Теоретические задачи

2.1 Bias-variance-noise decomposition

$$E_{x,y}E_{X^l}(y-a_{X^l}(x))^2 = E_{x,y}E_{X^l}(y-E_{X^l}a_{X_l}(x)+E_{X_l}a_{X_l}(x)-a_{X_l}(x) =$$

$$E_{x,y}E_{X^l}(y-E_{X^l}a_{X_l}(x))^2+E_{x,y}E_{X^l}2\cdot(y-E_{X^l}a_{X_l}(x))(E_{X_l}a_{X_l}(x)-a_{X_l}(x))+E_{x,y}E_{X^l}(E_{X_l}a_{X_l}(x)-a_{X_l}(x))^2$$

Найдем матожидание произведения:

$$E_{x,y}E_{X^l}2\cdot(y-E_{X^l}a_{X_l}(x))\cdot(E_{X_l}a_{X_l}(x)-a_{X_l}(x)) = 2\cdot(E_{x,y}E_{X^l}(y\cdot E_{X_l}a_{X_l}(x))-E_{x,y}(E_{X^l}a_{X_l}(x))^2-$$

$$-E_{x,y}(y\cdot E_{X_l}a_{X_l}(x)+E_{x,y}(E_{X^l}a_{X_l}(x))^2)=0$$

Тогда

$$E_{x,y}E_{X^l}(y-a_{X^l}(x))^2 = E_{x,y}E_{X^l}(y-E_{X^l}a_{X_l}(x))^2+Var = E_{x,y}E_{X^l}(y-E(y|x)+E(y|x)-E_{X^l}a_{X_l}(x))^2+Var =$$

$$= E_{x,y}(y-E(y|x))^2+E_{x,y}(E(y|x)-E_{X^l}a_{X_l}(x))^2+E_{x,y}2(y-E(y|x))(E(y|x)-E_{X^l}a_{X_l}(x))+Var$$

Матожидание произведения (использована независимость случайных величин и то, что $E(E(y|x)-E_{X^l}a_{X_l}(x))=0$):

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

Получили: $E_{x,y}E_{X^l}(y-a_{X^l}(x))^2 = Noise+Bias^2+Var$

2.2 Смещение и разброс в бэггинге

Ответы базовых алгоритмов распределены одинаково, значит bias разных алгоритмов равны.

$$B_i = E(y-a_i(x))$$

$$B = E(y-\frac{\sum_{i=1}^M a_i(x)}{M}) = \frac{1}{M}\sum_{i=1}^M E(y-a_i(x)) = \frac{1}{M}\sum_{i=1}^M B_i = B_1$$

Значит, bias не изменится.

$$Var_i = E(y-a_i(x))^2 = \sigma^2$$

$$Var = E\left(y-\frac{1}{M}\sum_{i=1}^M a_i(x)\right)^2 = \frac{1}{M^2}E\left(\sum_{i=1}^M(y-a_i(x))\right)^2 = \frac{1}{M^2}\left(\sum_{i=1}^M E(y-a_i(x))^2+\sum_{i\neq j}E(y-a_i(x))(y-a_j(x$$

В случае нескоррелированных деревьев $E(y-a_i(x))(y-a_j(x)) = E(y-a_i(x))\cdot E(y-a_j(x)) = 0$ и $Var = \frac{\sigma^2}{M}$, то есть уменьшится в M раз.

Если деревья скореллированы, воспользуемся неравенством Коши-Буняковского: $E(y-a_i(x))(y-a_j(x)) \le \sqrt{E(y-a_i(x))^2E(y-a_j(x))^2} = \sigma^2$

$$Var \le \frac{1}{M^2}(\sigma^2\cdot M + M(M-1)\sigma^2) = \sigma^2$$

То есть в случае скореллированных деревьев разброс может не измениться.

2.3 Корреляция ответов базовых алгоритмов

Пусть $\xi_1 ... \xi_M$ - одинаково распределённые величины, матожидание каждой величины - a, дисперсия - σ^2. Коэффициент корреляции - $\rho(\xi_i, \xi_j) = \frac{cov(\xi_i, \xi_j)}{\sqrt{D\xi_i} \cdot \sqrt{D\xi_j}} = \frac{cov(\xi_i, \xi_j)}{\sigma^2}$.

Найдём дисперсию среднего:

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

$$+ \sum_{i \neq j} E((\xi_i - E\xi_i)(\xi_j - E\xi_j))) = \frac{1}{M^2}(\sum_{i=1}^M D\xi_i + \sum_{i \neq j} cov(\xi_i, \xi_j)) = \frac{1}{M^2}(M \cdot \sigma^2 + M(M-1) \cdot \rho \cdot \sigma^2) =$$

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