

Numerai has a few high-impact ideas that we would like to immediately disseminate to the community to give you all some time to consider and comment. The topics are as follows:

- Better LGBM Parameters
- Signals V2 Data
- Reducing Signals Churn

Better LGBM Parameters

The [Atlas Data Release](#) brought with it new validation example predictions that can achieve a correlation Sharpe of 2 on our validation dataset. The way we achieved this performance was with the following LGBM parameters:

```
{ "n_estimators": 30000, "learning_rate": 0.001, "max_depth": 10, "num_leaves": 1024, "colsample_bytree": 0.1, "min_data_in_leaf": 10000 }
```

After grid-searching several parameters, we found that `max_depth` and `min_data_in_leaf` had the most impact on improving model performance. You may be able to get better results with your own modeling techniques. When we start accepting Atlas data submissions on September 13, we will also update our examples and tutorials to use these “deep” parameters so that new users can start at a higher level.

Signals V2 Data

Numerai is planning to release Signals V2 Data this quarter. V2 Data will significantly increase the size of the Signals Universe - a similar update to the Atlas release, but less of a breaking change than Atlas data due to the lack of encryption / obfuscation.

We don't currently have a timeline on when this universe expansion will occur as it's still in R&D, but we hope to have it released sometime in September. Something to note here is that the features and targets will change with the universe so it will be a good idea to retrain your models for the new Signals universe as soon as possible.

Reducing Signals Churn

Churn is a statistic describing how the alpha scores of a signal changes over time. If a submission to Numerai Signals has high churn, then Numerai can't trade the signal easily. We added a churn statistic to Signals Diagnostics in June 2023 (see [here](#) for details) to help users reduce their churn. Many models built on Numerai data have low churn organically, but Signals Churn is very high. Most Signals models have > 20% week-over-week Churn:

[
1600×831 90.4 KB
(<https://forum.numer.ai/uploads/default/original/2X/f/fd2f4b5037418174f62f9035165ed7e52b87e0bd.jpeg>)

We know that this negatively impacts the churn of the Signals Meta Model because the average individual churn of Signals models is nearly 70% correlated with the Signals Meta Model Churn:

[
946×162 9.6 KB
(<https://forum.numer.ai/uploads/default/original/2X/7/7190614cc1dd271bb5894014c357e6496d626a9c.png>)

[
1600×819 106 KB
(<https://forum.numer.ai/uploads/default/original/2X/6/626fa6a03bbfc4514b3ba7db6514203a51df3d87.jpeg>)

Thus the Signals Meta Model's churn is too high to be useful to Numerai:

[
1600×1130 166 KB
(<https://forum.numer.ai/uploads/default/original/2X/6/6cb772191887ad246911a4ff1d18ff81e73bfc73.png>)

Signals Meta Model churn is too high.

To lower the churn of the Signals Meta Model, we must lower the churn of all Signals - so we are implementing a strict churn threshold that operates as follows:

- Any model that has not submitted in the previous week will earn 0 payouts
- Any model that does not submit weekly will naturally cause high churn in the Meta Model
- Any model that does not submit weekly will naturally cause high churn in the Meta Model
- When you upload a new submission we:
 - Calculate churn with respect to each of this model's submissions from the previous week
 - Check if this submission has $\geq X\%$ churn with respect to any of this model's accepted submission in the previous week, then this submission will not be accepted - the API will return an error such as "Error: Churn is too high. Submission has $Y\%$ churn from [DATE]"
 - Calculate churn with respect to each of this model's submissions from the previous week
 - Check if this submission has $\geq X\%$ churn with respect to any of this model's accepted submission in the previous week, then this submission will not be accepted - the API will return an error such as "Error: Churn is too high. Submission has $Y\%$ churn from [DATE]"

We are going to roll this out over a 2 week:

1. Start enforcing a threshold of 30% on September 6, 2024
2. Reduce the threshold to 20% on September 13, 2024
3. Finally, set the threshold to 10% on September 20, 2024

You'll notice both our v43.cyrus_plus_teager model and Numerai Meta Model have breached the 10% churn threshold only twice, so we know this is an achievable level of churn that will guarantee a sufficient reduction in overall Signals Meta Model churn.

FAQs

How do I know what my churn is?

We will be open-sourcing the churn calculation we use in diagnostics so that you can calculate it yourself. Then, we will use this calculation to display the new statistic on the Signals website. Any submissions that breaches the threshold will be highlighted as "high churn" in the website.

When will this churn threshold take effect?

On September 6, 2024 the threshold will be 30%. On September 13, 2024 the threshold will be 20%. On September 20, 2024 the threshold will be 10%.

What about Numerai models?

This does not affect Numerai models as they cannot control their churn level due to the obfuscation of the dataset. Instead, we have crafted a dataset that naturally results in lower-churn models. Signals, on the other hand, can easily reduce their churn because Signals models can easily calculate it. Signals models can be trained to minimize churn just the way we did with our v43.cyrus_plus_teager model.

What if everyone breaches the threshold?

This is exceptionally improbable as there will always be a niche for reliably low Signals models. Furthermore, this new mechanic, any round with a large number of models that breach the churn threshold will have a higher payout factor - thus rewarding reliably low churn models.