A Metric to Improve the Robustness of Conformal Predictors in the Presence of Error Bars

Andrea Murari[1], Saeed Talebzadeh[2], Jesús Vega[3],
Emmanuele Peluso[2(✉)], Michela Gelfusa[2], Michele Lungaroni[2],
and Pasqualino Gaudio[2]

[1] Consorzio RFX (CNR, ENEA, INFN), Acciaierie Venete SpA,
Universita' di Padova, Corso Stati Uniti 4, 35127 Padua, Italy
murari@igi.cnr.it
[2] University of Rome "Tor Vergata", Via del Politecnico 1, 00133 Rome, Italy
talebzadeh.saeed@gmail.com,
{emmanuele.peluso,michele.lungaroni}@uniroma2.it,
{gelfusa,gaudio}@ing.uniroma2.it
[3] Asociación EURATOM/CIEMAT para Fusión, 28040 Madrid, Spain
jesus.vega@ciemat.es

Abstract. Conformal predictors, currently applied to many problems in various fields determine precise levels of confidence in new predictions on the basis only of the information present in the past data, without making recourse to any assumptions except that the examples are generated independently from the same probability distribution. In this paper, the robustness of their results is assessed for the cases in which the data are affected by error bars. This is the situation typical of the physical sciences, whose data are often the results of complex measurement procedures, unavoidably affected by noise. Assuming the noise presents a normal distribution, the Geodesic Distance on Gaussian Manifolds provides a statistical principled and quite effective method to handle the uncertainty in the data. A series of numerical tests prove that adopting this metric in conformal predictors improves significantly their performance, compared to the Euclidean distance, even for relatively low levels of noise.

Keywords: Conformal predictors · Geodesic distance · Inference methods · Error bars

1 Conformal Predictors and Measurement Errors

Machine-learning methods work often very well and have found many applications in both the public and the private sector. On the other hand, the reliability of their performance is typically proven asymptotically and is therefore not very useful in practice. Conformal predictors, which perform competitively in terms of success rates, include from their conception simple and useful measures of confidence [1]. Conformal prediction can be based on any technique of point prediction for classification or regression, including support-vector machines, decision trees, neural networks and Bayesian

methods. Starting from the point prediction tool, the conformal predictor consists of building a nonconformity measure, which determines how unusual an example is relative to previous examples. The conformal algorithm, based on the statistical concept of the p-values, turns this nonconformity measure into prediction regions. Given a nonconformity measure, the conformal algorithm produces a prediction region Uε for every probability of error ε. The region Uε is a (1−ε)-prediction region; it classifies the next example with probability at least 1−ε. Therefore conformal predictors are conservatively valid, which means that the probability they make a mistake when their output is at confidence level 1−ε is not greater than ε.

In most of the non-conformity measures utilised by conformal predictors, the Euclidean distance is implicitly assumed to be the proper metric to adopt in the calculation of the non-conformity measure and the p-values. The Euclidean distance has a precise geometrical meaning and a very long historical pedigree. However, it implicitly requires considering all data as single infinitely precise values. This assumption can be appropriate in other applications but it is obviously not the case in physics, since all the measurements typically present an error bar. An alternative idea is to use a new distance between data, which would take into account the measurement uncertainties. The causes of uncertainties in the measurements are typically many, which from a statistical point of view can be considered random variables. As a consequence, their global contribution can be often modelled as a noise of normal distribution. The idea, behind the approach proposed in this paper, consists therefore of considering the measurements not as points, but as Gaussian distributions [2]. Modelling measurements not as point values, but as Gaussian distributions, requires defining a distance between Gaussians. This distance must be the Geodesic on the Gaussian Manifold (GDGM) of the measurements and can be expressed as a closed formula (see Sect. 3) [3]. As shown in the rest of the paper, adopting this geodesic distance can increase significantly the accuracy of traditional conformal predictors, even when the data are affected by a very limited level of noise.

