Synergy Conformal Prediction for Regression

Niharika Gauraha[a], Ola Spjuth[a]

[a]*Department of Pharmaceutical Biosciences and Science for Life Laboratory Uppsala University, Uppsala, Sweden*

Abstract

Large and distributed data sets pose many challenges for machine learning, including requirements on computational resources and training time. One approach is to train multiple models in parallel on subsets of data and aggregate the resulting predictions. Large data sets can then be partitioned into smaller chunks, and for distributed data the need for pooling can be avoided. Combining results from conformal predictors using synergy rules has been shown to have advantageous properties for classification problems. In this paper we extend the methodology to regression problems, and we show that it produces valid and efficient predictors compared to inductive conformal predictors and cross-conformal predictors for 10 different data sets from the UCI machine learning repository using three different machine learning methods. The approach offers a straightforward and compelling alternative to pooling data, such as when working in distributed environments.

Keywords:
Conformal Prediction, Machine Learning, Regression, Synergy, Ensemble Methods

1. Introduction

Data has become one of the key assets of many organizations, and the amount of data continues to increase in virtually all domains; sometimes referred to as Big Data [1, 2, 3]. Training statistical (machine) learning models on large data sets can be challenging from different perspectives [4]. For example, large data sets require substantial computational hardware, and

Email addresses: niharika.gauraha@farmbio.uu.se (Niharika Gauraha), ola.spjuth@farmbio.uu.se (Ola Spjuth)

in some cases large data can be difficult to move into a single computer [5]. A multitude of approaches have been developed for machine learning on large data sets [5, 6]. The most popular methodology is probably to train a global model in distributed environments that preserves data locality, using frameworks such as Apache Spark [7]. Another methodology is to distribute (partition) data into smaller units and train multiple models in parallel and aggregate predictions [8].

Conformal Prediction is a relatively recent methodology where conformal predictors are built on top of standard machine learning algorithms and complement the predictions with valid measures of confidence [9]. The two main approaches are Transductive Conformal Prediction (TCP) [10] and Inductive Conformal Prediction (ICP) [11] and they can be used for both classification and regression problems. The main drawback of using TCP is that it is computationally demanding; for every test example a re-training of the model is required. ICP was developed to overcome this issue; it has little computational overhead to the underlying algorithm but there is some loss in the terms of informational efficiency due to a subset of training examples are set aside for calibration. To address this problem of information efficiency, ensembles of conformal predictors were introduced such as Cross Conformal Prediction (CCP) [12] and Aggregated Conformal Prediction (ACP) [13]. The validity of TCP and ICP is proven in that they produce $1 - \epsilon$ expectation tolerance regions, where ϵ is the selected significance level [9]. However, the validity of ensembles of CPs has not been theoretically proven and has been discussed in [13, 14].

Synergy Conformal Prediction (SCP) was recently proposed to address the validity problem of ensembles of conformal predictors that combines monotonic conformity scores instead of p-values [15]. Its applicability has been shown in two scenarios; where data is partitioned in order to reduce the total model training time, and where an ensemble of different machine learning methods is used to improve the overall efficiency of predictions.

In this paper, we extend the SCP methodology for regression problems and explore its usefulness when partitioning large data and aggregating results, and also for working with distributed data without pooling into a single dataset. The paper is organized in the following way. In section 2, we outline the background concepts and notations used throughout the paper. In Section 3 we introduce synergy conformal prediction for regression and discuss its properties. In Section 4 we perform numerical analysis on a set of real data sets. In Section 5 we summarize our results, and in Section 6 we

conclude and discuss implications and future outlook.

2. Background

In this paper we mainly focus on regression problems and assume exchangeability of observations. The object space is denoted by $\mathcal{X} \in \mathbb{R}^p$, where p is the number of features, and label space is denoted by $\mathcal{Y} \in \mathbb{R}$. We assume that each example consists of an object and its label, and its space is given as $\mathcal{Z} := \mathcal{X} \times \mathcal{Y}$. In a classical regression setting, given ℓ data points $Z = \{z_1, ..., z_\ell\}$ where each example $z_i = (x_i, y_i)$ is labeled, we want to predict the label of a new object x_{new}.

