Hedging bounds

March 8, 2018

Contents

1 Bounds **1**
1.1 Regular hedging bound 2
1.2 Hedging using only the current data 2
1.3 Hedging by live estimate update 2
1.4 **TODO** Hedging by preempting the estimate 3

1 Bounds

Suppose the true time series is $s(t)$. Due to data revisions, we first get a rough estimate $\hat{s}(t)$ and then a set of patches $\delta_j(t)$ at j time points after t where j goes from 1 to some k. Its mostly safe to assume that these patches are normally distributed with $|\mu_j| > |\mu_{j-1}|$ (though this information is not used anywhere as of now).

Because of these updates in s, we get certain uncertainty in estimating the model loss $m_i(t)$. Let $\hat{m}_i(t)$ be the first estimate of loss, then $m_i(t)$ (the real loss) is bounded as:

$$\hat{m}_i(t) - \sum_{j=1}^{k} \theta_i(j) \leq m_i(t) \leq \hat{m}_i(t) + \sum_{j=1}^{k} \theta_i(j)$$

Here $\theta_i(j)$ captures the effect of δ_j for the i^{th} model.

An online ensemble strategy here describes the weight update mechanism based on losses that each of the models and the ensemble as a whole receives. What we want is to bound the ensemble loss $L_H = \sum_{t=1}^{T} \sum_{i=1}^{n} p_i(t) m_i(t)$ in terms of the loss of any single expert $L_i = \sum_{t=1}^{T} m_i(t)$. Here $p_i(t)$ is the normalized weight $w_i(t)$ for expert i at time t.

1.1 Regular hedging bound

If we do regular hedging and have the real time series in hand, then the bound is similar to [1] and is given by:

$$L_H \leq \frac{-\ln w_i(1) - L_i \ln \beta}{1-\beta}$$

β is a hyper parameter in $[0, 1]$.

1.2 Hedging using only the current data

Since we don't have real data, we can only use the estimates. The weight update equation is $w_i(t+1) = w_i(t)\beta^{\hat{m}_i(t)}$. If we only use the latest estimate without utilizing the patches we get for earlier time points, we get the following bound which adds the extra uncertainty term:

$$L_H \leq \frac{-\ln w_i(1) - L_i \ln \beta}{1-\beta} - \frac{\ln \beta}{1-\beta} T(\sum_{j=1}^{k} \theta_i(j) + \max_{i' \in [1\ldots n]} \sum_{j=1}^{k} \theta_{i'}(j))$$

The bad thing here is that the extra term is dependent on T which makes it poorer as time increases.

1.3 Hedging by live estimate update

Here, we use all data available to us at any moment. This is equivalent to recalculating the weights from the start (using $w_i(1)$ values) at every time step. The weight updates here follow the following inequalities:

$$w_i(T+1) \geq w_i(1)\beta^{\sum_{t=1}^{T-k} m_i(t)} \beta^{\sum_{t=T-k+1}^{T} \hat{m}_i(t)} \beta^{\sum_{j=1}^{k-1} \theta_i(j)(k-j)}$$
$$w_i(T+1) \leq w_i(T)\beta^{m_i(T)} \beta^{-\sum_{j=1}^{k} \theta_i(j)j/T}$$

This time we get the following bound:

$$L_H \leq \frac{-\ln w_i(1) - L_i \ln \beta}{1-\beta} - \frac{\ln \beta}{1-\beta} \left(k\theta_i(k) + \left(\sum_{j=1}^{k-1} \theta_i(j)(2k-j) \right) + \max_{i' \in [1\ldots n]} \left(\sum_{j=1}^{k} \theta_{i'}(j)j \right) \right)$$

This doesn't involve T and thus is asymptotically better. Another thing to note here is that the term involving max says that we can do better by removing models with high θ values.

There are a few things to note regarding the component models:

- To reduce the loss L_H, one can put a good model in the mix (with low L_i) but since the models actually experience real time loss of $\hat{L}_i$, they will mostly have some θ values to trade off.

- A very accurate oracle model will have almost zero L_i but to do that, it will have to lower its $\hat{L}_i$ and thus will have higher θ values.

- A flatter model (like a uniform probability one) will have $\theta_i(j) = 0$ but will have high L_i.

- The model which is bad with lags (resulting in the max term) and is not the best one considering its L_i can be removed without any theoretical loss in performance.

1.4 TODO Hedging by preempting the estimate

References

[1] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.