With regard to the structure of the paper, next Section provides a short introduction to the general framework of conformal prediction. The mathematical background to the main mathematical tool introduced in the paper: the Geodesic Distance on Gaussian Manifolds, is the subject of Sect. 3. The proposed method is assessed with a series of numerical tests using a toy model described in Sect. 4. Section 5 reports in detail the results of the numerical tests. Conclusions and lines of future work are provided in the last Section of the paper.

2 The Framework of Conformal Prediction for Classification

The task of classification basically consists of attributing objects to different classes. Mathematically this can be formalised by considering successive ordered pairs $(\mathbf{x}_1, y_1)$, $(\mathbf{x}_2, y_2)$…….which are called examples. Each example consists of an object $\mathbf{x}_i$ and its label y_i, where the former represents the feature vector that describes the object i. The objects are elements of a measurable space X called the *object space*; the labels are elements of a measurable space Y called the *label space*. It is common practice to adopt a more compact notation, according to which z_i indicate the ordered pair $(\mathbf{x}_i, y_i)$, and $Z := X \times Y$ is defined as the *example space*.

Many machine learning tools are available to perform classification. On the other hand, as mentioned earlier, the vast majority of them cannot easily quantify the quality of their predictions. On the contrary, conformal predictors have been conceived explicitly to quantify the reliability of their predictions. They achieve this on the basis of the past examples. To this end, for each new sample to classify, it is necessary to measure how different the new one is from the old examples. In this perspective, a nonconformity measure is defined, which allows calculating a nonconformity score to estimate how different a new example is from a bag of old ones. A bag of size $n \in N$ is a collection of n elements some of which may be identical. In this paper, the notation $< z_1,\ldots, z_n >$ indicates a bag of n elements.

Given a nonconformity measure A and a bag $< z_1,\ldots,z_n >$, the nonconformity score can be calculated as:

$$\alpha_i := A(\langle z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n \rangle, z_i) \tag{1}$$

for each example z_i in the bag. Because nonconformity measures are not absolute but relative, the numerical value of α_i does not, by itself, determines how unusual z_i is according to the measure A. To really quantify how unusual a sample is, it is necessary to compare α_I with the nonconformity measures α_j of the other members of the bag. The p-value is a convenient and statistically sound way of calculating how anomalous a new example is. By definition the p-value is the fraction

$$Pval = \frac{\#\{j = 1, \ldots, n : \alpha_j \geq \alpha_i\}}{n} \tag{2}$$

This indicator, which lies between 1/n and 1, is the fraction of the examples in the bag as non conforming as z_i and in literature is called p-value of the element z_i ($p_{val}(z_i)$). The symbol "#" stands in fact for the number of elements "j" in the collection having a nonconformity score higher or at least the same nonconformity of the element "i". The lower the p-value, i.e. the closer to its lower bound 1/n ("j" includes "i" in fact) for
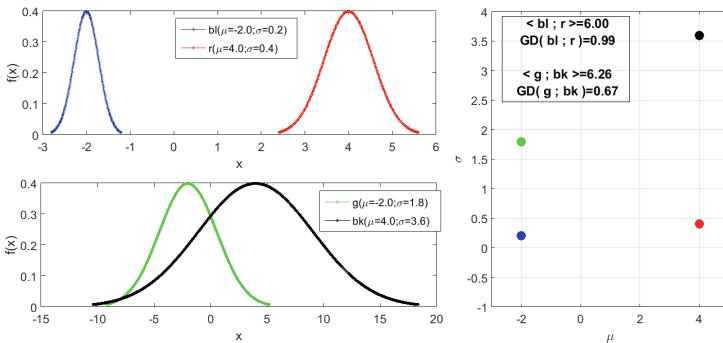

Fig. 1. Examples to illustrate how the GDGM determines the distance between two Gaussians. The two couples of pdf in the figure have the same mean but different σ. The geodesic distance between the two with higher σ is much smaller.

large n, the more non conforming z_i is and the more likely it can be considered as an outlier; this means that z_i is not representative of the typical member of the bag. If the p-value is large, i.e. close to its upper bound 1, then z_i is very conforming or very representative of the typical member of the bag. The new sample is attributed to the class with the highest p-value.

