Chapter 3: Conformal prediction

Volodymyr Vovk

v.vovk@rhul.ac.uk
Office Bedford 2-20

CS3920/CS5920 Machine Learning
Last edited: October 27, 2022

Version with solutions on slides 21–24, 26–27, and 31–32

Plan

1. Assumptions of machine learning
2. Conformal prediction
3. Conformal prediction based of Nearest Neighbour
4. Validity and efficiency of conformal predictors

History

- Statistics is a traditional data science, created mainly in England in the early 20th century. Based in maths. Main programming language: R.
- Machine learning: offshoot of computer science. Main programming language: Python.
- Different roots but a lot of connections nowadays.

Differences

- One difference is in the assumptions: in statistics, parametric assumptions (such as Gaussian) are ubiquitous, and in mainstream machine learning the assumption is IID.
- Machine learning is much more careful about computational efficiency.

Plan

1. Assumptions of machine learning
2. Conformal prediction
3. Conformal prediction based of Nearest Neighbour
4. Validity and efficiency of conformal predictors

Prediction with confidence

- We want guaranteed validity (such as guaranteed probability of error).
- Commonplace in statistics (confidence intervals, prediction intervals).
- Statisticians could do it because of their strong assumptions (such as Gaussianity).
- Relatively recent in machine learning (conformal prediction).
- New notation: $\mathbf{Y}$ is the set of all possible labels (label space).

Conformal prediction (1)

- Idea of conformal prediction: given a training set and a test sample, try in turn each potential label for the test sample.
- For each postulated label, we look at how plausible the extended (augmented) training set is (under the IID assumption).
- We can make a confident prediction if all but one completion look implausible.
- To evaluate the implausibility of the augmented training set we use the statistical notion of a p-value (to be defined).

Conformal prediction (2)

- The first step is to define a conformity measure.
- This is a function that maps any finite sequence of labelled samples

  $$z_1, \dots, z_m$$

  to the corresponding conformity scores

  $$\alpha_1, \dots, \alpha_m$$

  and is required to be equivariant: if we permute $z_1, \dots, z_m$, the corresponding $\alpha_1, \dots, \alpha_m$ will be permuted in the same way.

Conformal prediction (3)

- An equivalent way to express equivariance: α_i should be computable from z_i and the bag $\Lbag z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_m \Rbag$ (remember that a bag is a set with multiple copies of the same element allowed; also called a multiset).
- The intuition behind α_i: how well z_i conforms to the rest of the dataset.
- If α_i is small, we say that z_i is non-conforming, or strange.

Conformal prediction (4)

This is how conformal predictors work on a training set $z_1, \dots, z_n$ and test sample x^*:

- For each possible label $y \in \mathbf{Y}$ for x^*, compute the p-value

$$p(y) = \frac{\#\{i = 1, \dots, n+1 \mid \alpha_i^y \le \alpha_{n+1}^y\}}{n+1},$$

  where $\alpha_1^y, \dots, \alpha_n^y, \alpha_{n+1}^y$ are the conformity scores corresponding to $z_1, \dots, z_n, (x^*, y)$.

- If we are given a significance level $\epsilon > 0$ (our target probability of error), we can compute the corresponding prediction set

$$\Gamma^\epsilon = \{y \in \mathbf{Y} \mid p(y) > \epsilon\}.$$

Special cases

To understand the formula for $p(y)$ on the previous slide, consider special cases:

- If (x^*, y) is the strangest labelled sample among $z_1, \dots, z_n, (x^*, y)$ (which it might well be if y is a wrong label), then $p(y) = 1/(n+1)$.
  - $1/(n+1)$ is the smallest possible value for $p(y)$
- If (x^*, y) is the second strangest labelled sample among $z_1, \dots, z_n, (x^*, y)$, then $p(y) = 2/(n+1)$.
- If (x^*, y) is the most conforming labelled sample among $z_1, \dots, z_n, (x^*, y)$, then $p(y) = 1$.

Useful terminology

- I will sometimes refer to the numerator in

$$p(y) = \frac{\#\{i = 1, \ldots, n+1 \mid \alpha_i^y \le \alpha_{n+1}^y\}}{n+1}$$

  as the rank of α_{n+1}^y in the sequence (or bag) $\alpha_1^y, \ldots, \alpha_{n+1}^y$.
- So the p-value $p(y)$ is defined as the rank of α_{n+1}^y divided by $n+1$.
- Roughly, the rank of α_{n+1}^y is k if α_{n+1}^y is the kth smallest element in $\alpha_1^y, \ldots, \alpha_{n+1}^y$.

