

H2O-3 Open Source **Custom Loss Function** in Gradient Boosting Machines

Veronika Maurerová
Software Engineer, H2O.ai
veronika.maurerova@h2o.ai

ABOUT ME

H₂O.ai



Veronika is a techie who is interested in **Machine Learning, Mathematics and Statistic, Open Source, mentoring, and knowledge sharing.**

She holds a Master degree in Computer Science from Czech Technical University in Prague with specializations in **Knowledge Engineering.**

She worked as Software Engineer in **Ataccama** and **SEQENGI** before.

Veronika is currently a Software Engineer on the **H2O-3 Open Source ML library** at [H2O.ai](https://h2o.ai).

She is also a mentor in **Czechitas** project and manages cooperation H2O.ai with **Faculty of Information Technology** at CTU.

AGENDA

- **What is new in H2O-3 Open Source ML library**
- **Gradient Boosting Quick Intro**
- **Loss Function meaning**
- **Custom Loss Function**
 - Usage and definition
 - Example in H2O-3

What is new in H2O-3

Open Source ML library

H2O Release 3.26 (Yau)

LATEST STABLE

- Support-vector Machines
- 2D Partial Dependency plot
- Import MOJO with Metrics
- XGBoost Improvements
- TreeSHAP for Distributed Random Forest
- Target Encoding MOJO support
- **Custom Loss Function in Gradient Boosting Machines**

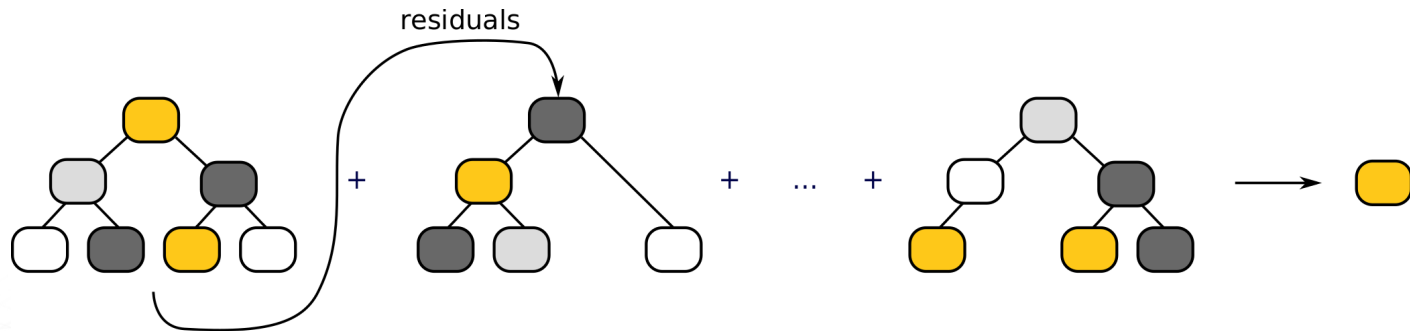
Gradient Boosting Machines

Quick intro

Gradient Boosting Machines

QUICK INTRO

- Ensemble of either **regression** or **classification** tree models.
- Weak prediction algorithms are **sequentially** applied to the incrementally changed data to create a **series of decision trees**, producing an **ensemble of weak prediction models**.



- More information in H2O GBM [Documentation](#) or [Booklet](#)

Loss function

Meaning, optimization

Distribution & Loss Function

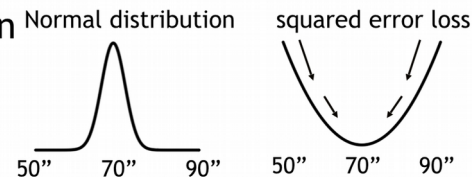
MEANING, OPTIMIZATION

- Distributions in H2O-3:

- Bernoulli
- Quasibinomial
- Multinomial
- Gaussian
- Poisson
- Gamma
- Laplace
- Quantile
- Huber
- Tweedie

