# Equity in Learning Problems: An OWA Approach

Juliette Ortholand, Sébastien Destercke(✉), and Khaled Belahcene

Université de Technologie de Compiègne, Heudiasyc, Compiègne, France
{juliette.ortholand,sebastien.destercke,khaled.belahcene}@hds.utc.fr

**Abstract.** It is well-known in computational social choice that the weighted average does not guarantee any equity or fairness in the share of goods. In a supervised learning problem, this translates into the fact that the empirical risk will lead to models that are good in average, but may have terrible performances for under-represented populations. Such a behaviour is quite damaging in some problems, such as the ones involving imbalanced data sets, in the inputs or the outputs (default prediction, ethical issues, ...). On the other hand, the OWA operator is known in computational social choice to be able to correct this unfairness. This paper proposes a means to transpose this feature to the supervised learning setting.

## 1 Introduction

The typical way to learn a predictive model from data is to search for the model that minimizes the average loss of the predictions made by this model on a set of training data. However, minimizing the average loss may well lead to poor results on some under-represented populations.

This is a well known fact, that happens in several settings that have proposed different solutions to the issue: in class imbalanced data sets, concerning for instance rare diseases or default (of payment, of production), the classical solution is to modify the sample sizes, for instance by over-sampling instances of the under-represented class [9]; in fairness issues [10], where the goal can be to protect sensitive populations or minorities, often by modifying not the sample but the loss function adequately; in extreme statistics [11], where one must guarantee that rare instances will be well predicted, for instance by learning a model specifically dedicated to them.

In this paper, we look at another aspects of misrepresentation of some data in the learning problem. Namely, we want to ensure that the loss incurred for data poorly represented in the feature space (whatever their class is) is not high. This is yet a different kind of under-representation of some population, whose closest related problem is the previously mentioned one of extreme statistics [11]. Our goal here is to propose a method ensuring that under-represented data will not suffer from a too high loss, while preserving a good average accuracy. To perform such a task, we will modify the classical expected loss by using the notion of ordered weighted averaging, an often used notion in fairness problems within computational social choice [13].


J. Davis and K. Tabia (Eds.): SUM 2020, LNAI 12322, pp. 187–199, 2020.
https://doi.org/10.1007/978-3-030-58449-8_13

More precisely, we will propose to give more weight to unknown zones. The paper is organised as follows: the formal mathematical framework can be found in Sect. 2, where we provide reminders, notations and preliminaries, and in Sect. 3, where we describe our proposal. This is followed by the experiments in Sect. 4 then by some related works in the Sect. 5. The paper ends with some conclusion and discussion on our work in Sect. 6.

## 2 Preliminaries

We consider a standard supervised problem where we have observations $(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}$, $i = 1, \ldots, n$ where $x_i$ are some inputs, and $y_i$ the observed outputs.

### 2.1 Supervised Classification via Empirical Risk Minimization

The general goal of supervised machine learning is to estimate a predictive model $h^* : \mathcal{X} \to \mathcal{Y}$, issued from a space $\mathcal{H}$ of hypothesis, such that the model delivers good prediction in average. This principle is most often translated by choosing the model minimizing the empirical risk, i.e.,

$$h^* = \arg\min_{h \in \mathcal{H}} R_{emp}(h)$$

where

$$R_{emp}(h) = \sum_{i=1}^{n} \ell(h(x_i), y_i) \tag{1}$$

with $\ell(h(x_i), y_i)$ the loss of predicting $h(x)$ when $y$ is the observed value. This empirical loss serves as an estimate of the true loss, i.e., $R(h) = \int_{\mathcal{X}\times\mathcal{Y}} \ell(h(x), y) dp(x, y)$, that is inaccessible as we do not know $p(x, y)$. Also, in many cases, $\mathcal{H}$ is a parametric family with parameters $\theta \in \Theta$, and in this case we will denote by $h_\theta$ the predictive function having $\theta$ for parameter.

