Evaluation

This competition evaluates predicted spectra (μ_user) and corresponding uncertainties (σ_user) for different wavelengths against the ground truth pixel level spectrum (y) using using the Gaussian Log-likelihood (GLL) function.

$$GLL = -\frac{1}{2}\left(\log(2\pi) + \log(\sigma_{user}^2) + \frac{(y-\mu_{user})^2}{\sigma_{user}^2}\right)$$

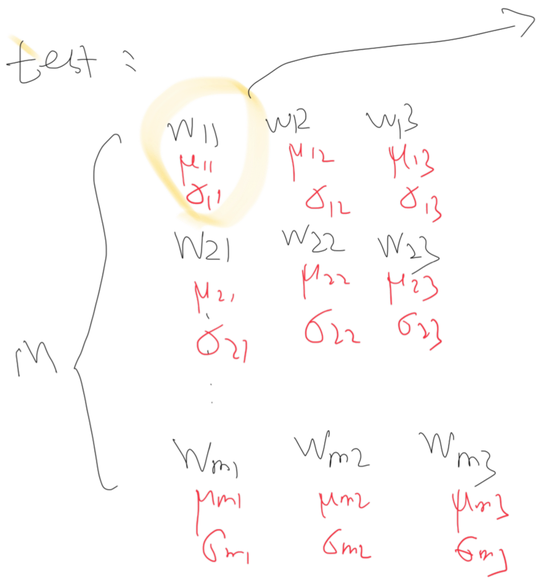
The GLL values from each pair will be summed across all wavelengths and across the entire test set to produce a final GLL value (L). The final GLL value will be transformed into a score using the following conversion function:

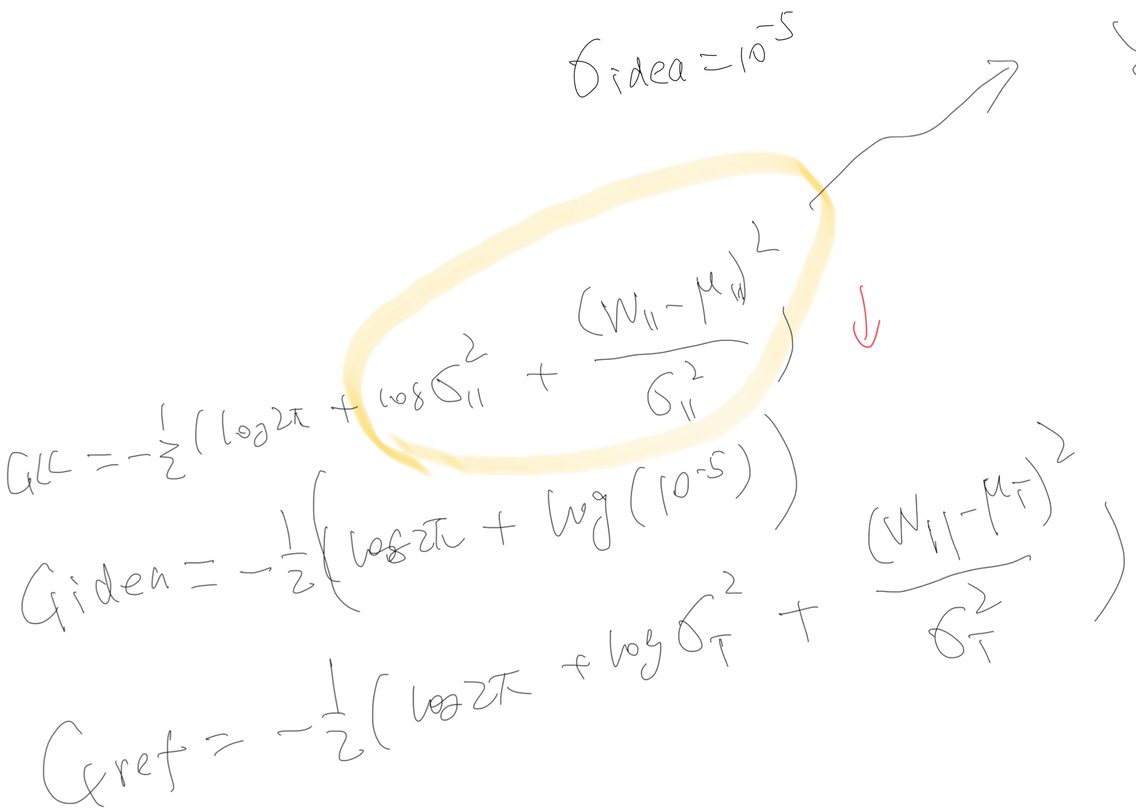
$$score = \frac{L - L_{ref}}{L_{ideal} - L_{ref}}$$

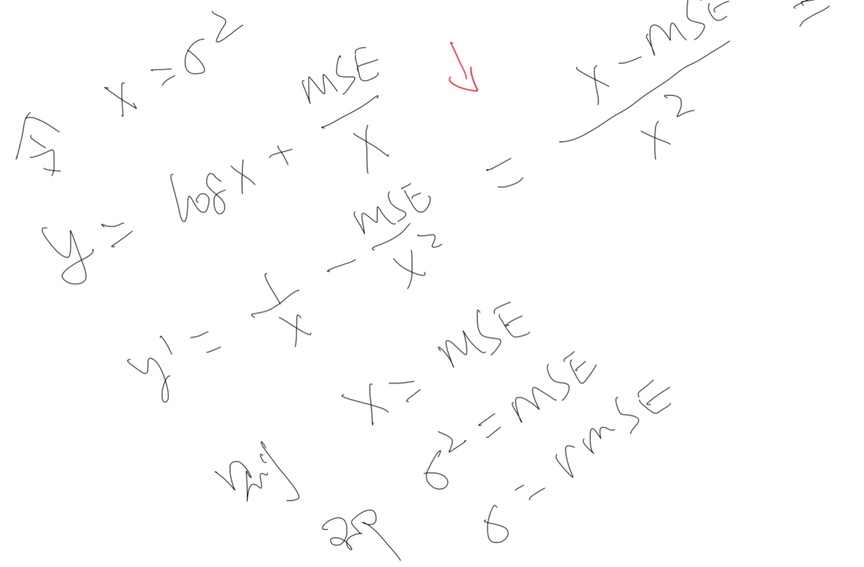
We define L_ideal as the case where the submission perfectly matches the ground truth values, with an uncertainty of 10 parts per million (ppm). This ideal case is defined based on Ariel's Stability Requirement. For L_ref is defined using the mean and variance of the training dataset as its prediction for all instances.

The score will return a float in the interval [0, 1], with higher scores corresponding to better performing models. Any score below 0 will be treated as 0.

The full metric implementation is available here.







μT σT

σ 优化策略

1° 全局一个 σ

2° 每个 w 一个 σ，共 23 个 σ

3° 每个样本一个 σ，共 m 个 σ → 如何做

4° 每个样本、每个 w 都有对应的 σ