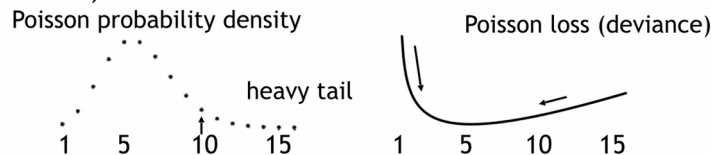
Gaussian “Normal” Distribution

- Height of Males in US



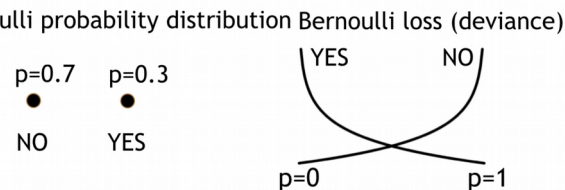
Poisson Distribution

- Probability of 10 red lights on your way to work (if the mean is 5)



Bernoulli Distribution

- Binary outcome (YES/NO)



Custom Loss Function

Usage and definition

Custom Loss Function

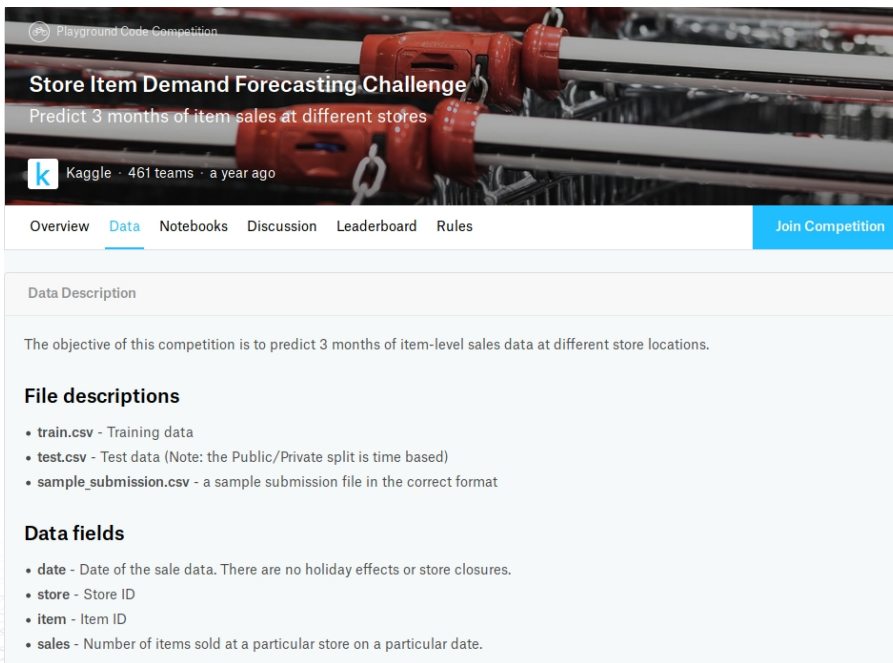
USAGE AND DEFINITION

- Sometimes an asymmetric or specific function is useful.
- Examples:
 - Flight departure prediction – to miss the flight could be a worse error than to come early;
 - Sold items prediction – to have a lack of items in the store could be a worse error than to have some items extra;
 - Disease prediction – to die could be a worse error than to get some extra medicaments.

Custom Loss Function

EXAMPLE: DATA

- [Store Item Demand Forecasting training dataset](#) (source: Kaggle)



The screenshot shows the Kaggle competition page for the 'Store Item Demand Forecasting Challenge'. The page header includes the competition title, a brief description ('Predict 3 months of item sales at different stores'), and the Kaggle logo with the text 'Kaggle · 461 teams · a year ago'. Below the header is a navigation bar with tabs for 'Overview', 'Data', 'Notebooks', 'Discussion', 'Leaderboard', and 'Rules', and a 'Join Competition' button. The 'Data' tab is selected. The main content area is titled 'Data Description' and contains the following text:

The objective of this competition is to predict 3 months of item-level sales data at different store locations.

