

# Машинное обучение. Задание 3

Сопильняк Ольга

28 марта 2017 г.

## 2.2 Смещение и разброс в бэггинге

Пусть композиция строится с помощью бэггинга:

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

Выясним, как соотносятся смещение и разброс для композиции с теми же параметрами для базовых алгоритмов.

$$E_{X,Y,x,y} a(x) = \frac{1}{M} M E_{X,Y,x,y} a_1(x) = E_{X,Y,x,y} a_1(x)$$

— смещение для одного алгоритма.

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

— разброс, где  $\rho = \text{cov}(a_i(x), a_j(x))$

Заметим, что коэффициент перед разбросом определяет соотношение смещения и разброса. Тогда при убывании  $\rho$   $\text{Var}$  также убывает, поэтому чем больше элементы коррелированы, тем больше разброс.

## 2.3 Корреляция ответов базовых алгоритмов

Пусть есть  $M$  одинаково распределенных случайных величин с дисперсией  $\sigma^2$ , любые две из которых имеют положительную корреляцию  $\rho$ .

Покажем, что дисперсия их среднего равна

$$\begin{aligned}\rho\sigma^2 + (1 - \rho)\frac{\sigma^2}{M} \\ cov(\xi_i, \xi_j) = \rho\sigma^2, \quad i \neq j \\ cov(\xi_i, \xi_i) = \sigma^2\end{aligned}$$

$$\begin{aligned}D\left(\frac{\xi_1 + \dots + \xi_M}{M}\right) &= \frac{1}{M^2} cov(\xi_1 + \dots + \xi_M, \xi_1 + \dots + \xi_M) = \\ &= \frac{1}{M^2} \sum_{i,j=1}^M cov(\xi_i, \xi_j) = \frac{1}{M^2} (M\sigma^2 + M(M-1)\rho\sigma^2) = \\ &= \frac{\sigma^2}{M} + \frac{M-1}{M}\rho\sigma^2 = \frac{\sigma^2}{M} + \rho\sigma^2 - \frac{\rho\sigma^2}{M} = \rho\sigma^2 + (1 - \rho)\frac{\sigma^2}{M}\end{aligned}$$

Получили искомое выражение.