On the basis of the p-values, conformal predictors allow calculating, for each new classification, two indicators, confidence and credibility, which quantify the reliability of the prediction. Credibility is defined as the largest p-value; confidence is defined as 1-2nd largest p-value. Confidence can be interpreted as the probability that the prediction, corresponding to the maximal *p*-value, is correct. A low credibility, typically less than 0.05, intuitively means that either the training set is non random or the test object is not representative of the training set. If the maximum *p*-value appears in more than one class, an ambiguity is present and the algorithm is not able to classify the sample. It is important to emphasize that confidence and credibility of the prediction play an analogous role to the observed level of significance in statistical parameter tests.

3   Geodesic Distance on Gaussian Manifolds

As mentioned in the previous section, in the natural sciences the data available are typically the result of experimental measurements. In this context, all measurements are affected by uncertainties referred to as error bars. The sources of this uncertainty are normally quite many and therefore it is more than reasonable to assume that the pdf of the noise is normal. Each measurement can therefore be modelled as a probability density function (pdf) of the Gaussian type, determined by its mean μ and its standard deviation σ:

$$p(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \qquad (3)$$

It is normal practice to assume that the experimental measured value is the mean of the pdf, since this is the most likely value of the pdf. The standard deviation can be determined independently from the knowledge of the instrumentation.

The set of normal distributions can therefore be modelled as a two dimensional space, or better a two dimensional manifold, parameterized by μ and σ. Modelling measurements not as point values, but as Gaussian distributions, requires defining a distance between Gaussians. The most appropriate definition of distance between Gaussian distributions is the geodesic distance (GDGM), on the probabilistic manifold containing the data, which is not a Euclidean but a Riemannian space. This geodesic distance on the Gaussian manifold can be calculated using the Fischer-Rao metric [3, 4]. For two univariate Gaussian distributions $(p_1(x|\mu_1, \sigma_1))$ and $(p_2(x|\mu_2, \sigma_2))$, parameterised by their mean μ_i and standard deviations $\sigma_i (i = 1, 2)$, the geodesic distance GDGM is given by:

$$GD(p_1||p_2) = \sqrt{2}ln\frac{1+\delta}{1-\delta} = 2\sqrt{2}tanh^{-1}\delta, \, where \, \delta = [\frac{(\mu_1-\mu_2)^2+2(\sigma_1-\sigma_2)^2}{(\mu_1-\mu_2)^2+2(\sigma_1+\sigma_2)^2}]^{\frac{1}{2}}$$

(4)

As will be shown in detail in the next sections, the replacement of the Euclidean distance with the GDGM improves significantly the robustness of the classification compared to the case of the Euclidean distance. In Fig. 1 a graphical example of the improvement obtained using the metric in Eq. (4) is shown. Considering a Cartesian coordinate system (μ, σ), where each point represents a Gaussian distribution, the Euclidean distance between the four points, so between the four distributions, is higher between the two wider distributions. On the other hand, considering the Geodesic Distance, the lower distance is obtained considering the wider distributions. This behaviour reflects the physical interpretation according to which physical quantities having higher error bars, are to be considered closer and more similar than those with narrower error bars.

4 A Toy Model

To exemplify and prove the usefulness of the method proposed in this paper, a series of numerical test has been performed. They are based on a toy model already introduced in [5]. The simplicity of the model allows appreciating both the nature of the problem and the advantages of adopting the proposed metric, the GDGM. The classification task consists of classifying points on a straight line, on which three classes have been defined. The problem is represented graphically in Fig. 2. The aim is to classify the new point Q with confidence and credibility.