In the conformal prediction setting, a non-conformity measure is the score from a function that measures the strangeness of an example in relation to the previous examples [9]. In our experiments, we use the standard regression non-conformity measure

$$\alpha_i = |y_i - \hat{y}_i|, \tag{1}$$

where $\hat{y}_i$ is the estimated output for the object x_i using regression algorithms such as Support Vector Regression (SVR) or Random Forests (RF).

Definition 1 Inductive Conformal Prediction (ICP) for Regression [16]
Given a training set of ℓ examples, $Z = \{z_1, ..., z_\ell\}$, drawn from an exchangeable distribution P, the training data is first divided into a proper training set $\{Z_T\}$ and a calibration set $\{Z_C\}$, where (T, C) is a partition of $\{1, ..., \ell\}$. The regression algorithm (i.e. SVR or RF) is applied to the proper training set, and using the decision rule a strangeness measure is associated with every example in the calibration set. In particular, we use the standard regression non-conformity measure

$$\alpha_i = |y_i - f(x_i)| = |y_i - \hat{y}_i|, i \in C \tag{2}$$

where $f : X \to Y$, is a prediction rule of the model trained on the proper training set Z_T, and $\hat{y}$ is the estimated output. To denote the dependency on the proper training set we write the decision rule as $f(Z_T, .)$. Let us denote by $\alpha_{(1)}, ..., \alpha_{(|C|)}$ the sequence of all α_i corresponding to the calibration set sorted in the ascending order. Let x_{new} (following the same distribution P) be the new object we want to predict, and let $\hat{y}_{new}$ be its estimated label

using the same function $f(Z_T, .)$. The prediction region (PR) for the new object x_{new} is then computed as

$$(\hat{y}_{new} - \alpha_{(s)},\ \hat{y}_{new} + \alpha_{(s)}), \tag{3}$$

where $s = \lfloor \epsilon(|C| + 1) \rfloor$, and $\epsilon \in (0, 1)$ is a chosen significance level, and $(1 - \epsilon)$ is the confidence level. We denote an ICP by a tuple $(Z_C, f(Z_T, .))$, which consists of the calibration Z_C set and the decision rule given by its training set, Z_T.

As ICP uses only part of the training examples for training its underlying algorithm, and part of the examples for calculating the α-scores, it may result in lower informational efficiency. The Cross Conformal Predictor (CCP) for regression was introduced in [17] and other ensembles of conformal predictors for classification discussed in [15] have also been effectively applied for regression problems. Most of these ensembles methods for regression problems aim to get more informational efficient conformal predictors by combining p-values (or by combining prediction regions). Since the combined p-values need not be uniformly distributed, as a result the final models are not guaranteed to be valid, see [14] for more details.

To assess a conformal predictor we consider validity and efficiency[18, 19]. In our experiments, validity is empirically assessed in terms of calibration plots; the plot of the percentage of errors against $\epsilon \in (0, 1)$ being close to the bisector of the first quadrant with some statistical fluctuations. We use size of the prediction region as the measure of efficiency; lower width implies better informational efficiency and hence indicates a 'better' model.

3. Regression Synergy Conformal Prediction

The objective with this paper is to expand SCP for regression problems. Similar to SCP for classification, for regression problems also we propose to combine non-conformity scores of the calibration set and test example computed from multiple trained models on the individual partitions of the proper training set.

Consider ℓ examples, $Z = \{z_1, ..., z_\ell\}$, drawn from an exchangeable distribution P. Similar to ICP, the training data is first divided into the proper training set $\{Z_T\}$ and the calibration set $\{Z_C\}$, where (T, C) is a partition of $\{1, ..., \ell\}$. Then the proper training data is further divided into M non-empty disjoint subsets and each subset $Z_{T_m}, m = 1, \ldots, M$ is then used for training.

Here $(T_1, ..., T_M)$ is a partition of T. The M predictive models trained on the individual partitions are then used to compute the non-conformity scores for the calibration set denoted by α_{mj}, for $j \in C$ and $m = 1, \ldots, M$. For example,

$$\alpha_{mj} = |y_i - f_m(x_i)|, \tag{4}$$

where $f_m(x)$ is the prediction rule defined by the predictive model trained on the m^{th} part of the training set. The aggregated non-conformity scores across models are then defined as

$$\alpha_j = \frac{1}{M} \sum_m \alpha_{mj}.$$

Let x_{new} (follows the same distribution P) be the object we want to predict, and let $\hat{y}$ be the aggregated estimated value across models. The synergy conformal predictor corresponding to the tuple $(Z_C, f_1(Z_{T_1}, .), ..., f_m A(Z_{T_M}, .))$ is defined as a prediction region (PR) as given in eq. (3).