### 2.2 Some Shortcomings Due to the Averaging of the Risk

Guaranteeing a low average loss does not prevent from having large losses for poorly represented groups of values, and even in some cases promote such large discrepancies [10].

*Example 1.* Figure 1 displays two different classes (i.e., $y \in \{0, 1\}$) represented in red and blue, that suffer from the problem we consider in this paper. Indeed, the two classes are balanced (there are about as much red as blue), but some region of the input space are less represented than others. More precisely, the data corresponding to each class have been generated by the following distributions:

$$X|y = 0 \sim \pi_0^1 \, \mathcal{N}(\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}3 & 0\\0 & 0.5\end{bmatrix}) + \pi_0^2 \; \mathcal{N}(\begin{bmatrix}16\\6\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 0.5\end{bmatrix})$$

$$X|y=1 \sim \pi_0^1 \mathcal{N}(\begin{bmatrix}0\\2\end{bmatrix}, \begin{bmatrix}3 & 0\\0 & 0.5\end{bmatrix}) + \pi_0^2 \ \mathcal{N}(\begin{bmatrix}15\\7\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 0.5\end{bmatrix})$$

with $\pi_0^1 = \pi_1^1 = 0.95$ and $\pi_0^2 = \pi_1^2 = 0.05$, meaning that the upper-right region is much less represented than the lower-left in Fig. 1. The frontier in this region corresponds to the one obtained by a logistic regression trained according to Eq. 1. It is easy to see that the model does a very bad job at predicting the data in the upper-right corner, as could be expected.

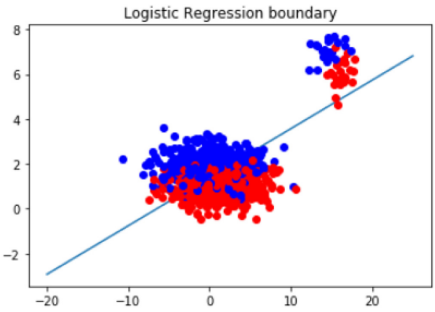



**Fig. 1.** Logistic Regression boundary (Color figure online)

### 2.3 A Short Introduction to Ordered Weighted Averages (OWA)

The OWA operators, initially introduced by Yager [14], apply a weighted aggregation functions on ordered values. Usually, the OWA is applied to a vector $\mathbf{a} = (a_1, \ldots, a_n)$ of real-valued quantities $a_i \in \mathbb{R}$, and is defined by a set $(w_1, \ldots, w_n)$ of positive weights $(w_i \geq 0)$ summing up to one $(\sum_i w_i = 1)$. Formally speaking, the OWA consists in first permuting the values $a_i$ in ascending order, i.e., such that for $i < j$, we have $a_{\sigma(i)} \leq a_{\sigma(j)}$, with $\sigma$ denoting the permutation. The classical OWA operator is then

$$OWA(a_1, \ldots, a_n) = \sum_{i=1}^{n} w_i a_{\sigma(i)}. \tag{2}$$

The OWA operator therefore allows to put different strength on lower or higher values, irrespectively of where they appear in the vector $\mathbf{a}$. We retrieve the arithmetic mean for $w_i = 1/n$, and $k$th percentiles when we have $w_{k/n} = 1$ for some $i$. In particular, the minimum and maximum values are respectively retrieved for $w_1 = 1$ and $w_n = 1$. They characterize what Yager called extreme behaviour of "andness" and "orness", respectively.

## 3 Our Proposal

In this paper, we consider the use of OWA operators [8], that propose to make weighted averages not on the initial observations $(x_i, y_i)$ and their associated

losses, but on a re-ordering of them, with the idea that a higher weight should be given to poorly represented instances, to be ranked first in our proposal. We will denote by $(x_{\sigma(i)}, y_{\sigma(i)})$ the corresponding permutation on observations. More precisely, we propose not to optimise $R_{emp}(h)$, but rather $R_{OWA}(h)$, where

$$R_{OWA}(h) = \sum_{i=1}^{n} w_i \ell(h(x_{\sigma(i)}), y_{\sigma(i)}) \tag{3}$$

with the idea that when $i \leq j$, the instances $x_{\sigma(i)}$ is not as well represented as $x_{\sigma(j)}$ in the data set. It should be noted that in contrast to usual OWA and Eq. (3), we will not considering the re-ordering of values $\ell(h(x_{\sigma(i)}), y_{\sigma(i)})$, but a re-ordering based on the representativeness of the instances $x_i$.

