2 Теоретические задачи

2.1 Bias-variance-noise decomposition

Пусть $y = f(x) + \epsilon$, где ϵ - шум, независящий от x, $E\epsilon = 0, D\epsilon = \sigma^2 = E(y - f(x))^2$, $f(x) = E(y|x)$. Хотим по выборке X получить функцию $a_X(x) = f(x)$. Найдем матожидание ошибки (MSE), т.е. $E(y - a_X(x))^2$.

$$
\begin{aligned}
E((y - a_X(x))^2|x) = \\
&= E(E(f(x) - \epsilon - a_X(x))^2|x) = \\
&= E((a_X(x) - f(x))^2 - 2 \cdot \epsilon \cdot (a_X(x) - f(x)) + \epsilon^2|x) = \\
&= E((a_X(x) - f(x))^2|x) - 2 \cdot E(\epsilon) \cdot (a_X(x) - f(x)) + E(\epsilon^2) = \\
&= E((a_X(x) - f(x))^2|x) + \sigma^2 = \\
&= E((a_X(x) - f(x) + (Ea_X(x) - Ea_X(x)))^2|x) + \sigma^2 = \\
&= E((a_X(x) - Ea_X(x))^2 + (f(x) - Ea_X(x))^2 - \\
&\quad - 2 \cdot (a_X(x) - Ea_X(x)) \cdot (f(x) - Ea_X(x))|x) + \sigma^2 = \\
&= (a_X(x) - Ea_X(x))^2 + (f(x) - Ea_X(x))^2 + \sigma^2 = \\
&= (a_X(x) - Ea_X(x))^2 + (E(y|x) - Ea_X(x))^2 + E(y - E(y|x))^2
\end{aligned}
$$

$$
\begin{aligned}
E(y - a_X(x))^2 &= E(E(y - a_X(x))^2|x)) = \\
&= E(a_X(x) - Ea_X(x))^2 + E(E(y|x) - Ea_X(x))^2 + E(y - E(y|x))^2 = \\
&= variance + bias + noise
\end{aligned}
$$

2.2 Смещение и разброс в бэггинге

Для композиции с одинаковыми параметрами: пусть $a_m(x)$ - ответ m-го алгоритма на объекте x. В предположении, что a_m - одинаково распределены, $m \in \overline{1, M}$.

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

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

Поэтому bias не изменится от использования беггинга.

$$
\begin{aligned}
D_{X,Y,x,y}(a(x)) &= \frac{1}{M^2} D\left(\sum_{m=1}^{M} a_m(x)\right) = \\
&= \frac{1}{M^2}\left(M \cdot D(a_1(x)) + \sum_{i \neq j} cov(a_i(x), a_j(x))\right)
\end{aligned}
$$

Если коэффициент корреляции одинаковый (а раз параметры совпадают, это логично) и равен ρ.

$$\begin{aligned}D_{X,Y,x,y}(a(x)) =&\frac{1}{M^2}\left(M\cdot D(a_1(x)) + M\cdot(M-1)\cdot\rho\cdot D(a_1(x)\right) =\\ =&\left(\rho+\frac{1-\rho}{M}\right)\cdot D(a_1(x))\end{aligned}$$

Поэтому чем меньше ρ, тем меньше variance.

2.3 Корреляция ответов базовых алгоритмов

Пусть $X_1, X_2, \cdots, X_M$ - одинаково распределенные случайные величины с дисперсией σ^2, любые две из которых имеют положительную корреляцию ρ. Посчитаем дисперсию их среднего.

$$\begin{aligned}D\left(\frac{\sum_{i=1}^M X_i}{M}\right) =&\frac{1}{M^2}\cdot cov\left(\sum_{i=1}^M X_i, \sum_{i=1}^M X_i\right) =\\ =&\frac{1}{M^2}\left(\sum_{i=1}^M D(X_i) + \sum_{i\neq j} cov(X_i, X_j)\right) =\\ =&\frac{1}{M^2}\left(M\cdot\sigma^2 + M\cdot(M-1)\cdot\rho\cdot\sigma^2\right) =\\ =&\frac{\sigma^2}{M}+\frac{M-1}{M}\cdot\rho\cdot\sigma^2 =\\ =&\rho\cdot\sigma^2+\frac{1-\rho}{M}\cdot\sigma^2\end{aligned}$$