File descriptions

- train.csv - Training data
- test.csv - Test data (Note: the Public/Private split is time based)
- sample_submission.csv - a sample submission file in the correct format

Data fields

- date - Date of the sale data. There are no holiday effects or store closures.
- store - Store ID
- item - Item ID
- sales - Number of items sold at a particular store on a particular date.

Custom Loss Function

EXAMPLE: DEFINITION

- Custom Loss Function demo - [iPython notebook](#).
- Goal: penalize 'lack of items in the store' error.
- Use **Gaussian Distribution** for easy demonstration

Implementation of Gaussian Distribution

```
class DistributionGaussian(CustomDistributionGeneric):
    """
        Predefined distribution class for regression problems.
    """
    def link(self):
        return "identity"

    def init(self, w, o, y):
        return [w * (y - o), w]

    def gradient(self, y, f):
        return y - f

    def gamma(self, w, y, z, f):
        return [w * z, w]
```

Implementation of Custom Gaussian Distribution

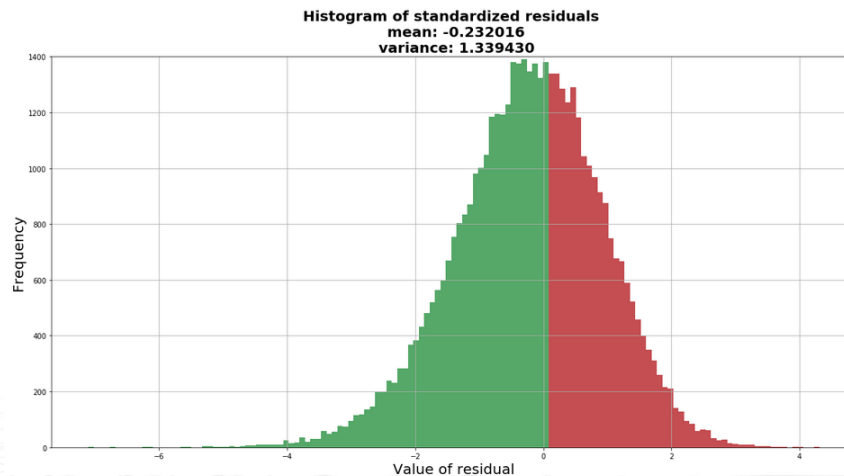
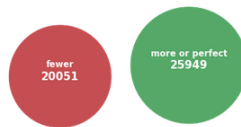
[illegible]

Custom Loss Function

EXAMPLE : RESULTS

Acceptable vs. Unacceptable Predictions

Gaussian Distribution



Custom Loss Function

EXAMPLE: DEFINITION

- Custom Loss Function demo - [iPython notebook](#).
- Goal: penalize 'lack of items in the store' error.
- Use **Gaussian Distribution** for easy demonstration

Implementation of Gaussian Distribution

```
class DistributionGaussian(CustomDistributionGeneric):
    """
        Predefined distribution class for regression problems.
    """
    def link(self):
        return "identity"

    def init(self, w, o, y):
        return [w * (y - o), w]

    def gradient(self, y, f):
        return y - f

    def gamma(self, w, y, z, f):
        return [w * z, w]
```

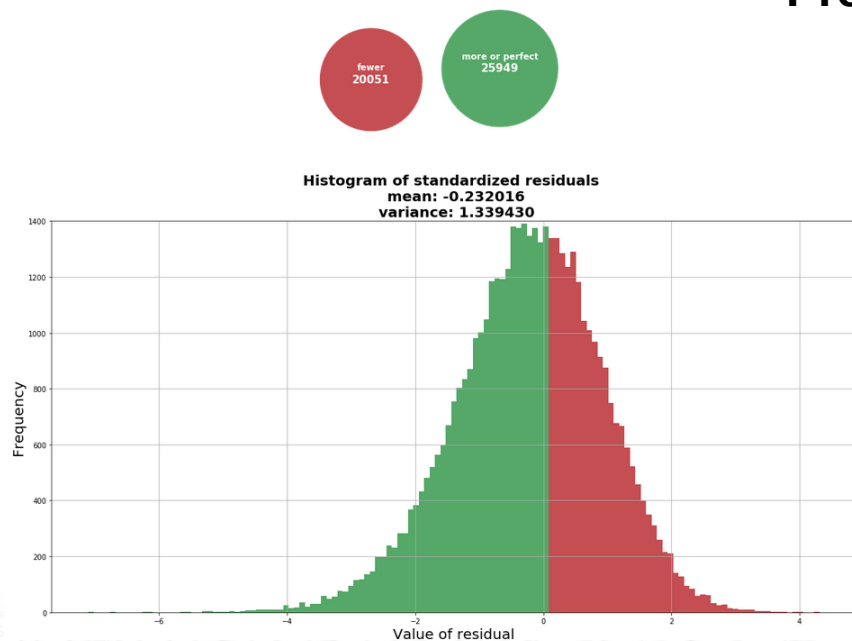
Implementation of Custom Gaussian Distribution