For the purpose of this example, the classification is based on the nearest neighbour. Mathematically, given a "*bag*" $\{z_1,...,z_{n-1}\}$, where each z_i consists of a feature vector $\mathbf{x}_i$ and a non-numerical label y_i, when a new example $z_n = (\mathbf{x}_n, y_n)$ becomes available for classification, its feature vector $\mathbf{x}_n$ is known but its label y_n is not. The nearest-neighbour method finds the $\mathbf{x}_i$ closest to $\mathbf{x}_n$ and its label y_i becomes the prediction of y_n. A natural way to measure the nonconformity of the new example z_n with respect to the old examples z_i consists of comparing $\mathbf{x}$'s distances to old objects with the same label to its distance to old objects with a different label. For example, the nonconformity scores can be defined as:

$$\alpha_i = \frac{d_{sl}}{d_{dl}}$$

(5)

$$d_{sl} = \min\{|x_j - x_i| : 1 \leq j \leq n \, \& \, j \neq i \, \& \, y_j = y_i\}$$

$$d_{dl} = \min\{|x_j - x_i| : 1 \leq j \leq n \, \& \, j \neq i \, \& \, y_j \neq y_i\}$$

For the new point Q = 14.85 shown in Fig. 2 (the non-conformity measurement is presented in Table 1), therefore the P values credibility and confidence assume the values 1 and 0.9844, respectively; and point Q belongs to Class C.
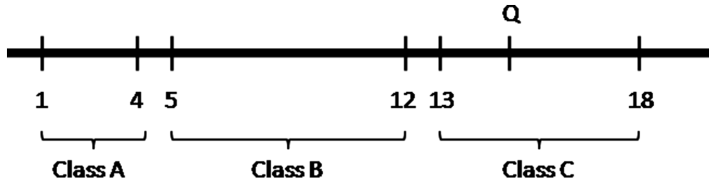
Fig. 2. The toy model. A new point Q must be classified as belonging to class A, B or C.

Table 1. Non-conformity measurements for point Q = 14.85

Object i	α if Y = A	α if Y = B	α if Y = C		Object i	α if Y = A	α if Y = B	α if Y = C
1	0.0625	0.0625	0.0625		10	0.0833	0.0833	0.0833
1.25	0.0667	0.0667	0.0667		10.25	0.0909	0.0909	0.0909
1.5	0.0714	0.0714	0.0714		10.5	0.1	0.1	0.1
1.75	0.0769	0.0769	0.0769		10.75	0.1111	0.1111	0.1111
2	0.0833	0.0833	0.0833		11	0.125	0.125	0.125
2.25	0.0909	0.0909	0.0909		11.25	0.1429	0.1429	0.1429
2.5	0.1	0.1	0.1		11.5	0.1667	0.1667	0.1667
2.75	0.1111	0.1111	0.1111		11.75	0.2	0.2	0.2
3	0.125	0.125	0.125		12	0.25	0.25	0.25
3.25	0.1429	0.1429	0.1429		13	0.25	0.25	0.25
3.5	0.1667	0.1667	0.1667		13.25	0.2	0.2	0.2
3.75	0.2	0.2	0.2		13.5	0.1852	0.1852	0.1667
4	0.25	0.25	0.25		13.75	0.2273	0.2273	0.1429
5	0.25	0.25	0.25		14	0.2941	0.2941	0.125
5.25	0.2	0.2	0.2		14.25	0.4167	0.4167	0.1111
5.5	0.1667	0.1667	0.1667		14.5	0.7143	0.7143	0.1
5.75	0.1429	0.1429	0.1429		14.75	2.5	2.5	0.0364
6	0.125	0.125	0.125		15	1.6667	1.6667	0.05
6.25	0.1111	0.1111	0.1111		15.25	0.625	0.625	0.0769
6.5	0.1	0.1	0.1		15.5	0.3846	0.3846	0.0714
6.75	0.0909	0.0909	0.0909		15.75	0.2778	0.2778	0.0667
7	0.0833	0.0833	0.0833		16	0.2174	0.2174	0.0625
7.25	0.0769	0.0769	0.0769		16.25	0.1786	0.1786	0.0588
7.5	0.0714	0.0714	0.0714		16.5	0.1515	0.1515	0.0556
7.75	0.0667	0.0667	0.0667		16.75	0.1316	0.1316	0.0526
8	0.0625	0.0625	0.0625		17	0.1163	0.1163	0.05
8.25	0.0588	0.0588	0.0588		17.25	0.1042	0.1042	0.0476
8.5	0.0556	0.0556	0.0556		17.5	0.0943	0.0943	0.0455
8.75	0.0588	0.0588	0.0588		17.75	0.0862	0.0862	0.0435
9	0.0625	0.0625	0.0625		18	0.0794	0.0794	0.0417
9.25	0.0667	0.0667	0.0667		14.85	108.5	...	...
9.5	0.0714	0.0714	0.0714		14.85	...	28.5	...
9.75	0.0769	0.0769	0.0769		14.85	...	...	0.0351
Continue -------->					p-value	0.0156	0.0156	1