### 3.1 Ranking Instances by Representativeness

A first task to apply Eq. (3) to our setting is to order the instances by their representativeness in the data set. To do this, we can order them by measuring, for instance, the epistemic uncertainty, or lack of knowledge concerning each instance $x_i$, e.g., following ideas from [12] to obtain a score $E_i$ for each instance $x_i$, and then ordering them accordingly in Eq. (3), i.e., $\sigma(i) \leq \sigma(j)$ if $E_i \geq E_j$.

One simple idea that we will apply in the following is to measure the density of points around a neighbourhood of fixed size around $x_i$ to compute its associated epistemic uncertainty. For this, an easy technique one can use is to simply perform a kernel density estimation through the formula

$$f(x) = \frac{1}{n} \sum_{i=1}^{n} K_\epsilon(x - x_i)$$

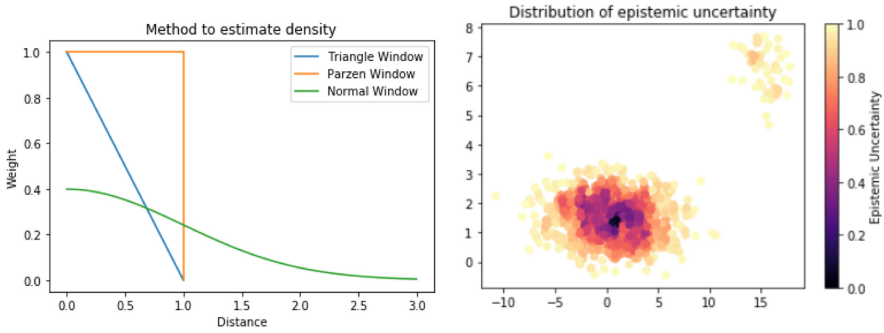
with $K_\epsilon$ a kernel function. Common choices of kernel functions include:

- the Parzen window, defined as $K_\epsilon(x - x_i) = \frac{1}{2\epsilon} 1_{|x-x_i|<\epsilon}$, that simply comes down to count the number of points that are at a distance below a certain $\epsilon$;
- the triangle kernel, defined as $K_\epsilon(x - x_i) = (\frac{1}{\epsilon} - |x - x_i|)1_{|x-x_i|<\epsilon}$, for which weights decrease linearly from one to zero depending on the distance;
- the normal window, defined as $K_\epsilon(x - x_i) = \frac{1}{\epsilon\sqrt{2\pi}} e^{\frac{-(x-x_i)^2}{2\epsilon^2}}$, for which the weights for the points depend on a normal distribution around the chosen point with a mean of zero and a standard deviation of $\epsilon$.

The three kernels are pictured in Fig. 2a.

Once a kernel is chosen, we can then simply use $f(x_i) = E_i$ as a score quantifying epistemic uncertainty. Note that in our case, it is not important to have a reliable estimate of the density (a very difficult problem), but to just have a reliable ordering between the different points, as $f(x_i)$ will only be used to order values in OWA operators.

Figure 2b represents the distribution of the epistemic uncertainty of points given in Fig. 1, computed for a triangular kernel with $\epsilon$ being the mean distance between points. One can readily see that the most uncertain points (hence the first in the re-ordering) are those in the upper right corner, that is precisely those for which we would like to increase accuracy, followed by the ones on the border of the big cluster.