Validity and efficiency of conformal prediction

- Conformal predictors satisfy the following property of validity automatically: $y^* \notin \Gamma^\epsilon$ (the predictor makes a mistake) with probability at most ϵ (provided the labelled samples are IID).
- The property is easy to achieve (set $\Gamma^\epsilon = \mathbf{Y}$). We also want efficiency: in addition to validity, the prediction set should be small.

Plan

1. Assumptions of machine learning
2. Conformal prediction
3. Conformal prediction based of Nearest Neighbour
4. Validity and efficiency of conformal predictors

Suitable conformity measures for 1-Nearest Neighbour

- The distance to the nearest sample of a different class.
- One over the distance to the nearest sample of the same class:

$$\frac{1}{\text{the distance to the nearest sample of the same class}}.$$

- Or we can combine the two ideas: the distance to the nearest sample of a different class divided by the distance to the nearest sample of the same class:

$$\frac{\text{the distance to the nearest sample of a different class}}{\text{the distance to the nearest sample of the same class}}.$$

Example (1)

- Remember the training set in Chapter 2:
  - positive samples: $(0, 3)$, $(2, 2)$, $(3, 3)$
  - negative samples: $(-1, 1)$, $(-1, -1)$, $(0, 1)$.
- But now the test sample is $(0, 0)$, in the middle of the negative samples.
- What are the two p-values?
- As conformity measure, use the distance to the nearest sample of a different class divided by the distance to the nearest sample of the same class.

Example (2)

First assume the label of $(0,0)$ is $+1$.

Sample	Label	Conformity score
$(0,3)$	$+1$	$2/\sqrt{4+1} \approx 0.894$
$(2,2)$	$+1$	$\sqrt{4+1}/\sqrt{1+1} \approx 1.581$
$(3,3)$	$+1$	$\sqrt{9+4}/\sqrt{1+1} \approx 2.550$
$(-1,1)$	-1	$\sqrt{1+1}/1 \approx 1.414$
$(-1,-1)$	-1	$\sqrt{1+1}/2 \approx 0.707$
$(0,1)$	-1	$1/1 = 1$
$(0,0)$	$+1$ (?)	$1/\sqrt{4+4} \approx 0.354$

The test sample is the strangest, and the p-value is $1/7 \approx 0.143$.

Example (3)

Next assume the label of $(0,0)$ is -1.

Sample	Label	Conformity score
$(0,3)$	$+1$	$2/\sqrt{4+1} \approx 0.894$
$(2,2)$	$+1$	$\sqrt{4+1}/\sqrt{1+1} \approx 1.581$
$(3,3)$	$+1$	$\sqrt{9+4}/\sqrt{1+1} \approx 2.550$
$(-1,1)$	-1	$\sqrt{4+1}/1 \approx 2.236$
$(-1,-1)$	-1	$\sqrt{16+1}/\sqrt{2} \approx 2.915$
$(0,1)$	-1	$2/1 = 2$
$(0,0)$	-1 (?)	$\sqrt{4+4}/1 \approx 2.828$

The test sample is the second most conforming; its rank is 6, and the p-value is $6/7 \approx 0.857$.

Example (4)

- Notice: changing the postulated label changes plenty of conformity scores, not just one.
- The p-values are 0.143 (for $+1$) and 0.857 (for -1).
- We can predict -1, but our prediction does not achieve statistical significance (5%, as discussed below; for that, we would need at least 19 labelled samples in the training set).

Exercises for home

- Compute the two p-values taking as conformity score the distance to the nearest sample of a different class. Answer: 0.286 (for $+1$) and 0.714 (for -1).
- Compute the two p-values taking as conformity score one over the distance to the nearest sample of the same class. Answer: 0.143 (for $+1$) and 1 (for -1).
- What is your conclusion (if any) about the efficiency of the three conformity measures?

First exercise for $+1$

Sample	*Label*	*Conformity score* $=$ *Euclidean distance*
$(0,3)$	$+1$	2
$(2,2)$	$+1$	$\sqrt{5}$
$(3,3)$	$+1$	$\sqrt{13}$
$(-1,1)$	-1	$\sqrt{2}$
$(-1,-1)$	-1	$\sqrt{2}$
$(0,1)$	-1	1
$(0,0)$	$+1$ *(?)*	1

The test sample is one of the two strangest, and the p-value is $2/7 \approx 0.286$.