In the previous example, the conformity measure of Eq. (5) has been calculated using the Euclidean distance between the various points. All the derived quantities are therefore also based on this metric. In the case of measurements affected by noise, the Euclidean metric is not adequate and adopting the GDGM provides several improvements as discussed in the next section.

5   Results of the Numerical Tests

In order to assess the potential of the GDGM metric to counteract the effect of noise, a series of systematic tests has been performed using the toy model introduced in the previous section. To this end, a series of points have been automatically generated along the straight line of Fig. 2. These are to be considered the right values of the physical quantity to measure. Then Gaussian noise, with zero mean and standard deviations equals to a percentage (10 %,20 %,…) of the value itself, has been added to the previously generated points. Adding this noise to the data provides the actual values to be considered as the available measurements, affected by additive noise of Gaussian distribution. These points have been then classified with the nonconformity measure based on the next neighbour criterion using both the Euclidean distance and the GDGM. The results have been reported in Table 2 for the Euclidean distance as metric and in Table 3 for the GDGM as metric.

Table 2. Classification using the Euclidean distance to calculate the nearest neighbour. The first column reports the accuracy (Acc.); the second the credibility (Cred.) and the third the confidence (Conf). The following column reports the same quantities but for different levels of noise. The top of the table reports the average values for all the 50 points.

Euclidean								
Noise=10%			Noise=20%			Noise=30%		
Acc.	Cred.	Conf.	Acc.	Cred.	Conf.	Acc.	Cred.	Conf.
Average values for all runs:								
84.67	0.80	0.89	72.77	0.77	0.86	65.03	0.76	0.84
Average values for each run:								
87.18	0.84	0.98	71.79	0.71	0.83	69.23	0.81	0.81
76.92	0.77	0.81	79.49	0.82	0.91	69.23	0.76	0.83
89.74	0.81	0.93	79.49	0.78	0.86	58.97	0.71	0.76
89.74	0.78	0.88	66.67	0.83	0.86	53.85	0.72	0.73
79.49	0.80	0.93	58.97	0.74	0.78	71.79	0.69	0.83
84.62	0.83	0.93	74.36	0.74	0.86	64.10	0.75	0.81
84.62	0.77	0.91	82.05	0.78	0.91	71.79	0.76	0.81
84.62	0.81	0.96	66.67	0.64	0.81	53.85	0.70	0.76
92.31	0.81	0.93	66.67	0.82	0.88	58.97	0.81	0.83
79.49	0.78	0.81	82.05	0.79	0.91	66.67	0.73	0.83
79.49	0.81	0.88	76.92	0.80	0.86	69.23	0.80	0.86

Table 2. (*Continued*)

