2 + E_{x,y}(E(y|x) - E_{X^\ell}[a_{X^\ell}(x)])^2 + E_{x,y} 2(y - E(y|x))(E(y|x) -$$
$$- E_{X^\ell}[a_{X^\ell}(x)]) + Var$$

Матожидание произведения:

$$E_{x,y} 2(y - E(y|x))(E(y|x) - E_{X^\ell}[a_{X^\ell}(x)]) = 2 E_{x,y}(y - E(y|x))\cdot E_{x,y}(E(y|x) - E_{X^\ell} a_{X^\ell}(x)) = 0$$

Использована независимость сл. величин и то, что $E(E(y|x) - E_{X^\ell}[a_{X^\ell}(x)]) = 0$

Получили

$$E_{x,y} E_{X^\ell}(y - a_{X^\ell}(x))^2 = E_{x,y}(y - E(y|x))^2 + E_{x,y}(E(y|x) - E_{X^\ell}[a_{X^\ell}(x)])^2 +$$
$$+ E_{x,y} E_{X^\ell}(E_{X^\ell}[a_{X^\ell}(x)] - a_{X^\ell}(x))^2 = Noise^2 + Bias^2 + Var.$$

2.2. Смещение и разброс при бэггинге

$$a(x) = \frac{1}{M}\sum_{m=1}^{M} a_m(x)$$

По условию все базовые алгоритмы распределены одинаково. Найдём смещение алгоритма $a(x)$:

$$\mathrm{Bias}^2 a(x) = E_{x,y}\left(E(y|x) - E_{X^l}[a(x)]\right)^2$$

Выразим $E_{X^l}[a(x)]$ через $a_1, \dots, a_m$:

$$E_{X^l}[a(x)] = E_{X^l}\left(\frac{1}{M}\sum_{m=1}^{M} a_m(x)\right) = \frac{1}{M} M \, E_{X^l}[a_1(x)] = E_{X^l}[a_1(x)] \Rightarrow$$ смещение не изменится.

Найдём разброс нового алгоритма, положив смещение алгоритма равным 0:

$$\mathrm{Var}\, a(x) = E_{x,y} E_{X^l}\left(a(x) - E_{X^l} a(x)\right)^2 = E_{x,y} E_{X^l}\left(a(x) - E_{X^l} a_1(x)\right)^2 = E_{x,y} E_{X^l}[a(x)]^2$$

Заменим $a(x)$ на $\frac{1}{M}\sum_{m=1}^{M} a_m(x)$ в $E_{X^l} a(x)^2$:

$$E_{X^l} a(x)^2 = E_{X^l}\left(\frac{1}{M}\sum_{m=1}^{M} a_m(x)\right)^2 = \frac{1}{M^2} M \cdot E_{X^l} a_1(x)^2 + \frac{1}{M^2}\sum_{i \neq j} E_{X^l} a_i(x) a_j(x) = \frac{1}{M} E_{X^l} a_1(x)^2 + \frac{1}{M^2}\sum_{i \neq j} r_{ij} \, E_{X^l} a_i(x) \cdot E_{X^l} a_j(x).$$

Таким образом,

$$\mathrm{Var}\, a(x) \leq \frac{1}{M}\mathrm{Var}\, a_1(x) + \frac{1}{M^2} M(M-1)\,\mathrm{Var}\, a_1(x) = \mathrm{Var}\, a_1(x)$$

неравенство следует из оценки корреляций алгоритмов ≤ 1. Если хотя бы между несколькими алгоритмами корреляция будет меньше 1, то получим строгое нер-во. Т.е. разброс алгоритма уменьшается при бэггинге.

2.3. Корреляция ответов базовых алгоритмов

Пусть $\xi_1, \dots, \xi_M$ — одинаково распределённые величины, матожидание каждой величины — a, дисперсия — σ^2.

Коэффициент корреляции: $\rho(\xi_i, \xi_j) = \dfrac{\mathrm{cov}(\xi_i, \xi_j)}{\sqrt{D\xi_i}\sqrt{D\xi_j}} = \dfrac{\mathrm{cov}(\xi_i, \xi_j)}{\sigma^2}$.

Найдём дисперсию среднего:

$$D\frac{\sum_{i=1}^{M}\xi_i}{M} = E\left(\frac{\sum_{i=1}^{M}\xi_i}{M} - E\left(\frac{\sum_{i=1}^{M}\xi_i}{M}\right)\right)^2 = \frac{1}{M^2} E\left[\sum_{i=1}^{M}(\xi_i - E\xi_i)\right]^2 = \frac{1}{M^2}\left[\sum_{i=1}^{M} E(\xi_i - E\xi_i)^2 + \right.$$

$$\left. + \sum_{i \neq j} E\left((\xi_i - E\xi_i)(\xi_j - E\xi_j)\right)\right] = \frac{1}{M^2}\left[\sum_{i=1}^{M} D\xi_i + \sum_{i \neq j}\mathrm{cov}(\xi_i, \xi_j)\right] =$$

$$= \frac{1}{M^2}\left(M\sigma^2 + M(M-1)\rho\sigma^2\right) = \rho\sigma^2 + \frac{1}{M}(1-\rho)\sigma^2.$$