First exercise for -1

Sample	*Label*	*Conformity score* $=$ *Euclidean distance*
$(0,3)$	$+1$	2
$(2,2)$	$+1$	$\sqrt{5} \approx 2.236$
$(3,3)$	$+1$	$\sqrt{13} \approx 3.606$
$(-1,1)$	-1	$\sqrt{5} \approx 2.236$
$(-1,-1)$	-1	$\sqrt{17} \approx 4.123$
$(0,1)$	-1	2
$(0,0)$	-1 *(?)*	$\sqrt{8} \approx 2.828$

The test sample is the fifth strangest, and the p-value is $5/7 \approx 0.714$.

Second exercise for $+1$

For simplicity, I will use Euclidean distance as nonconformity measure.

Sample	*Label*	*Nonconformity score*
$(0,3)$	$+1$	$\sqrt{5}$
$(2,2)$	$+1$	$\sqrt{2}$
$(3,3)$	$+1$	$\sqrt{2}$
$(-1,1)$	-1	1
$(-1,-1)$	-1	2
$(0,1)$	-1	1
$(0,0)$	$+1$ *(?)*	$\sqrt{8}$

The test sample is the strangest, and the p-value is $1/7 \approx 0.143$.

Second exercise for -1

I will again use nonconformity scores.

Sample	*Label*	*Nonconformity score*
$(0,3)$	$+1$	$\sqrt{5}$
$(2,2)$	$+1$	$\sqrt{2}$
$(3,3)$	$+1$	$\sqrt{2}$
$(-1,1)$	-1	1
$(-1,-1)$	-1	$\sqrt{2}$
$(0,1)$	-1	1
$(0,0)$	-1 *(?)*	1

The test sample is one of the three least strange, and the p-value is 1.

Exercise for now

- The training set has only one feature:
    - positive samples: 0 and 1
    - negative samples: 10 and 11.
- The test sample is 12 (which seems to be in the negative area).
- What are the two p-values?
- As conformity measure, take the distance to the nearest sample of a different class.

Solution (1)

First assume the label of 12 *is* $+1$.

Sample	*Label*	*Conformity score*
0	$+1$	10
1	$+1$	9
10	-1	2
11	-1	1
12	$+1$ *(?)*	1

The test sample is one of the two strangest, and the p-value is $2/5 = 0.4$.

Solution (2)

Next assume the label of 12 *is* -1*.*

Sample	*Label*	*Conformity score*
0	$+1$	10
1	$+1$	9
10	-1	9
11	-1	10
12	-1 *(?)*	11

The test sample is the least strange, and the p-value is $5/5 = 1$*.*

Nonconformity measures (1)

- Formally, nonconformity measures are defined in the same way as conformity measures.
- But their interpretation is different: α_i measures how strange (rather than how conforming) z_i is.
- The formula for computing p-values on slide 10, becomes

$$p(y) = \frac{\#\{i = 1, \ldots, n+1 \mid \alpha_i^y \geq \alpha_{n+1}^y\}}{n+1}$$

(the only difference is that $\leq$ becomes $\geq$).

Nonconformity measures (2)

- In principle, it does not matter whether you use nonconformity measures (the Royal Holloway convention) or conformity measures (the Carnegie Mellon convention).
- Instead of using nonconformity scores α_i, you can instead use conformity scores $-\alpha_i$ (or $1/\alpha_i$ if α_i are positive).
- But in application to regression nonconformity scores are often more convenient.

Exercise for now

- Consider the same training set as before (slide 25):
  - positive samples: 0 and 1
  - negative samples: 10 and 11.
- The test sample is still 12.
- What are the two p-values?
- But now we use a nonconformity measure, namely: the distance to the nearest sample of the same class.

Solution (1)

First assume the label of 12 *is* $+1$.

Sample	*Label*	*Nonconformity score*
0	$+1$	1
1	$+1$	1
10	-1	1
11	-1	1
12	$+1$ *(?)*	11

The test sample is the strangest one, and so the p-value is $1/5 = 0.2$.

Solution (2)

Next assume the label of 12 *is* -1.

Sample	*Label*	*Nonconformity score*
0	$+1$	1
1	$+1$	1
10	-1	1
11	-1	1
12	-1 *(?)*	1

The test sample is one of the least strange, and the p-value is $5/5 = 1$.

