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,

$$E =$$

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

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

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

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

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

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

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

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

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

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

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:

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

Note that this value is the original non-gradient value of β_n, μ_n, minus some value. Thus, we can utilize a learning rate, ϵ, to get that

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

We must recalculate the original error calculation