[illegible]

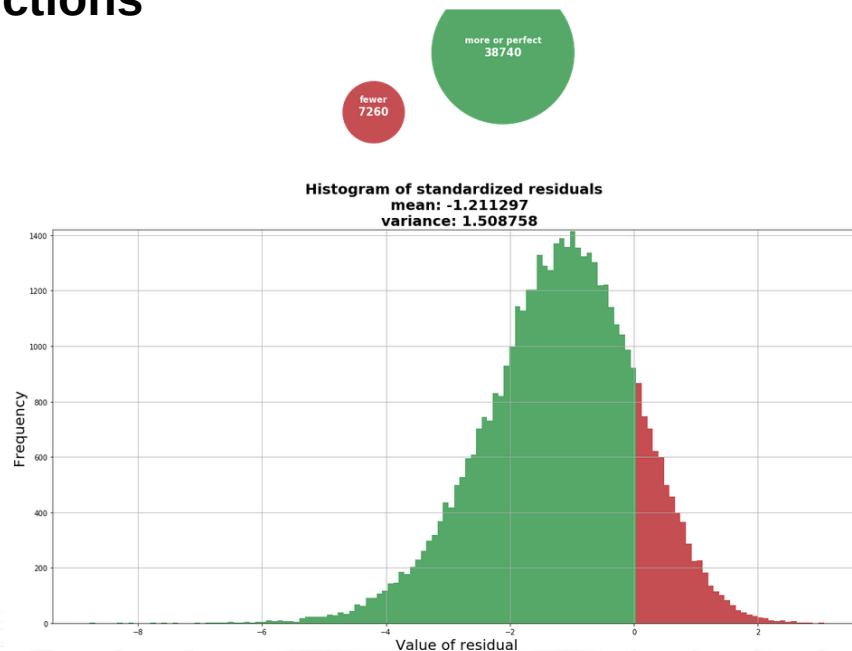
Custom Loss Function

EXAMPLE : RESULTS

Acceptable vs. Unacceptable Predictions

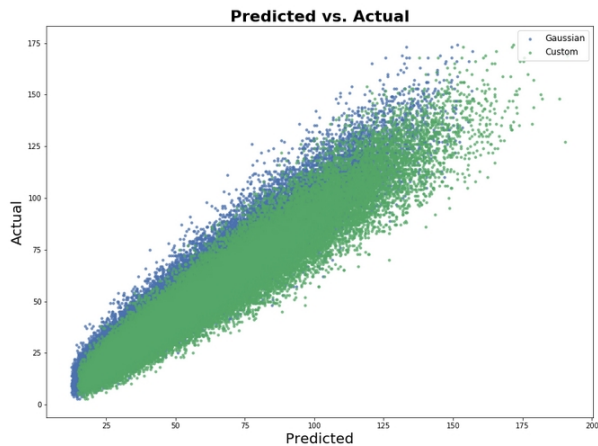
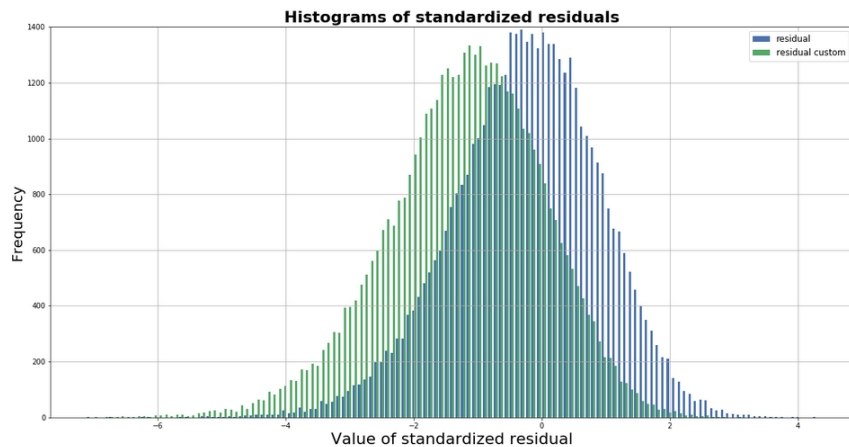