87.18	0.84	0.96	87.18	0.83	0.96	66.67	0.75	0.86
87.18	0.78	0.88	61.54	0.68	0.76	66.67	0.75	0.88
87.18	0.79	0.91	69.23	0.87	0.91	64.10	0.86	0.91
82.05	0.78	0.88	61.54	0.81	0.83	69.23	0.86	0.91
82.05	0.76	0.86	69.23	0.72	0.78	61.54	0.74	0.81
74.36	0.76	0.86	64.10	0.81	0.88	69.23	0.89	0.93
74.36	0.80	0.83	79.49	0.80	0.88	64.10	0.74	0.86
82.05	0.80	0.86	82.05	0.79	0.86	56.41	0.84	0.88
87.18	0.78	0.88	76.92	0.83	0.91	74.36	0.72	0.81
87.18	0.75	0.86	66.67	0.77	0.88	64.10	0.74	0.81
89.74	0.88	0.96	69.23	0.81	0.88	71.79	0.83	0.86
79.49	0.83	0.86	74.36	0.71	0.86	61.54	0.64	0.83
92.31	0.81	0.88	66.67	0.75	0.83	66.67	0.78	0.88
87.18	0.87	0.96	82.05	0.75	0.88	69.23	0.71	0.83
84.62	0.78	0.88	69.23	0.83	0.86	66.67	0.75	0.83
79.49	0.82	0.91	74.36	0.74	0.91	58.97	0.72	0.78
79.49	0.80	0.83	69.23	0.73	0.83	69.23	0.83	0.91
84.62	0.75	0.93	74.36	0.80	0.88	76.92	0.82	0.91
87.18	0.84	0.88	66.67	0.80	0.86	58.97	0.65	0.78
92.31	0.81	0.96	76.92	0.82	0.91	66.67	0.71	0.78
94.87	0.83	0.96	84.62	0.86	0.88	71.79	0.75	0.86
76.92	0.75	0.78	69.23	0.76	0.88	64.10	0.76	0.83
84.62	0.86	0.91	76.92	0.82	0.91	51.28	0.65	0.76
79.49	0.79	0.86	79.49	0.76	0.91	58.97	0.75	0.78
87.18	0.81	0.93	74.36	0.75	0.88	69.23	0.80	0.88
87.18	0.85	0.91	74.36	0.78	0.81	61.54	0.68	0.73
84.62	0.75	0.91	69.23	0.72	0.76	58.97	0.71	0.76
79.49	0.79	0.86	79.49	0.76	0.91	58.97	0.75	0.78
87.18	0.81	0.93	74.36	0.75	0.88	69.23	0.80	0.88
87.18	0.85	0.91	74.36	0.78	0.81	61.54	0.68	0.73
84.62	0.75	0.91	69.23	0.72	0.76	58.97	0.71	0.76
89.74	0.77	0.88	66.67	0.74	0.86	79.49	0.81	0.88
87.18	0.78	0.86	76.92	0.70	0.86	76.92	0.78	0.86
92.31	0.85	0.96	64.10	0.80	0.86	69.23	0.80	0.88
69.23	0.65	0.76	71.79	0.75	0.86	69.23	0.76	0.86
84.62	0.83	0.88	74.36	0.78	0.81	61.54	0.81	0.91
79.49	0.65	0.78	79.49	0.78	0.88	69.23	0.76	0.91
89.74	0.80	0.96	71.79	0.80	0.88	64.10	0.78	0.91
79.49	0.79	0.81	61.54	0.67	0.73	51.28	0.75	0.83
89.74	0.87	0.93	76.92	0.84	0.88	64.10	0.76	0.86
87.18	0.80	0.86	64.10	0.66	0.76	51.28	0.75	0.81
84.62	0.86	0.91	79.49	0.73	0.88	76.92	0.79	0.86
89.74	0.78	0.91	76.92	0.81	0.88	51.28	0.68	0.81

Table 3. Classification using the Geodesic Distance on Gaussian Manifolds distance to calculate the nearest neighbour. The first column reports the accuracy (Acc.); the second the credibility (Cred.) and the third the confidence (Conf.). The following column reports the same quantities but for different levels of noise. The top of the table reports the average values for all the 50 points.