Algorithm 1: Synergy Conformal Predictor for Regression with data partitioning

Input: training dataset:Z, object to predict:x_{new}, a regression algorithm:A, number of partitions: M
Output: Prediction Region (PR)
Step1: Split the training set into two smaller sets, $\{Z_T\}$ and the calibration set $\{Z_C\}$. (T, C) is a partition of $\{1, ..., |Z|\}$.
Step2: Split the set Z_T into M proper training sets, $\{Z_{T_m}, m = 1, \ldots, M\}$. $(T_1, ..., T_M)$ is a partition of T.
Step3: For each part Z_{T_m}, train and construct the rule to generate non-conformity scores.
Step4: Compute the aggregated non-conformity scores across M models for each example in the calibration set. α_j, for $j \in C$.
Step5: Compute the aggregated estimated value across M regression models for the new object x_{new}, which results in $\hat{y}_{new}$.
Step6: Compute prediction region PR using eq 3.
return PR

The SCP method differs from the ensemble methods discussed previously that combine conformal p-values or prediction regions obtained from different

ICPs, whereas SCP combines non-conformity scores. In the following, we discuss the validity property of Regression SCP. For the regression settings as discussed previously, we can compute non-conformity score for a test point x_{new} for its every potential label y as

$$\alpha_{new} = |y - \hat{y}|, \tag{5}$$

where $\hat{y}$ is the estimated value of x_{new} as defined previously. We denote the p-value associated with the potential label y as

$$p(y) = \frac{|\alpha_i \geq \alpha_{new}, i \in C| + 1}{|C| + 1}, \tag{6}$$

Proposition 1 For every probability distribution P in $\mathbb{R}^p \times \mathbb{R}$ and every significance level $\epsilon > 0$, the following holds true:

$$P\{p(y) \leq \epsilon\} \leq \epsilon. \tag{7}$$

Proof. For unpartitioned data, when the proper training set as a whole is used for training, in that case SCP is exactly ICP and hence valid. We give the same argument as in proof of the Proposition 1 in [16]. Keeping the proper-training set fixed, and the randomization is done over the augmented set, the set consisting of calibration set $z_i, i \in C$ and the new example $z_{new} = (x_{new}, y)$, where y is a postulated label of x_{new} (under the assumption of exchangeability). Without loss of information we can assume that first $|C|$ examples belong to the calibration set then we write the augmented set as $\{z_1, ..., z_{|C|}, z_{|C|+1}\}$, where $z_{|C|+1} = z_{new}$. The inequality $p(y) \leq \epsilon$ holds true if and only if α_{new} is in the largest $\lfloor \epsilon(|C| + 1) \rfloor$ α_i. Since every permutation of $\{z_{\pi(1)}, ..., z_{\pi(|C|+1)}\}$ (from $(|C|+1)!$ possible orderings) is equally probable hence $P\{p(y) \leq \epsilon\} \leq \epsilon$ holds true.

For partitioned training data, SCP can be viewed as a single ICP (hence valid), when the ensemble of M regression methods is considered as one function producing the (aggregated) non-conformity scores. To illustrate this, let us consider partition of the set Z_T, into M subsets, $Z_{T_1}, \ldots, Z_{T_M}$, and let their corresponding decision rules be $f_1(Z_{T_1}, .) \ldots, f_m(Z_{T_M}, .)$ respectively. Define a new decision rule $f(Z_T, .)$ which aggregates the estimated values of an example z, $f(Z_T, z) = \frac{1}{M} \sum_{m=1}^{M} f_m(Z_{T_m}, z)$. Then the non-conformity scores can be computed using the aggregated estimated values. The pair $\{Z_C, f(Z_T, .)\}$ forms an ICP corresponding to the new decision rule $f(Z_T, .)$, hence valid. □

4. Experiments

We evaluate Regression SCP on ten classification datasets from UCI machine learning repository [20], see Table 1. The following four experiments were carried out by randomly dividing each dataset into training and test subset of sizes using a $80:20$ ratio. The training set was further (randomly) divided into proper training set and calibration set of sizes using a $70:30$ ratio. Then the proper training set was randomly partitioned into M equal parts, and a model was trained on each individual part. We used three machine learning algorithms: Support Vector Regression (SVR) (implemented in Python with LIBLINEAR and LIBSVM) and Random Forest (RF) (implemented in Python). The corresponding hyper-parameters were learned using 5-fold cross-validation. The Non-conformity Score (NCS) was computed for each example in the calibration set and test set and this NCS was averaged across models. Then finally, using the averaged NCSs the predictions regions were computed and their width were reported as a measure of efficiency. The whole process was repeated 10 times and the results obtained from various runs were averaged.