Gaussian Distribution



Custom Gaussian Distribution

Gaussian (blue) vs. Custom Gaussian (green)



Custom Loss Function

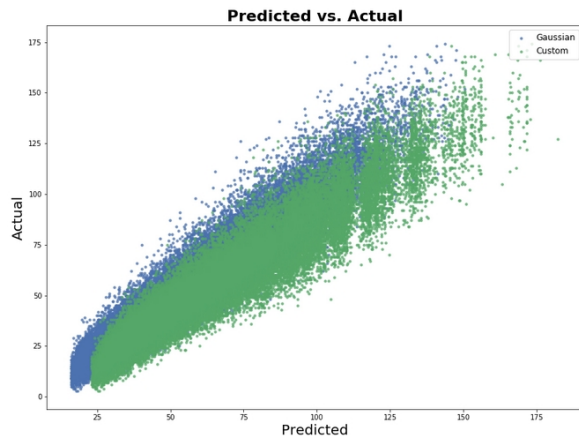
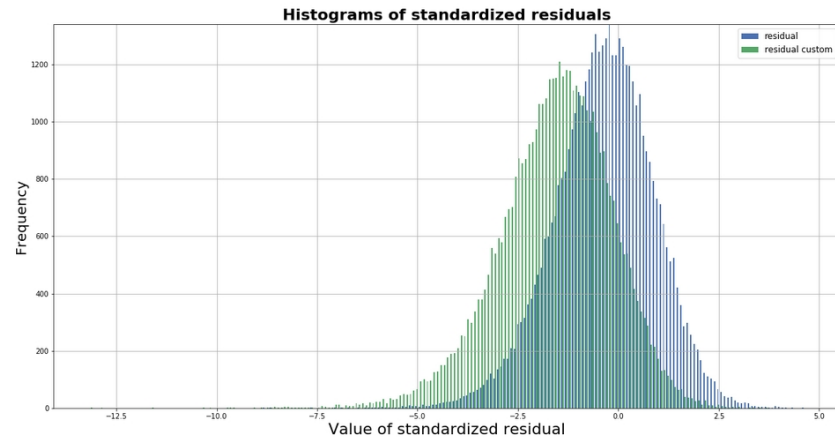
CUSTOM DISTRIBUTION VS. CUSTOM METRIC

- **Customize loss function and stopping metric together**

Implementation of Custom MSE metric

[illegible]

Gaussian (blue) vs. Custom Gaussian (green) with Custom Metric



Conclusion

Takeaways

TAKEAWAYS

- Customization of Loss Function available in H2O-3 from **release 3.26 (Yau)**.
- Custom Loss Function could be useful when **asymmetry** in prediction is needed.
- **Combination** of Custom Distribution and Custom Metric allows more flexibility in customization of training and validation.

The logo for H2O.ai, featuring the text "H2O.ai" in a bold, black, sans-serif font. The "2" is a subscript. The logo is contained within a solid yellow square. The background of the slide is dark gray with a complex pattern of thin, white, curved lines that create a sense of depth and movement, resembling a wireframe or a stylized landscape.

H₂O.ai

Thank you!