Two families of nonconformity measures

- Perhaps the most popular class of nonconformity measures in the case of regression is

  $$\alpha_i = \left|y_i - \hat{y}_i\right|,$$

  where $\hat{y}_i$ is a prediction for y_i. Advantage: mathematical simplicity; facilitates efficient computations.
- Another popular class of nonconformity measures in the case of regression is

  $$\alpha_i = \left|y_i - \hat{y}_i\right| / \sigma_i,$$

  where $\hat{y}_i$ is a prediction for y_i and $\sigma_i > 0$ is an estimate of its accuracy. Advantage: the size of the prediction set is more adaptive.

A nonconformity measure based on Nearest Neighbour

In this chapter we consider an element of the first family: the nonconformity scores

$$\alpha_1, \ldots, \alpha_m$$

of labelled samples

$$(x_1, y_1), \ldots, (x_m, y_m)$$

are defined by $\alpha_i = |y_i - \hat{y}_i|$, where $\hat{y}_i$ is the label of the nearest neighbour of x_i among $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m$.

Difficulty

- The main difficulty for regression problems is that there are infinitely many potential labels to consider.
- One possible solution is to consider a dense finite grid of potential labels, and then for each possible label in the grid compute its p-value.
- Occasionally another solution is possible: we can derive a formula (or a very efficient algorithm) for the prediction set. This is the case for K Nearest Neighbours, Least Squares regression, Ridge Regression, and Lasso (the last three algorithms will be discussed in later chapters).

Example of querying specific labels

- Consider the training set

$$
\begin{aligned}
(x_1, y_1) &= (2, 0), & (x_2, y_2) &= (1.2, 2), \\
(x_3, y_3) &= (1, 1), & (x_4, y_4) &= (0, 2)
\end{aligned}
$$

  consisting of four labelled samples.
- Find the p-values for the test labelled samples $(x, y) = (4, 0)$ and $(x, y) = (2.5, 1)$.

Solution for $(x, y) = (4, 0)$

Sample	Label	Label of the NN	NS
2	0	2	$\lvert 0-2\rvert = 2$
1.2	2	1	$\lvert 2-1\rvert = 1$
1	1	2	$\lvert 1-2\rvert = 1$
0	2	1	$\lvert 2-1\rvert = 1$
4	0 (?)	0	$\lvert 0-0\rvert = 0$

Here NN stands for "Nearest Neighbour" (in the augmented training set) and NS stands for "nonconformity score". The test sample is the least strange, and the p-value is 1.

Solution for $(x, y) = (2.5, 1)$

Sample	Label	Label of the NN	NS
2	0	1 (?)	$\lvert 0 - 1 \rvert = 1$
1.2	2	1	$\lvert 2 - 1 \rvert = 1$
1	1	2	$\lvert 1 - 2 \rvert = 1$
0	2	1	$\lvert 2 - 1 \rvert = 1$
2.5	1 (?)	0	$\lvert 1 - 0 \rvert = 1$

- The test sample is one of the least strange (and also one of the most strange), and the p-value is still 1.
- The question mark now indicates the dependence on the postulated label.

Using a grid (1)

- In regression problems, Γ^ϵ is typically an interval, $\Gamma^\epsilon = [a, b]$.
- A crude way to compute it is to choose a large interval $[A, B]$ (containing all interesting values for the label y) and choose a dense grid in it: $\{A, A + \texttt{step}, A + 2\,\texttt{step}, \ldots, B\}$ for a small $\texttt{step} > 0$.
- Go over all y in the grid and output the range $[a, b]$ of y for which $p(y) > \epsilon$.

Using a grid (2)

This is a possible snippet of Python code, where

- p is a function computing $p(y)$ given y,
- arange(A,B,step) is (in NumPy) $\{A, A+\texttt{step}, A+2\,\texttt{step}, \dots, B\}$ (not including B),
- epsilon is ϵ,
- and NaN is an undefined value (“not a number”).

```
a = NaN
b = NaN
for y in arange(A,B,step):
  if p(y) > epsilon:
    b = y
    if a == NaN:
      a = y
```

A shortcut

- Consider the same training set

$$\begin{aligned} (x_1, y_1) &= (2, 0), & (x_2, y_2) &= (1.2, 2), \\ (x_3, y_3) &= (1, 1), & (x_4, y_4) &= (0, 2) \end{aligned}$$

  consisting of four labelled samples.
- The test sample is $x^* = 4$. What is the prediction set at the significance level 20%?