Table 1: Datasets from UCI repository that are used in the evaluation. Training refers to the number of examples in the training set, Calibration refers to the number of examples in the calibration set, Test refers to the number of objects in the test set, Features refer to the number of features in the dataset.

Dataset	Training	Calibration	Test	Features
Boston Housing	269	135	102	13
Wine	2612	1306	980	10
Parkinsons	554	278	208	26
Power Plant	5102	2552	1914	4
Energy Efficiency	409	205	154	8
Concrete	549	275	206	8
Grid Stability	5333	2667	2000	12
Super Conduct	11340	5670	4253	81
Condition Based Maintenance	6364	3183	2387	16
Skill Craft	1780	890	668	18

4.1. Experiment 1: Synergy of Conformal Prediction using the Same Machine Learning Algorithm on Partitioned Data

The objective of this experiment is to compare efficiency of SCP with ICP on partitioned and unpartitioned data, using linear Support Vector Regression (linear SVR) as the underlying machine learning method. Inspired by the experiments by [8] we partition the proper training set into three equal partitions in the SCP method, and results for the test sets are shown in Table 2. The first column is the average efficiency of the individual ICP trained on partitioned data with the lowest mean width of the prediction region, referred to as ICP_p. The second column is the mean width of ICP where the proper training set as a whole is used for modeling. The mean width of SCP across 10 repetitions is reported in the third column.

Table 2: Mean width of prediction region calculated at confidence level 90% where linear SVR is used as the underlying machine learning algorithm. ICP_p corresponds to the partition with the lowest mean width, ICP to the mean width where the whole dataset is used, and SCP to the Synergy Conformal Prediction with Regression method.

Dataset	ICP_p	ICP	SCP
Boston	2.02	1.741	1.765
Wine	0.935	0.914	0.923
PD	3.863	3.825	3.78
PowerPlant	0.834	0.83	0.829
Energy	1.215	1.18	1.191
Concrete	2.048	2.085	2.076
GridStability	0.076	0.076	0.074
SuperConduct	1.742	1.728	1.739
CBM	0.161	0.073	0.122
Game	2.238	2.22	2.239

To illustrate the quantitative difference between ICP_p, ICP and SCP, box plots are presented in Figure 1 for Super Conduct dataset. Similar plots for other data sets are available in the supplementary material.

4.2. Experiment 2: Synergy of Conformal Prediction using Different Machine Learning Algorithms on Partitioned Data

The objective of Experiment 2 is similar to Experiment 1, with the difference that multiple machine learning methods were used for the different

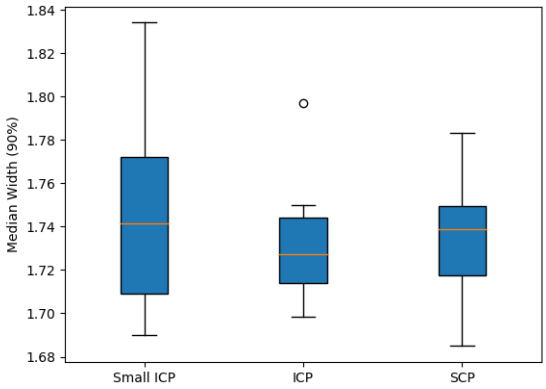


Figure 1: Illustration of the variance in mean width between ICP_p, ICP and SCP for Super Conduct dataset. The linear SVR is used as the underlying machine learning algorithm

partitions. We use the same setup as in the previous experiment with three equal partitions of the proper training set, and with three different machine learning algorithms: linear SVR, Random Forest (RF) and SVR using RBF kernel (RBF-SVR) for each partition. The results are reported in Table 3. To illustrate the quantitative difference between ICP_p and SCP, box plots are presented in Figure 2 for Super Conduct dataset. Similar plots for other data sets are available in the supplementary material.