(a) Methods to estimate density

(b) Distribution of the epistemic uncertainty

**Fig. 2.** Epistemic uncertainty

### 3.2 OWA Weights to Induce Equity

The next step is how we can choose the weights in Eq. (3) so as to balance the accuracy in the model between well-represented and poorly represented instances. Clearly, if we pick $w_i = 1/n$, this re-ordering is a useless step and leads us to the usual solution. However, we can easily pick weights that will enforce giving more importance to poorly represented instances. More precisely, we can pick a function $\phi : [0, 1] \to [0, 1]$ and take as weights

$$w_i = \phi(i/n) - \phi(i-1/n)$$

if $\phi(x) = x$, then we retrieve the weighted average. If $\phi$ is concave, then we start giving more weights to first ordered instances, and less weight to last ordered instances. In terms of OWA, we increase the "andness" of the function, that we can then parameterize to be more or less fair. Ideally, this number of parameters should not be too high, and example of such functions include:

- The $L_p$ norm on with $p \in [0, 1]$, which function is $\phi(x) = x^p$. The lower $p$, the more we increase the "andness"
- piece-wise linear functions made of two linear parts, that can be define with two parameters $p$ and $prop$ as follows:

$$\phi(x) = \begin{cases} px & if \quad x \leq prop \\ \frac{1-p\times prop}{1-prop}x + \frac{prop(p-1)}{1-prop} & if \ prop \leq x \leq 1 \end{cases} \tag{4}$$

  where $p$ defines the slope before the abscissa value $prop$.

Both are represented on Fig. 3 with the $L_{1/2}$ norm and the linear by part function with $prop = 0.1$ and $p = 4$.

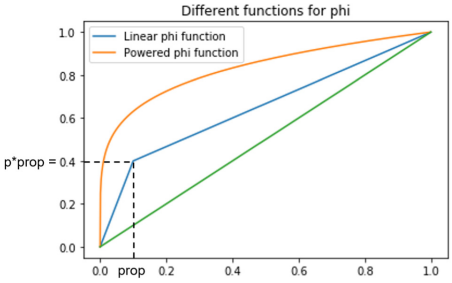



**Fig. 3.** Different functions for $\phi$

*Remark 1 (Model optimisation).* Note that as we do not modify the nature of the loss function $\ell$, most optimisation techniques used for the arithmetic mean will be straightforward to extend. Thanks to the versatility of $\phi$, we can also think of other kind of behaviours than concave ones. For instance, an S-shaped function would amount to try to balance between being quite good on poorly represented as well as quite good of very-well represented groups, thus protecting minorities and majorities.

## 4 Experiments

This section presents some experiments using our approach to try to augment the accuracy on the poorly represented data, that we will call *minority*, while preserving a good average accuracy. After quickly describing the chosen model, we will provide results first on synthetic data sets, second on real-world data sets. In this latter case, since there are no benchmark data sets focusing on the problem we try to solve, we will try to adapt common UCI data sets [6] to our setting.

### 4.1 Implementing the Proposal

We will apply our approach to standard regularized logistic regression in binary classification problems, with the output class $Y \in \{0, 1\}$. Let us simply recall that in this case, we learn a probabilistic model of the shape

$$h_\theta(x) = \frac{1}{1 + \exp -\theta x} \tag{5}$$

with $h_\theta(x) = p(y = 1|x)$. The associated loss is

$$\ell(y, h_\theta(x)) = -(ylog(h_\theta(x)) + (1 - y)log(1 - h_\theta(x)) + \theta^2 \tag{6}$$

that corresponds to a logarithmic loss with a $L_2$ regularisation term. In experiments, we use python sklearn package to fit the different models, with which it is straightforward to add weights to samples.

In the experiments, we used the triangle kernel applied to data with an Euclidean distance computed between them. The reason for this choice is that it gives no ties between values $f(x_i)$ in practice (while Parzen windows delivers the same value when having the same number of data within it), and that it has a finite support, therefore being more coherent than the normal kernel with the fact that epistemic uncertainty is mostly a local property. However, our tests with other kernel functions show no significant differences.