Geodesic								
Noise=10%			Noise=20%			Noise=30%		
Acc.	Cred.	Conf.	Acc.	Cred.	Conf.	Acc.	Cred.	Conf.
Average values for all runs:								
89.33	**0.75**	**0.95**	**80.67**	**0.71**	**0.94**	**74.46**	**0.68**	**0.93**
Average values for each run:								
87.18	0.79	0.98	79.49	0.71	0.91	82.05	0.76	0.93
89.74	0.74	0.93	87.18	0.75	0.98	79.49	0.63	0.96
94.87	0.75	0.98	87.18	0.70	0.93	71.79	0.74	0.91
92.31	0.76	0.91	74.36	0.65	0.93	64.10	0.57	0.83
82.05	0.76	0.98	76.92	0.72	0.98	74.36	0.62	0.93
87.18	0.78	0.96	82.05	0.71	0.93	79.49	0.65	0.96
89.74	0.72	0.96	87.18	0.75	0.96	87.18	0.73	0.96
87.18	0.77	0.98	76.92	0.65	0.91	69.23	0.63	0.91
92.31	0.75	0.96	71.79	0.71	0.93	71.79	0.74	0.96
89.74	0.79	0.91	87.18	0.70	0.96	82.05	0.70	0.98
89.74	0.79	0.98	89.74	0.81	0.98	76.92	0.69	0.93
89.74	0.73	0.98	89.74	0.70	0.98	76.92	0.69	0.96
94.87	0.74	0.96	76.92	0.68	0.91	66.67	0.63	0.86
89.74	0.77	0.93	71.79	0.66	0.93	69.23	0.65	0.96
87.18	0.75	0.93	69.23	0.64	0.91	71.79	0.67	0.93
92.31	0.73	0.96	82.05	0.73	0.91	71.79	0.71	0.91
82.05	0.72	0.96	69.23	0.66	0.93	71.79	0.70	0.96
84.62	0.77	0.93	84.62	0.72	0.98	71.79	0.64	0.93
89.74	0.76	0.93	89.74	0.79	0.93	66.67	0.79	0.98
92.31	0.75	0.93	84.62	0.72	0.98	89.74	0.78	0.96
92.31	0.78	0.91	71.79	0.72	0.93	74.36	0.69	0.91
92.31	0.79	0.98	76.92	0.69	0.96	82.05	0.79	0.96
87.18	0.73	0.93	82.05	0.72	0.93	66.67	0.62	0.91
97.44	0.75	0.93	82.05	0.75	0.98	74.36	0.62	0.96
87.18	0.76	0.96	89.74	0.70	0.96	82.05	0.71	0.96

Table 3. (*Continued*)

92.31	0.75	0.96	74.36	0.66	0.91	79.49	0.72	0.96
82.05	0.74	0.93	74.36	0.64	0.93	76.92	0.69	0.96
82.05	0.76	0.91	76.92	0.66	0.91	69.23	0.60	0.91
89.74	0.68	0.98	74.36	0.66	0.93	84.62	0.67	0.98
92.31	0.78	0.93	74.36	0.70	0.93	71.79	0.63	0.88
89.74	0.77	0.98	82.05	0.67	0.96	76.92	0.69	0.88
94.87	0.73	0.96	89.74	0.69	0.93	79.49	0.66	0.93
87.18	0.73	0.93	76.92	0.65	0.96	69.23	0.66	0.88
87.18	0.75	0.93	82.05	0.67	0.96	61.54	0.56	0.88
84.62	0.72	0.91	84.62	0.70	0.96	71.79	0.69	0.91
92.31	0.74	0.98	79.49	0.74	0.93	74.36	0.64	0.91
87.18	0.72	0.93	87.18	0.75	0.93	74.36	0.70	0.86
87.18	0.73	0.93	84.62	0.67	0.93	71.79	0.63	0.88
92.31	0.74	0.96	71.79	0.70	0.93	87.18	0.72	0.96
89.74	0.74	0.93	84.62	0.69	0.93	82.05	0.68	0.91
94.87	0.73	0.98	74.36	0.76	0.96	74.36	0.63	0.93
82.05	0.71	0.91	79.49	0.71	0.93	79.49	0.71	0.93
87.18	0.76	0.91	82.05	0.74	0.96	66.67	0.73	0.96
87.18	0.68	0.88	89.74	0.80	0.98	71.79	0.66	0.91
92.31	0.74	0.98	79.49	0.75	0.96	69.23	0.61	0.96
89.74	0.76	0.93	71.79	0.61	0.83	64.10	0.72	0.93
94.87	0.77	0.98	87.18	0.75	0.98	76.92	0.72	0.98
87.18	0.74	0.96	76.92	0.65	0.91	66.67	0.65	0.96
87.18	0.78	0.93	87.18	0.72	0.96	84.62	0.75	0.93
92.31	0.76	0.98	87.18	0.77	0.98	64.10	0.67	0.93