4.3. Experiment 3: Synergy of Conformal Prediction using Different Machine Learning Algorithms on Unpartitioned Data

The objective of this study is to compare SCP with ICP and CCP trained on the whole training set, in order to show that using different machine learning algorithms for training the same data (proper training set) in SCP also improves the informational efficiency. In this experiment we use three different machine learning algorithms: linear SVR, RF and RBF-SVR; the results are reported in Table 4. The results of the individual ICP with the lowest mean width and SCP are given in first and second column respectively. The third column shows the average informational efficiency of Cross Conformal Prediction (CCP) applied on the whole training set with three fold cross conformal prediction using linear SVR as an underlying machine learning algorithm. The prediction regions for CCP was combined by taking the me-

Table 3: Mean width of prediction region calculated at confidence level 90% for ICP_p and SCP with synergy of three different machine learning algorithms: linear SVR, RF and RBF-SVR. ICP_p corresponds to the partition with the lowest mean width and SCP to the Synergy Conformal Prediction with Regression method.

Dataset	ICP_p	SCP
Boston	1.829	1.47
Wine	0.962	0.918
PD	3.432	3.261
PowerPlant	0.835	0.714
Energy	1.226	0.737
Concrete	2.138	1.497
GridStability	0.077	0.057
SuperConduct	1.775	1.161
CBM	0.142	0.042
Game	2.22	2.132

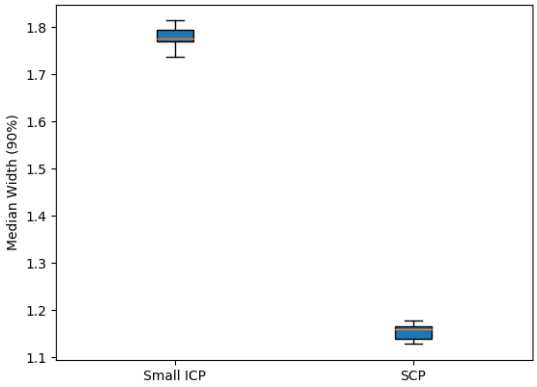


Figure 2: Illustration of the variance in mean width between ICP_p and SCP for Super Conduct dataset. Three different machine learning algorithms were used: linear SVR, RF and RBF-SVR. ICP_p corresponds to the partition with the lowest mean width.

Table 4: Mean width of prediction region calculated at confidence level 90% for ICP, SCP and CCP on unpartitioned data. The first column shows results from the ICP with the lowest mean width, the second column shows results for SCP for an ensemble of different machine learning algorithms (linear, RF and RBF-SVR), and the third column shows results for three fold CCP using linear SVR.

Dataset	ICP	SCP	CCP
Boston	1.486	1.061	1.668
Wine	0.944	0.838	0.94
PD	3.922	3.401	3.568
PowerPlant	0.826	0.678	0.832
Energy	1.179	0.429	1.195
Concrete	2.108	1.245	2.105
GridStability	0.076	0.052	0.076
SuperConduct	1.743	1.013	1.758
CBM	0.064	0.023	0.054
Game	2.015	1.986	2.249

dians of the lower and upper bounds as suggested in [21]. To illustrate the quantitative difference between the ICP, SCP and CCP, box plots are presented in Figure 3 for Super Conduct dataset. Similar plots for other data sets are available in the supplementary material.

4.4. Experiment 4: Calibration of Conformal Predictors

In this section we study the validity of ICP, SCP and CCP. We use the same setup as in Experiment 1, using Random Forest (RF) with 10 trees as the underlying machine learning algorithm. We also train three fold CCP using RF with 10 trees. Calibration plots are available in Figure 4, indicating that all models show little deviation from validity.