Regarding the parametric shape of the OWA, we picked the shape given by Eq. (4), as in our experiments the use of the $L_p$ norm tended to give too quickly too much importance to poorly represented data, introducing sometimes important discontinuities in our results for small changes of the parameter. This can already be seen in Fig. 3. Thus, the $\phi$ function depends on two parameters $p$, the slope of the first linear part of the function and *prop*, the abscissa of the slope breaking point.

As our aim is to improve accuracy on minorities while keeping a good average accuracy, this means that our performances will be measures according to two values: the average accuracy on the minority samples only, $acc \in [0, 1]$, and the classical average accuracy, $ACC \in [0, 1]$. For this reasons, we will present our experimental results a Pareto front on the space $[0, 1] \times [0, 1]$, as for a given couple $(p, prop)_k$ of the proposal, we will obtain a pair $(acc_k, ACC_k)$. This means that we will present the results for all non-dominated values, that is all $(acc_k, ACC_k)$ such that there will be no other pairs $(p, prop)_{k'}$ with $acc_{k'} \geq acc_k$ and $ACC_{k'} \geq ACC_k$.

### 4.2 Synthetic Data Set

In the first set of experiments, we simply consider the same distributions as the ones described in Example 1 for a binary problem.

As in Example 1, minorities of each class represent about 5% of the total quantity of samples from that class. For each set of experiments, we generate 1000 points for the training set, and as much for the testing set.

In the experiments, we proceeded with a simple grid search to fix $p$ and *prop*. We let $p$ vary between 1 and 5 with a 0.5 increment, and *prop* between $[0.1, 0.3]$ with a step of 0.05. Every test is made on 10 different sets of data and the mean is taken to obtain reliable estimates. The total accuracy and the accuracy of the minority are studied.

Figure 4 illustrates the results obtained for a particular run and give the model obtained for the parameters $p = 5.0$ and $prop = 0.2$. One can easily see that the obtained model is much more relevant on the minorities, as it starts to discriminate the two classes in this region.

One question though is to know whether this potential benefit on the minority region does not alter too much the overall accuracy. The answer is provided by Fig. 5 that displays the obtained Pareto front as well as the results of the basic

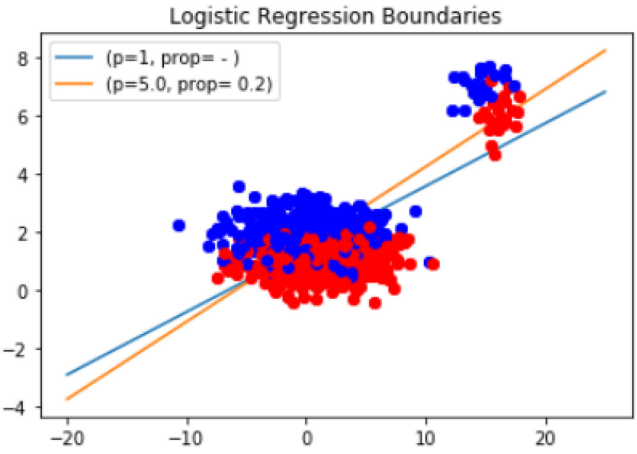



**Fig. 4.** Logistic Regression Boundaries for extreme values

model $(1, -)$. One can observe that the accuracy on the minority region can increase by more than 10%, going from 0.53 to 0.66, while the loss on the average accuracy is below 3%. Note that the Pareto allows a possible decision maker to finely choose the trade-off between minority protection and overall performances.

### 4.3 UCI Data Sets

As we said, there are to our knowledge no benchmark data set that explicitly deals with the problem of within-class imbalance, the situation described by our synthetic example. For this reason, we tried to apply it to UCI data sets susceptible to display similar behaviours.