Solution for $x^* = 4$

Sample	Label	Label of the NN	NS
2	0	2	$\lvert 0-2\rvert = 2$
1.2	2	1	$\lvert 2-1\rvert = 1$
1	1	2	$\lvert 1-2\rvert = 1$
0	2	1	$\lvert 2-1\rvert = 1$
4	y	0	$\lvert y-0\rvert = \lvert y\rvert$

The p-value is 20% (or less) if the test labelled sample is the strangest. In other words: if $|y| > 2$. The prediction set at 20%:

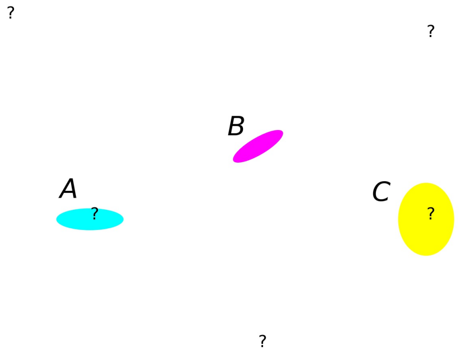
$$\Gamma^{20\%} = [-2, 2].$$

It contains 0, as we already know (see slide 37).

Conformal prediction for anomaly detection

- It is possible that no postulated label for the test sample is plausible (all p-values are small). What does it mean?
- Consider three families of computer viruses as shown on the next slide (three compact clouds) and think what happens if the new virus is far from any of the clouds.
- The slide after that: a numeric illustration.
- For simplicity, take the distance to the same class as nonconformity measure.

Three families and several test samples



Exercise (1)

- Consider the following training set with three classes (such as virus families) A, B, and C:
  - 0 and 1 are labelled A;
  - 5 and 6 are labelled B;
  - 10 and 11 are labelled C.
- Using the distance to the same class as nonconformity measure, compute the p-values for 0.5, 5.5, and 3.
  - Answer for 0.5: $p_A = 1$, $p_B = 1/7$, $p_C = 1/7$.

Exercise (2)

- What should be our conclusions?
- In the case of the test samples 0.5 and 5.5 we can make a confident prediction: the families are A and B, respectively.
- For the test sample 3, all p-values are low. It looks as if we have a new family (or, otherwise, the test sample is an unusual representative of an old family).

Confidence and credibility (1)

- Let p_A, p_B, and p_C be the three p-values for a test sample.
- We can make a confident prediction if all p-values apart from one are very small.
- We can summarize our prediction as follows:
  - the point prediction is the label with the largest p-value (A for the test sample 0.5 on slide 45);
  - our confidence is one minus the second largest p-value (6/7 for the test sample 0.5);
  - the credibility is the largest p-value (1 for the test sample 0.5).

Confidence and credibility (2)

- Therefore, we can make a confident prediction if the confidence is high (close to 1) and the credibility is not low.
- If the credibility is very low: are we witnessing a new class?
- But it is not a good idea to measure the performance of your conformal predictor by, say, the average confidence on the test set. We will see a much better way in the next section.

Plan

1. Assumptions of machine learning
2. Conformal prediction
3. Conformal prediction based of Nearest Neighbour
4. Validity and efficiency of conformal predictors

Proof of validity for $\epsilon = 1/(n+1)$

- What is the probability that $y^* \notin \Gamma^{1/(n+1)}$? (Cf. slide 13.)
- Notice that $y^* \notin \Gamma^{1/(n+1)}$ means that $z^* = (x^*, y^*)$ is the strangest labelled sample in the set $z_1, \ldots, z_n, z^*$.
- By the IID assumption, all permutations of $z_1, \ldots, z_n, z^*$ (and the corresponding permutations of $\alpha_1, \ldots, \alpha_n, \alpha_{n+1}$) have the same probability.
- So the probability that the smallest conformity score in the bag $\wr \alpha_1, \ldots, \alpha_{n+1} \int$ will be the last one is exactly $1/(n+1)$ (provided there is only one smallest element in the bag; otherwise the probability of error will be less, 0).

Exercises for home

- Exercise 1: Show that the probability of $y^* \notin \Gamma^{2/(n+1)}$ does not exceed $2/(n+1)$. *For an answer, see the version of* `q03.tex` *with solutions.*
- Exercise 2 (optional): Show that the probability of $y^* \notin \Gamma^{k/(n+1)}$ does not exceed $k/(n+1)$, for any $k \in \{1, 2, \dots, n\}$.

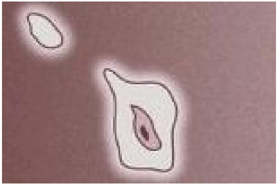
Statistical significance

- In statistics, p-values are used for testing statistical hypotheses.
- If we obtain a p-value $\leq 5\%$, the result is statistically significant.
- If we obtain a p-value $\leq 1\%$, the result is highly statistically significant.
- In conformal prediction, we are testing the IID assumption.
- The standard statistical conventions call for paying particular attention to $\Gamma^{5\%}$ and $\Gamma^{1\%}$.