The results reported in Tables 2 and 3 indicate that the GDGM provides a clear improvement in the success rate of the classification. Table 2 shows how the performance of conformal predictors degrade with increasing levels of noise. It is important also to notice how the indicators of the quality of the prediction, confidence and credibility, tend to overestimate the reliability of the classification when significant level of noise is present. Table 3 reports the clear improvement in both performance and reliability of the quality indicators when the Euclidean distance is replaced with GDGM. Another important consideration is the fact that, adopting the GDGM metric does not cause any degradation of performance when the data are not affected by noise.

6   Conclusions

In many applications of conformal predictors, the Euclidean distance is explicitly or implicitly adopted as the right metric. In the case of experimental measurements typical of the physical sciences, the data are affected by noise of normal distribution. In this situation, the GDGM proves to be a better metric, to be used in the definition of the non-conformity measure. The calculation of the nonconformity measure and of the p-values using the GDGM provides significantly more reliable classifications, by reducing the adverse effects of the noise. The reported results using the GDGM have been obtained using a desktop computer with two Xeon E5520 @2.27 GHz processors and 24 GB of RAM, and required an average of one minute for each test performed, for a total of 50 min for all 50 points. The computational cost is therefore very similar to the one required to perform the calculations with the Euclidean distance.

With regard to future developments, it would be important to apply the same approach to different pdfs: particularly relevant would be the case of the Poisson distribution, since in practice many detectors work in photon counting or particle counting mode. Another very interesting application would be the case in which the pdf of the noise is not known. This situation has practical applications because in many experimental situations the uncertainties in the measurements can be quantified with an interval but without any additional specification. Therefore the real value is expected to fall in a certain interval but no additional information is available. In this case the implementation of an appropriate form of uncertain probability is expected to produce improvements in the classification of conformal predictors comparable to the case of the GDGM for the case of measurements affected by Gaussian noise.

In terms of practical applications, the mathematics of conformal predictors can be applied to most classifiers, including Fuzzy ones [6]. Therefore the approach can be of extreme help in all the cases, such as disruptions in Tokamaks, where classification is a particularly problematic and difficult task also due to the uncertainties in the measurements [7, 8].

References

1. Vovk, V., Gammerman, A., Shafer, G.: Algorithmic Learning in a Random World. Springer, New York (2005)
2. Katz, J.O., Rohlf, F.J.: Function-point cluster analysis. Syst. Zool. **22**(3), 295–301 (1973)
3. Amari, S., Nagaoka, H.: Methods of Information Geometry. Oxford University Press, Oxford (1993)
4. Murari, A., et al.: Nucl. Fusion **53,** 033006, 9 (2013)
5. Vega, J., et al.: Rev Sci. Instrum. **81**, 10E118 (2010)
6. Murari, A., et al.: Ann. Math. Artif. Intell. **74**(1), 155–180 (2015)
7. Murari, A., et al.: Nucl. Fusion **49,** 055028, 11 (2009)
8. Murari, A., Peluso, E., Gelfusa, M., Lungaroni, M., Gaudio, P.: How to handle error bars in symbolic regression for data mining in scientific applications. In: Gammerman, A., Vovk, V., Papadopoulos, H. (eds.) SLDS 2015. LNCS, vol. 9047, pp. 347–355. Springer, Heidelberg (2015). doi:10.1007/978-3-319-17091-6_29