To test whether this is the case, a simple procedure is adopted: we split the data set into training and test sets, and order the elements of the test set according to their epistemic uncertainty, computed by using the samples of the training set. A logistic regression is then fitted to the data, and we check the difference between the global average accuracy ($ACC$) and the average accuracy of the first $\alpha$% of the ordered test samples ($acc$). If the difference $ACC - acc$ is big enough, we retain the data set.

In our experiment, we fixed the value $\alpha$ to 10%, and similarly to the previous case, proceeded to apply our method by letting $p$ vary between 1 and 5 with a 0.5 increment, and $prop$ between $[0.1, 0.3]$ with a step of 0.05. Each training/testing experiment is made by taking 50% of the data set as training, and the process is repeated a hundred times for each configuration, the mean being kept as a representative point.

Perhaps surprisingly, it proved quite hard in this manner to find suitable data sets. Finally, we retained three binary classification data sets that are summarised in Table 1.

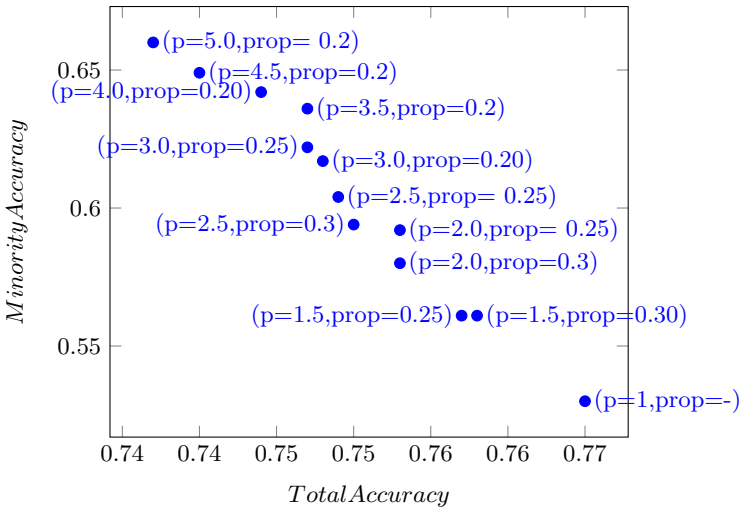



**Fig. 5.** Pareto front on synthetic data set

**Table 1.** Data set descriptions

| Data set | Samples | Percentage of positive class |
|---|---|---|
| Istanbul Stock Exchange [4] | 536 | 50% |
| Credit Approval [5] | 653 | 45% |
| Vertebral Column [5] | 310 | 68% |

