

Influence of hyperparameters on aggregating predictions of infinite number of experts

Kunin-Bogoiavlenskii Sergey

Expert: R. D. Zukhba

Consultant: A. V. Zukhba

Moscow Institute of Physics and Technology

kunin-bogoiavlenskii.sm@phystech.edu

April 2, 2024

Contents

- 1 Introduction
 - What can be forecast?
 - Targets
- 2 Problem statement
 - Data
 - Terms
 - Algorithm
- 3 Experiments

What can be forecast?

Tell us what the future holds, so
we may know that you are gods.

Isaiah 41:23

- Weather conditions
- Economic trends
- Technology advancements
- Consumer behavior
- Population growth
- Political elections outcomes

Targets

Prediction is very difficult,
especially if it's about the future.

Niels Bohr

- ① Time series generator implementation
- ② Aggregating algorithm implementation
- ③ Experiments with various hyperparameters

Problem statement

There are two kinds of forecasters: those who don't know, and those who don't know they don't know.

John Kenneth Galbraith

Data

It is assumed that there are multiple generators, whose structure is unknown to the predictors. These generators switch, producing a time series that is subdivided into a sequence of segments - areas of stationarity, which can be studied using machine learning methods.

Gerators implemented:

- Linear
- ARMA

Problem statement

Terms

- X — signals space
- Y — responses space
- \mathcal{N} — set of experts, indexed by natural numbers
- D — decision space, to which predictions belong
- $\lambda : D \times Y \rightarrow \mathbb{R}_+$ — nonnegative loss function
- $L_T^i = \sum_{t=1}^T l_t^i$ — cumulative loss of expert i during the first T steps
- $H_T = \sum_{t=1}^T h_t$ — master's cumulative loss during the first T steps
- $R_T^i = H_T - L_T^i$ — master's regret relative to the expert i

Problem statement

Algorithm

FOR $t = 1, 2, \dots$:

1. Expert f^t initialization
2. Experts' predictions $f_t^i = f_t^i(x_t)$, $1 \leq i \leq t$
3. Master's prediction evaluation $\gamma_t = \text{Subst}(\mathbf{f}_t, \hat{\mathbf{w}}_t)$
4. Computation of master's loss $h_t = \lambda(p_t, y_t)$ and experts' losses l_t^i
5. **Loss Update** weights modification
6. **Mixing Update** weights modification

ENDFOR

Experiments

Initialization weights

Default weights: $w_1^i = \frac{1}{(i+1) \ln^2(i+1)}$

Experimental: $\frac{1}{i^\alpha}$, $\frac{1}{(i+1) \ln(i+1) \ln^2 \ln(i+1)}$, etc.

Window size

As the algorithm does not know the locations of generator switches, finding an optimal training window is also a challenge.

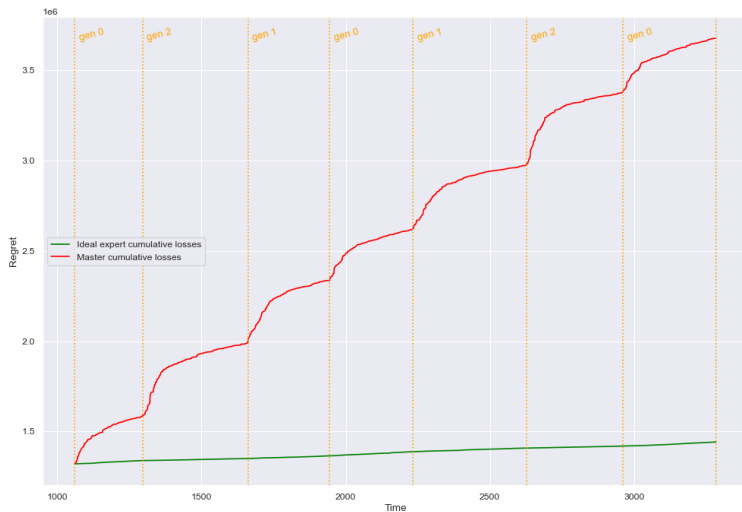
Mixing update coefficients

Default coefficient: $\alpha_t = \frac{1}{t+1}$

Experimental: $\frac{1}{(t+1)^\beta}$, etc.

Metric - the regret

Losses plot



The End