Шеншиной Юлии 491

1. Показать справедливость

$$E_{xy}E_{X^\ell}(y - a_{X^\ell}(x))^2 \overset{?}{=} E_{xy}(y - E(y|x))^2 +$$
$$+ E_{xy}(E(y|x) - E_{X^\ell}a_{X^\ell}(x))^2 + E_{xy}E_{X^\ell}(a_{X^\ell}(x) - E_{X^\ell}a_{X^\ell}(x))^2$$

Рассмотрим $E_{xy}E_{X^\ell}(y - a_{X^\ell}(x))^2 =$

$= E_{xy}E_{X^\ell}(E_{xy}E_{X^\ell}[(a_{X^\ell}(x) - y)^2|X]) \ominus$ положим $y = f(x) + \varepsilon$, $E\varepsilon = 0$, Variance $\varepsilon = \sigma^2 = noise^2$

$$\ominus E_{xy}E_{X^\ell}(E_{xy}E_{X^\ell}[(a_{X^\ell}(x) - f(x) - \varepsilon)^2|X]) =$$
$$= E_{xy}E_{X^\ell}((a_{X^\ell}(x) - f(x))^2|X) - 2E_{xy}E_{X^\ell}[(a_{X^\ell}(x) - f(x))|X] \cdot E_{xy}E_{X^\ell}(\varepsilon|X) +$$

$+ E_{xy}E_{X^\ell}(\varepsilon^2|X) \ominus$ в силу независимости ε от X $\ominus$

$$\ominus E_{xy}E_{X^\ell}[(a_{X^\ell}(x) - f(x))^2|X] - 2 \cdot \overbrace{E_{xy}E_{X^\ell}\varepsilon}^{=0} \cdot E_{xy}E_{X^\ell}[(a_{X^\ell}(x) - f(x))|X] +$$
$$+ \underbrace{E_{xy}E_{X^\ell}\varepsilon^2}_{\sigma^2} = E_{xy}E_{X^\ell}[(a_{X^\ell}(x) - f(x))^2|X] + \sigma^2$$

Рассмотрим отдельно

$$E_{xy}E_{X^\ell}[(a_{X^\ell}(x) - f(x))^2|X] = E_{xy}E_{X^\ell}[(a_{X^\ell}(x) - E_{xy}E_{X^\ell}a_{X^\ell}(x) +$$
$$+ E_{xy}E_{X^\ell}a_{X^\ell}(x) - f(x))^2|X] = E_{xy}E_{X^\ell}[(a_{X^\ell}(x) - E_{xy}E_{X^\ell}a_{X^\ell})^2|X] -$$
$$- 2 \cdot E_{xy}E_{X^\ell}[(a_{X^\ell}(x) - E_{xy}E_{X^\ell}a_{X^\ell}(x))|X] \cdot E_{xy}E_{X^\ell}[(a_{X^\ell}(x) - f(x))|X] +$$
$$+ E_{xy}E_{X^\ell}[(a_{X^\ell}(x) - f(x))^2|X] = Variance(a_{X^\ell}) + Bias^2(a_{X^\ell})$$

Тогда, возвращаясь к началу задачи, имеем:

$$E_{xy}E_{X^\ell}(y - a_{X^\ell}(x))^2 = Variance + Bias^2 + Noise^2$$

Леншиной Юлии 491

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

Считать, что ответы всех базовых алгоритмов распределены одинаково

$$Bias = E_{xy} E_{x^e}(a(x)) = E_{xy} E_{x^e}\left(\frac{1}{M}\sum_{m=1}^{M} a_m(x)\right) = \frac{1}{M} E_{xy} E_{x^e}\left(\sum_{m=1}^{M} a_m(x)\right) =$$

$$= \frac{1}{M}\sum_{m=1}^{M}\left(E_{xy} E_{x^e}(a_m(x))\right) = \frac{1}{M}\cdot M \cdot E_{xy} E_{x^e}\, a_1(x) \ominus$$

в силу одинаковой распределения

$\ominus E_{xy} E_{x^e}\, a_1(x) \Rightarrow$ смещение не изменится

$$Variance = E_{xy} E_{x^e}(a(x)^2) = \frac{1}{M^2} E_{xy} E_{x^e}\left(\sum_{i,j} a_i(x)\cdot a_j(x)\right) =$$

$$= \frac{1}{M^2} E_{xy} E_{x^e}\left(\underbrace{\sum_{i=j} a_i(x) a_j(x)}_{"a_i(x)^2} + \sum_{i\neq j} a_i(x) a_j(x)\right) =$$

$$= \frac{1}{M}\underbrace{E_{xy} E_{x^e}\, a_1^2(x)}_{Variance(a_1)} + \frac{1}{M^2}\sum_{i\neq j} cov(a_i(x), a_j(x)) =$$

$$= \frac{\sigma^2}{M} + \frac{1}{M^2}\sum_{i\neq j} cov(a_i(x), a_j(x)), \text{ где } \sigma^2 = Variance(a_1)$$

$\Rightarrow$ при уменьшении корреляции между деревьями уменьшается и разброс.

Шеншиной Юлии 491

2.3. Пусть $\{x_i\}_{i=1}^{M}$ – одинаково распределенные случайные величины.

$$D(\overline{x}) = D\left(\frac{1}{M}\sum_{i=1}^{M} x_i\right) = \frac{1}{M^2} E\left(\sum_{i=1}^{M} x_i - E\sum_{i=1}^{M} x_i\right)^2 =$$

$$= \frac{1}{M^2} E\left(\sum_{i=1}^{M}(x_i - Ex_i)\right)^2 =$$

$$= \frac{1}{M^2}\left(\sum_{i=j}^{M} E(x_i - Ex_i)^2 + \sum_{i\neq j}^{M} E(x_i - Ex_i)(x_j - Ex_j)\right) =$$

$$= \frac{1}{M^2}\left(\sum_{i=1}^{M} Dx_i + \sum_{i\neq j}^{M} cov(x_i, x_j)\right) =$$

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

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