5. Discussions

The aim of this paper was to develop SCP for regression and to explore its performance in different settings. In Experiment 1, we considered partitioned proper-training data with the same machine learning algorithm applied on each partition. Similar to SCP for classification, SCP for regression also succeeded in combining the models and in most of the cases obtain improved

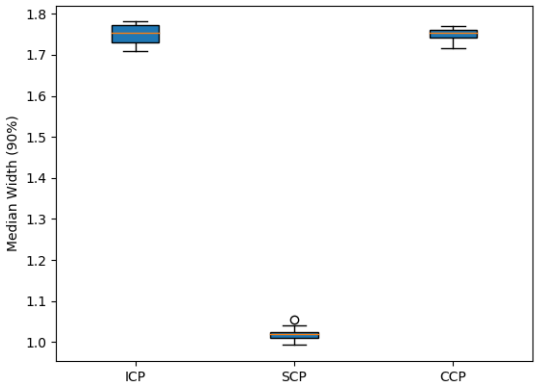
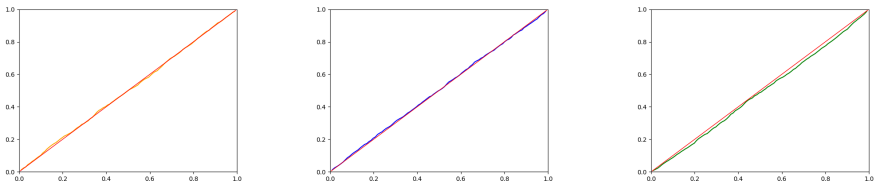


Figure 3: Illustration of the variance in mean width between ICP, SCP and CCP for Super Conduct dataset. SCP used different machine learning methods for each partition (linear SVR, RF and RBF-SVR) whereas ICP and CCP used linear SVR.



(a) ICP (b) SCP (c) CCP

Figure 4: Calibration plots of ICP, SCP and CCP for Super Conduct dataset using random forest with 10 trees for training all the underlying models

efficiency as compared with each individual ICP, and we also observe that SCP is comparable with an ICP trained on the entire dataset (see Table 2). In Experiment 2, we considered partitioned proper training data with different machine learning algorithms applied on each partition. We showed that SCP has lower mean width when comparing with ICPs trained on partitioned data. In Experiment 3, we considered unpartitioned proper training data with different machine learning algorithms and showed that SCP on unpartitioned data has lower mean width than an ICP or CCP trained on the entire dataset using linear SVR. In Experiment 4, we showed that ICP, SCP and CCP indicates valid models (close to the bisector of the first quadrant). The property to be able to produce valid models while aggregating results over multiple models contrasts to previous aggregation methods, such as reported in [14, 22].

Our results suggests that the SCP method can be applied to problem settings where large datasets can be partitioned, and benefit from reduced modeling time while still getting efficient and valid prediction intervals. The results further promotes the applicability of SCP in federated setting, where data is located in different locations, and the data cannot be pooled due to privacy, regulatory, or practical reasons. We also envision that SCP could be a foundation upon which implementations can train individual partition in parallel, such as in locality-aware Big Data frameworks. The main drawback of SCP is that it requires a calibration set that is shared between individual partitions which is likely not a big concern when partitioning data, but which does constitute a an additional step when working with data sources that are distributed.

6. Conclusions and Future Directions

We introduced Synergy Conformal Prediction for regression and evaluated its validity and informational efficiency using various underlying machine learning algorithms. The key outcome is that SCP offers an alternative to pooling distributed data when using inductive conformal predictors, with reduced training time as models can be trained in parallel and predictions aggregated, while still preserving validity. We also demonstrated that SCP has comparable efficiency to commonly used ICP and CCP approaches, making it widely applicable as a valid confidence predictor. Future directions when working on partitioned data include (i) studying the effect of the number and size of data partitions as well as overlapping partitions (ii) evaluating

the effect of different non-conformity scores and different underlying machine learning algorithms with individual partitions.

Acknowledgements

This project received financial support from the Swedish Foundation for Strategic Research (SSF) as part of the HASTE project under the call 'Big Data and Computational Science'. The computations were performed on resources provided by SNIC through Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under project SNIC 2019/8 − 15.

References

[1] I. A. T. Hashem, I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani, S. U. Khan, The rise of big data on cloud computing: Review and open research issues, Information systems 47 (2015) 98–115.

[2] Z. D. Stephens, S. Y. Lee, F. Faghri, R. H. Campbell, C. Zhai, M. J. Efron, R. Iyer, M. C. Schatz, S. Sinha, G. E. Robinson, Big data: astronomical or genomical?, PLoS biology 13 (7) (2015) e1002195.

[3] A. Gandomi, M. Haider, Beyond the hype: Big data concepts, methods, and analytics, International Journal of Information Management 35 (2) (2015) 137–144.

