Cumulative Sum

1 Mathematical Definition

We know that, for general random forests, the left-side error is

$$E =$$
$$\sum_{i=1}^{n}(y_i - \frac{\sum_{i=1}^{n} y_i}{n})^2 =$$
$$\sum_{i=1}^{n}(y_i - \mu_n)^2 =$$
$$\sum_{i=1}^{n} y_i^2 - 2y_i\mu_n + \mu_n^2 =$$
$$\sum_{i=1}^{n} y_i^2 - 2\mu_n \sum_{i=1}^{n} y_i + n\mu_n^2 =$$
$$\sum_{i=1}^{n} y_i^2 - 2\mu_n n\mu_n + n\mu_n^2 =$$
$$\sum_{i=1}^{n} y_i^2 - n\mu_n^2 =$$
$$\sum_{i=1}^{n} y_i^2 - \frac{\sum_{i=1}^{n} y_i}{n}$$

Which is useful because it can easily be calculated via cumsum.

However, for our general forest, the left-side error also takes into account the previous predictions. We call the previous mean prediction for value i o_i, and let there be α other predictions included in the mean. Thus,

$$
\begin{aligned}
E &= \\
\sum_{i=1}^{n}(y_i - \frac{\alpha o_i + \frac{1}{n}\sum_{i=1}^{n} y_i}{\alpha+1})^2 &= \\
\sum_{i=1}^{n}(y_i - \frac{\alpha o_i + \mu_n}{\alpha+1})^2 &= \\
\sum_{i=1}^{n}(y_i - \frac{\alpha o_i + \mu_n}{\alpha+1})^2 &= \\
\sum_{i=1}^{n} y_i^2 - \frac{2\alpha y_i o_i + 2y_i\mu_n}{\alpha+1} + \frac{(\alpha o_i + \mu_n)^2}{(\alpha+1)^2} &= \\
\sum_{i=1}^{n} y_i^2 - \frac{2\alpha}{\alpha+1}\sum_{i=1}^{n} y_i o_i - \frac{2}{\alpha+1} n\mu_n^2 + \frac{\alpha^2}{(\alpha+1)^2}\sum_{i=1}^{n} o_i^2 + \frac{2\alpha}{(\alpha+1)^2}\mu_n \sum_{i=1}^{n} o_i + \frac{\sum_{i=1}^{n} u_n^2}{(\alpha+1)^2} &= \\
\sum_{i=1}^{n} y_i^2 - \frac{2\alpha}{\alpha+1}\sum_{i=1}^{n} y_i o_i - \frac{2}{\alpha+1} n\mu_n^2 + \frac{\alpha^2}{(\alpha+1)^2}\sum_{i=1}^{n} o_i^2 + \frac{2\alpha}{(\alpha+1)^2}\mu_n \sum_{i=1}^{n} o_i + \frac{n u_n^2}{(\alpha+1)^2} &= \\
\sum_{i=1}^{n} y_i^2 - \frac{2\alpha}{\alpha+1}\sum_{i=1}^{n} y_i o_i - \frac{2}{\alpha+1} n\mu_n^2 + \frac{\alpha^2}{(\alpha+1)^2}\sum_{i=1}^{n} o_i^2 + \frac{2\alpha}{(\alpha+1)^2}\mu_n \sum_{i=1}^{n} o_i + \frac{n\frac{(\sum_{i=1}^{n} y_i)^2}{n^2}}{(\alpha+1)^2} &= \\
\sum_{i=1}^{n} y_i^2 - \frac{2\alpha}{\alpha+1}\sum_{i=1}^{n} y_i o_i - \frac{2}{\alpha+1} n\mu_n^2 + \frac{\alpha^2}{(\alpha+1)^2}\sum_{i=1}^{n} o_i^2 + \frac{2\alpha}{(\alpha+1)^2}\mu_n \sum_{i=1}^{n} o_i + \frac{(\sum_{i=1}^{n} y_i)^2}{n(\alpha+1)^2} &= \\
\sum_{i=1}^{n} y_i^2 - \frac{2\alpha}{\alpha+1}\sum_{i=1}^{n} y_i o_i - \frac{2(\sum_{i=1}^{n} y_i)^2}{n(\alpha+1)} + \frac{\alpha^2}{(\alpha+1)^2}\sum_{i=1}^{n} o_i^2 + \frac{2\alpha}{(\alpha+1)^2}\mu_n \sum_{i=1}^{n} o_i + \frac{(\sum_{i=1}^{n} y_i)^2}{n(\alpha+1)^2} &= \\
\sum_{i=1}^{n} y_i^2 - \frac{2\alpha}{\alpha+1}\sum_{i=1}^{n} y_i o_i - \frac{2(\sum_{i=1}^{n} y_i)^2}{n(\alpha+1)} + \frac{\alpha^2}{(\alpha+1)^2}\sum_{i=1}^{n} o_i^2 + \frac{2\alpha\sum_{i=1}^{n} y_i}{n(\alpha+1)^2}\sum_{i=1}^{n} o_i + \frac{(\sum_{i=1}^{n} y_i)^2}{n(\alpha+1)^2} &= \\
(Y^2)_n - \frac{2\alpha}{\alpha+1}(YO)_n - \frac{2}{n(\alpha+1)}Y_n^2 + \frac{\alpha^2}{(\alpha+1)^2}O_n + \frac{2\alpha}{n(\alpha+1)^2}Y_n O_n + \frac{1}{n(\alpha+1)^2}Y_n^2 &=
\end{aligned}
$$

