whybaggingworks,numericprediction

Alessia

2023-06-16

Let each (y,$\mathbf{x}$) case in L be independently drawn from the probability distribution P. Suppose y is numerical and $\Phi(\mathbf{x}, L)$ the predictor. The aggregated predictor is $\Phi_A(\mathbf{x}, P) = E_L\Phi(x, L)$. Now assume $z = y - \Phi(\mathbf{x}, L)$, apply the inequality $E(Z)^2 \geq (EZ)^2$ and take the expectation over L(mean squared error):

$$E_L(y - \Phi(\mathbf{x}, L))^2 \geq (E_L(y - \Phi(\mathbf{x}, L))^2$$

Now we take the $E_{y,\mathbf{x}}$ of both members of the inequality and so we get average prediction error:

1. $E_{y,\mathbf{x}}[E_L(y - \Phi(\mathbf{x}, L))^2] = E_{y,\mathbf{x}}E_L(y^2 - 2y\Phi(\mathbf{x}, L) + \Phi^2(\mathbf{x}, L)) = E_{y,x}E_Ly^2 - 2E_{y,\mathbf{x}}E_Ly\Phi(\mathbf{x}, L) + E_{y,\mathbf{x}}E_L\Phi^2(\mathbf{x}, L) = E_{y,\mathbf{x}}y^2 - 2E_{y,\mathbf{x}}y\Phi_A + E_{y,\mathbf{x}}E_L\Phi^2(x, L)$
2. $E_{y,\mathbf{x}}(E_L(y - \Phi(\mathbf{x}, L))^2 = E_{y,\mathbf{x}}(y - \Phi_A)^2 = E_{y,\mathbf{x}}(y^2 - 2y\Phi_A + \Phi_A^2) = E_{y,\mathbf{x}}y^2 - 2E_{y,\mathbf{x}}\Phi_A + E_{y,\mathbf{x}}\Phi_a^2$

Substituting this two values in the inequality we get:

$$E_{y,\mathbf{x}}y^2 - 2E_{y,\mathbf{x}}y\Phi_A + E_{y,\mathbf{x}}E_L\Phi^2(x, L) \geq (E_{y,\mathbf{x}}y^2 - 2E_{y,\mathbf{x}}\Phi_A + E_{y,\mathbf{x}}\Phi_A^2)$$

So how much lower the mean squared error of Φ_A is depends on how unequal the two sides of

$$[E_L\Phi(\mathbf{x}, L)]^2 \leq [E_L\Phi^2(x, L)]$$

are. The effect of instability is clear;

- if $\Phi(x, L)$ does not change too much with replicate L the two sides will be nearly equal, and aggregation will not help;
- the more highly variable the $\Phi(x, L)$ are, the more improvement may produce. But $\Phi_A(x, L)$ always improves on Φ.

To clarify better this last point we have to remind that if the predictors remain constant, then the expected value will also remain constant over repeated observations. In this case, both sides of the inequality will be equal because the expected value squared will be the same as the expected value of the squared predictors. Instead, when the predictors are highly variable the square of the mean might understimate the average value of the squared predictors because it doesn't account for the variability.