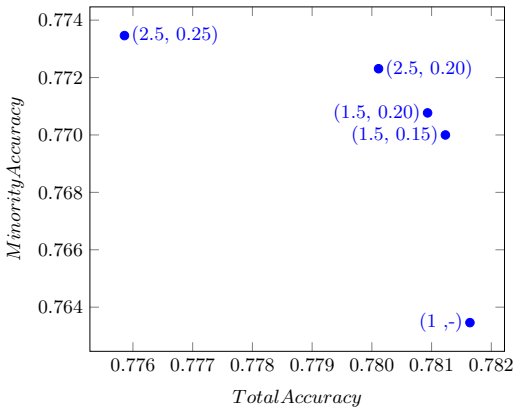



**Fig. 6.** Istanbul Stock Exchange Pareto Front with $(p, prop)$ values

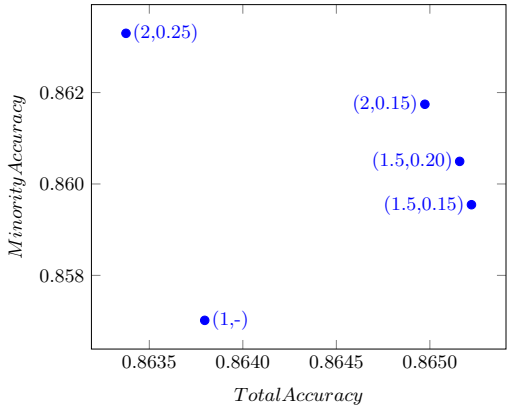



**Fig. 7.** Credit Approval Pareto Front with $(p, prop)$ values

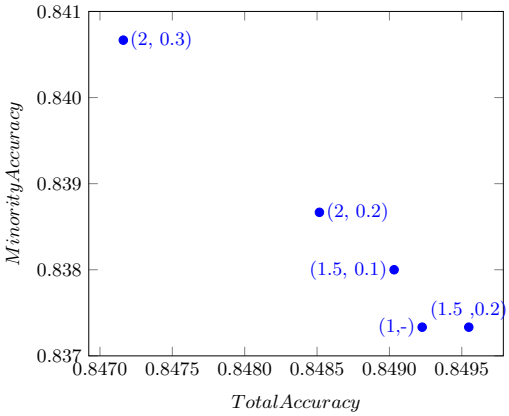



**Fig. 8.** Vertebral Column with $(p, prop)$ values

The resulting Pareto fronts of our experiments, along with the accuracies of the base model corresponding to $(1, -)$, are given in Figs. 6 (Istanbul), 7 (Credit) and 8 (Vertebral). In each of them, we can see a possible increase in the minority accuracy that out-weights the drop in global, average accuracy. However, it is clear that since for these data sets the difference between $acc$ and $ACC$ is already quite low for the basic model (given by the point $(1, -)$ in the Figures), we cannot hope to achieve a gain as significant as the one of the synthetic data sets.

So, while the presented results confirm that the proposed approach is working, future works should focus on exhibiting similar behaviours in existing data sets, maybe by revisiting the ordering we use, the considered learning algorithm or by focusing on specific data sets such as class imbalanced data sets, hoping that the imbalance in the classes is transferred to the input space.

## 5 Links with Other Learning and Estimation Approaches

To our knowledge, the learning approach presented here is quite original, in the sense that applying OWA to learning problems in order to solve inequities has, to our knowledge, not been done before.

### 5.1 From OWA to Probability Sets

A now well-established trend in the learning literature is the so-called distributionally robust approach [1,7] that consider the problem of finding the minimax model over a possible set $\mathcal{P}$ of probability distributions, mostly defined as a neighbourhood of the empirical distribution of the observations $(x_i, y_i)$. Such approaches have been applied, for instance, to fairness issues in machine learning [10] or to transfer learning problems [2].

Since it is well-known that the OWA operator correspond to apply a Choquet integral with a specific Choquet measure [8], and that such Choquet integral can correspond to lower/upper expectations computed for specific probability sets, it would be interesting to study under which conditions and to which extent our current approach could be interpreted as the solution of a minimax problem for some specific set $\mathcal{P}$.

### 5.2 From OWA Loss to Weighted Likelihood

Another classical way to learn a model, and particularly probabilistic models, is through the maximisation of a likelihood function. In such a case, each parameter value $\theta$ determine a probability distribution of a random variable $X$. First recall that the likelihood of a parameter value $\theta$, given a set of observations $x_i \in \mathcal{X}$ is $\mathcal{L}(\theta|x_i) = \prod_i p_\theta(x_i)$.

Provided we consider the logarithmic loss in Eq. (3), one can easily express our weighting scheme in terms of likelihood. For this, it is sufficient to consider the log-likelihood

$$\mathcal{C}(\theta|x_i) = -ln(\mathcal{L}(\theta|x_i)) = -\sum_i ln(p_\theta(x_i)) \tag{7}$$

where the loss is $ln(p_\theta(x_i))$. We can then apply the OWA loss instead of the current loss to get $\mathcal{C}_{OWA}(\theta|x_i) = -\sum_i w_i \ln(p_\theta(x_{\sigma(i)}))$ with $w_i$ the OWA weights. It is then possible to go back to the formula of the likelihood, obtaining

$$\mathcal{L}(X) = \prod_i p_\theta(x_{\sigma(i)})^{w_i}$$

as our new, weighted likelihood. Thus the OWA transformation on the loss which corresponds to multiply the terms by weights, is equivalent to a exponent operation with weight for the likelihood. While such an exponent weight may appear odd at first, it should be noticed that it has been proposed and used before like in [3], where it has been used to down weight anomalous point in Bayesian prediction.