Thus, though it is considerably more complicated, we can still calculate everything we need with the following cumsums:

$$\sum_{i=1}^{n} y_i$$
$$\sum_{i=1}^{n} o_i$$
$$\sum_{i=1}^{n} o_i y_i$$
$$\sum_{i=1}^{n} y_i^2$$
$$\sum_{i=1}^{n} o_i^2$$

What if we instead use gradient descent? Instead of setting the predictions to μ_n, can we dynamically calculate it?

Let

$$E = \sum_{i=1}^{n} (y_i - \frac{\alpha o_i + \beta_n}{\alpha + 1})^2$$

Then, we can calculate the best β_n via gradient, to get:

$$0 = \sum_{i=1}^{n} y_i - \frac{\alpha o_i + \beta_n}{\alpha + 1} \implies$$
$$\sum_{i=1}^{n} y_i - \frac{\alpha}{\alpha + 1} \sum_{i=1}^{n} o_i = \frac{n}{\alpha + 1} \beta_n \implies$$
$$(\alpha + 1)\mu_n - \frac{\alpha}{n} \sum_{i=1}^{n} o_i = \beta_n$$

We represent the true value of $\phi = \frac{(1-\epsilon)Y_n}{n} + \epsilon\beta_n$, which represents some ϵ-importance of the other predictions to the value of the split.

$$E =$$

$$\sum_{i=1}^{n}\left(y_i - \frac{\alpha o_i + \frac{(1-\epsilon)Y_n}{n} + \epsilon\beta_n}{\alpha+1}\right)^2 =$$

$$\sum_{i=1}^{n}\left(y_i - \frac{\alpha o_i + \frac{(1-\epsilon)Y_n}{n} + \epsilon\frac{Y_n(\alpha+1)}{n} - \epsilon\frac{\alpha O_n}{n}}{\alpha+1}\right)^2 =$$

$$\sum_{i=1}^{n}\left(y_i - \frac{\alpha o_i + \frac{Y_n - \epsilon Y_n + \epsilon\alpha Y_n + \epsilon Y_n}{n} - \epsilon\frac{\alpha O_n}{n}}{\alpha+1}\right)^2 =$$

$$\sum_{i=1}^{n}\left(y_i - \frac{\alpha o_i + \frac{Y_n + \epsilon\alpha Y_n}{n} - \epsilon\frac{\alpha O_n}{n}}{\alpha+1}\right)^2 =$$

$$\sum_{i=1}^{n}\left(\left(y_i - \frac{\alpha}{\alpha+1}o_i\right) - \frac{\frac{Y_n + \epsilon\alpha Y_n}{n} - \epsilon\frac{\alpha O_n}{n}}{\alpha+1}\right)^2 =$$

$$\sum_{i=1}^{n}\left(y_i - \frac{\alpha}{\alpha+1}o_i\right)^2 - 2\frac{\frac{Y_n + \epsilon\alpha Y_n}{n} - \epsilon\frac{\alpha O_n}{n}}{\alpha+1}\left(y_i - \frac{\alpha}{\alpha+1}o_i\right) + \left(\frac{\frac{Y_n + \epsilon\alpha Y_n}{n} - \epsilon\frac{\alpha O_n}{n}}{\alpha+1}\right)^2$$

We can define

$$A_i = y_i - \frac{\alpha}{\alpha+1}o_i$$

$$B_n = \frac{Y_n + \epsilon\alpha Y_n - \epsilon\alpha O_n}{\alpha+1} = \frac{(1+\epsilon\alpha)\sum_{i=1}^{n} y_i - \epsilon\alpha\sum_{i=1}^{n} o_i}{\alpha+1}$$