[4] M. C. Elish, danah boyd, Situating methods in the magic of big data and ai, Communication Monographs 85 (1) (2018) 57–80. arXiv:https://doi.org/10.1080/03637751.2017.1375130, doi:10.1080/03637751.2017.1375130.
URL `https://doi.org/10.1080/03637751.2017.1375130`

[5] L. Zhou, S. Pan, J. Wang, A. V. Vasilakos, Machine learning on big data: Opportunities and challenges, Neurocomputing 237 (2017) 350 – 361. doi:https://doi.org/10.1016/j.neucom.2017.01.026.
URL `http://www.sciencedirect.com/science/article/pii/S0925231217300577`

[6] D. Peteiro-Barral, B. Guijarro-Berdiñas, A survey of methods for distributed machine learning, Progress in Artificial Intelligence 2 (1) (2013) 1–11.

[7] X. Meng, J. Bradley, B. Yavuz, E. Sparks, S. Venkataraman, D. Liu, J. Freeman, D. Tsai, M. Amde, S. Owen, et al., Mllib: Machine learning in apache spark, The Journal of Machine Learning Research 17 (1) (2016) 1235–1241.

[8] V. Vapnik, R. Izmailov, Synergy of monotonic rules, The Journal of Machine Learning Research 17 (1) (2016) 4722–4754.

[9] V. Vovk, A. Gammerman, G. Shafer, Algorithmic learning in a random world, Springer Science & Business Media, 2005.

[10] V. Vovk, Transductive conformal predictors, in: IFIP International Conference on Artificial Intelligence Applications and Innovations, Springer, 2013, pp. 348–360.

[11] H. Papadopoulos, Inductive conformal prediction: Theory and application to neural networks, in: Tools in artificial intelligence, InTech, 2008.

[12] V. Vovk, Cross-conformal predictors, Annals of Mathematics and Artificial Intelligence 74 (1-2) (2015) 9–28.

[13] L. Carlsson, M. Eklund, U. Norinder, Aggregated conformal prediction, in: IFIP International Conference on Artificial Intelligence Applications and Innovations, Springer, 2014, pp. 231–240.

[14] H. Linusson, U. Norinder, H. Boström, U. Johansson, T. Löfström, On the calibration of aggregated conformal predictors, in: A. Gammerman, V. Vovk, Z. Luo, H. Papadopoulos (Eds.), Proceedings of the Sixth Workshop on Conformal and Probabilistic Prediction and Applications, Vol. 60 of Proceedings of Machine Learning Research, PMLR, Stockholm, Sweden, 2017, pp. 154–173.
URL `http://proceedings.mlr.press/v60/linusson17a.html`

[15] N. Gauraha, O. Spjuth, Synergy conformal prediction.
URL `urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-360504`

[16] H. Papadopoulos, K. Proedrou, V. Vovk, A. Gammerman, Inductive confidence machines for regression, in: European Conference on Machine Learning, Springer, 2002, pp. 345–356.

[17] H. Papadopoulos, Cross-conformal prediction with ridge regression, in: International Symposium on Statistical Learning and Data Sciences, Springer, 2015, pp. 260–270.

[18] V. Vovk, V. Fedorova, I. Nouretdinov, A. Gammerman, Criteria of efficiency for conformal prediction, in: Symposium on Conformal and Probabilistic Prediction with Applications, Springer, 2016, pp. 23–39.

[19] F. Svensson, N. Aniceto, U. Norinder, I. Cortes-Ciriano, O. Spjuth, L. Carlsson, A. Bender, Conformal regression for quantitative structure–activity relationship modeling—quantifying prediction uncertainty, Journal of Chemical Information and Modeling 58 (5) (2018) 1132–1140. doi:10.1021/acs.jcim.8b00054.
URL https://doi.org/10.1021/acs.jcim.8b00054

[20] M. Lichman, et al., Uci machine learning repository (2013).

[21] S. Park, D. V. Budescu, Aggregating multiple probability intervals to improve calibration, Judgment and Decision Making 10 (2) (2015) 130.

[22] L. Carlsson, M. Eklund, U. Norinder, Aggregated conformal prediction, in: L. Iliadis, I. Maglogiannis, H. Papadopoulos, S. Sioutas, C. Makris (Eds.), Artificial Intelligence Applications and Innovations, Springer Berlin Heidelberg, Berlin, Heidelberg, 2014, pp. 231–240.