## 6 Conclusion

In this paper, we have presented an original approach, based on OWA operators, to obtain more balanced and equitable classifiers in those problems where data can be scarce in some regions of the input space. Such an approach aims at ensuring that even poorly represented instances will be treated fairly, in the sense that we will not allow them to suffer huge losses, while keeping an average loss comparable to the one obtained without including such equity requirement.

Our illustrative experiments on synthetic data sets indeed show that the method is appropriate, and allows one to obtain a more balanced model. We have also shown that the same observation can be made on UCI data sets, albeit the improvement is here much more general, due to the fact that there is no benchmark data sets explicitly suffering from the problem we have considered in this paper.

We nevertheless believe that the idea of using aggregation operators issued from the social choice literature to solve inequities and unfairness in supervised learning procedure is a promising idea, that needs to be developed. This study is simply a first proposal going in this direction, and many aspects remain to be studied, such as the nature of the ordering between instances or whether there are algorithms where our approach can make a bigger difference, notably in the case of multi-class problems, as we only used logistic regression on binary problems here.

## References

1. Abadeh, S.S., Esfahani, P.M.M., Kuhn, D.: Distributionally robust logistic regression. In: Advances in Neural Information Processing Systems, pp. 1576–1584 (2015)
2. Abadeh, S.S., Nguyen, V.A., Kuhn, D., Esfahani, P.M.M.: Wasserstein distributionally robust Kalman filtering. In: Advances in Neural Information Processing Systems, pp. 8474–8483 (2018)
3. Agostinelli, C., Greco, L.: Weighted likelihood in Bayesian inference. In: Proceedings of the 46th Scientific Meeting of the Italian Statistical Society, pp. 746–757 (2012)
4. Akbilgic, O., Bozdogan, H., Balaban, M.E.: A novel hybrid RBF neural networks model as a forecaster. Stat. Comput. **24**(3), 365–375 (2014)
5. Dua, D., Graff, C.: UCI Machine Learning Repository (2017)
6. Frank, A., Asuncion, A.: UCI Machine Learning Repository (2010)
7. Gabrel, V., Murat, C., Thiele, A.: Recent advances in robust optimization: an overview. Eur. J. Oper. Res. **235**(3), 471–483 (2014)
8. Grabisch, M.: OWA operators and nonadditive integrals. In: Yager, R.R., Kacprzyk, J., Beliakov, G. (eds.) Recent Developments in the Ordered Weighted Averaging Operators: Theory and Practice. Studies in Fuzziness and Soft Computing, vol. 265. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-17910-5_1
9. Han, H., Wang, W.-Y., Mao, B.-H.: Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) ICIC 2005. LNCS, vol. 3644, pp. 878–887. Springer, Heidelberg (2005). https://doi.org/10.1007/11538059_91

10. Hashimoto, T., Srivastava, M., Namkoong, H., Liang, P.: Fairness without demographics in repeated loss minimization. In: International Conference on Machine Learning, pp. 1929–1938 (2018)
11. Jalalzai, H., Clémençon, S., Sabourin, A.: On binary classification in extreme regions. In: Advances in Neural Information Processing Systems, pp. 3092–3100 (2018)
12. Senge, R., et al.: Reliable classification: learning classifiers that distinguish aleatoric and epistemic uncertainty. Inf. Sci. **255**, 16–29 (2014)
13. Shams, P., Beynier, A., Bouveret, S., Maudet, N.: Minimizing and balancing envy among agents using ordered weighted average (2019)
14. Yager, R.R.: On ordered weighted averaging aggregation operators in multicriteria decisionmaking. IEEE Trans. Syst. Man Cybern. **18**(1), 183–190 (1988)