Then,

$$E =$$

$$\sum_{i=1}^{n} A_i^2 - 2\frac{1}{n}B_n\sum_{i=1}^{n} A_i + \frac{B_n^2}{n}$$

All of which can be calculated with cumulative sums.

$$
\begin{aligned}
&E = \\
&\sum_{i=1}^{n} (y_i - \frac{\alpha o_i + \beta_n}{\alpha+1})^2 = \\
&\sum_{i=1}^{n} (y_i - \frac{\alpha o_i + \mu_n - \frac{\epsilon\alpha}{n(\alpha+1)} \sum_{j=1}^{n} o_j}{\alpha+1})^2 = \\
&\sum_{i=1}^{n} y_i^2 - \frac{2\alpha y_i o_i + 2 y_i \mu_n - \frac{2 y_i \epsilon\alpha}{n(\alpha+1)} \sum_{j=1}^{n} o_j}{\alpha+1} + \frac{\left(\alpha o_i + \mu_n - \frac{\epsilon\alpha}{n(\alpha+1)} \sum_{i=1}^{n} o_i\right)^2}{(\alpha+1)^2} = \\
&(Y^2)_N - \frac{2\alpha}{\alpha+1}(YO)_N - \frac{2}{n(\alpha+1)} Y_N^2 + \frac{2\epsilon\alpha}{n(\alpha+1)^2} Y_N O_N + \frac{1}{n(\alpha+1)^2} Y_N^2 + \frac{\alpha^2}{(\alpha+1)^2}(O^2)_n + \\
&\frac{2\alpha}{n(\alpha+1)^2} Y_n O_n + \sum_{i=1}^{n} \frac{\epsilon^2\alpha^2}{n^2(\alpha+1)^4} O_n^2 - \sum_{i=1}^{n} \frac{2\alpha^2 \epsilon o_i O_n}{n(\alpha+1)^3} - \sum_{i=1}^{n} \frac{2\alpha\epsilon}{n^2(\alpha+1)^3} Y_n O_n = \\
&(Y^2)_N - \frac{2\alpha}{\alpha+1}(YO)_N - \frac{2}{n(\alpha+1)} Y_N^2 + \frac{2\epsilon\alpha}{n(\alpha+1)^2} Y_N O_N + \frac{1}{n(\alpha+1)^2} Y_N^2 + \frac{\alpha^2}{(\alpha+1)^2}(O^2)_n + \\
&\frac{2\alpha}{n(\alpha+1)^2} Y_n O_n + \frac{\epsilon^2\alpha^2}{n(\alpha+1)^4} O_n^2 - \frac{2\alpha^2\epsilon}{n(\alpha+1)^3} O_n^2 - \frac{2\alpha\epsilon}{n(\alpha+1)^3} Y_n O_n = \\
&(Y^2)_n - \frac{2\alpha}{\alpha+1}(YO)_n + \frac{1}{n(\alpha+1)} Y_n^2 \left(\frac{1}{\alpha+1} - 2\right) + \frac{\alpha^2}{(\alpha+1)^2}(O^2)_n + \\
&\frac{\epsilon\alpha^2}{n(\alpha+1)^3} O_n^2 \left(\frac{\epsilon}{\alpha+1} - 2\right) + \frac{2\alpha}{n(\alpha+1)^2} Y_n O_n \left(\epsilon + 1 - \frac{\epsilon}{\alpha+1}\right)
\end{aligned}
$$

Thus, the final error function as as follows:

$$
\begin{aligned}
E_n = (Y^2)_n - \frac{2\alpha}{\alpha+1}(YO)_n + \frac{1}{n(\alpha+1)} Y_n^2 \left(\frac{1}{\alpha+1} - 2\right) + \frac{\alpha^2}{(\alpha+1)^2}(O^2)_n + \\
\frac{\epsilon\alpha^2}{n(\alpha+1)^3} O_n^2 \left(\frac{\epsilon}{\alpha+1} - 2\right) + \frac{2\alpha}{n(\alpha+1)^2} Y_n O_n \left(\epsilon + 1 - \frac{\epsilon}{\alpha+1}\right)
\end{aligned}
$$

While seemingly complicated, this can easily be calculated with pre-calculated cumulative sums. It is also useful to precaclulate $\frac{1}{\alpha+1}$ given its common usage.