Nested prediction sets

But it's best to look at what happens at more than two significance levels. For example: we can call

- $\Gamma^{20\%}$ casual prediction,
- $\Gamma^{5\%}$ confident prediction,
- $\Gamma^{1\%}$ highly confident prediction.



Randomized p-values

- In theoretical (never, or almost never, experimental) work people usually use randomized p-values

  $$p(y,\tau) = \frac{\#\{i : \alpha_i^y < \alpha_{n+1}^y\} + \tau\#\{i : \alpha_i^y = \alpha_{n+1}^y\}}{n+1},$$

  where $i = 1, \ldots, n+1$ and $\tau \in [0,1]$ is chosen independently from the uniform distribution on $[0,1]$.
- Notice: $p(y,\tau) \leq p(y,1) = p(y)$; now $p(y,\tau) < 1/(n+1)$ is possible.
- Using randomized p-values can only make our prediction sets smaller (and so validity will be more difficult to achieve).

Prediction in the online mode

Let P be a probability measure on $\mathbf{Z}$ (the labelled samples) and U be the uniform probability measure on $[0, 1]$.

Protocol

```
ONLINE MODE OF PREDICTION
    generate a labelled sample $z_1 = (x_1, y_1) \sim P$
    **for** $n = 1, 2, \ldots$ **do**
        generate a new labelled sample $z_{n+1} = (x_{n+1}, y_{n+1}) \sim P$
            independently
        generate a new random number $\tau_n \sim U$ independently
        compute $p(y, \tau_n)$ for each potential label $y$ for $x_{n+1}$ as test
            sample from $z_1, \ldots, z_n$ as training set
        set $p_n = p(y_{n+1}, \tau_n)$
    **end for**
```

The property of validity

Theorem

In the online mode of prediction, the consecutive p-values $p_1, p_2, \dots$ are independent and distributed uniformly on $[0, 1]$.

- In particular, conformal predictors make errors at different steps independently with probability ϵ.
- Optional remark 1: the independence allows us to use the law of large numbers (so that the percentage of errors over the first n steps is close to ϵ with high probability when n is large).
- Optional remark 2: the proof of the theorem uses a backward argument.

Average false p-value

How do we measure the efficiency of conformal predictors?

- In the case of regression, we could look at the area of the nested prediction sets such as those given on slide 53 (details omitted).
- In this subsection we only discuss the case of classification.
- The average false p-value: the average of the p-values for all postulated labels in the test set except for the true labels.

Exercise

- Suppose we have obtained the following p-values when applying a conformal predictor to a test set of size 4 and with the label space $\mathbf{Y} = \{A, B, C\}$:

True label	p_A	p_B	p_C
B	0.05	0.3	0.05
A	0.7	0.02	0.08
B	0.04	0.4	0.06
C	0.01	0.09	1

- Find the average false p-value. Answer: 0.05.
- This is the way you are asked to measure the efficiency of your conformal predictor in Assignment 1.

Python packages

Conformal Prediction add-on in Orange 3 (Version 1.1.3, May 2019)
`https://pypi.org/project/Orange3-Conformal/`

Henrik Linusson, `nonconformist` (Version 2.1.0, June 2017)
`https://pypi.org/project/nonconformist/`

In Assignment 1, you can use them only as a source of ideas (but it's easier not to use them at all).

Further information

- V05: Chapter 2.
  See also the events and working papers at the book's web site, `http://alrw.net`.
- B14: Chapter 1.
  Starting from Chapter 3: reviews of applications of conformal prediction by area.
- Wikipedia article "Ranking",
  `https://en.wikipedia.org/wiki/Ranking`.
  Conformal prediction uses "modified competition ranking" (or "1334" ranking).

Further information: research literature

- Wenyu Chen, Kelli-Jean Chun, & Rina Foygel Barber (2018).
  Discretized conformal prediction for efficient distribution-free inference.
  Stat 7:e173.
  How to do conformal regression using a grid rigorously.

- Leying Guan and Robert Tibshirani (2022).
  Prediction and outlier detection in classification problems.
  Journal of the Royal Statistical Society B 84:524–546.
  Also available as arXiv report.
  Combining classification and anomaly-detection capabilities of conformal prediction.