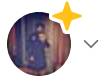Open in app ↗

Search Medium   Write   🔔

Sklearn Probability Calibration (in two schemes)

Alex Kataev · Following
Sep 9, 2021

👏 2   💬   🔖   ▶   ⬆   •••

> *Inspired by: https://scikit-learn.org/stable/modules/calibration.html*

I'll show you two schemes to make the probability calibration with scikit-learn illustrative and clear.
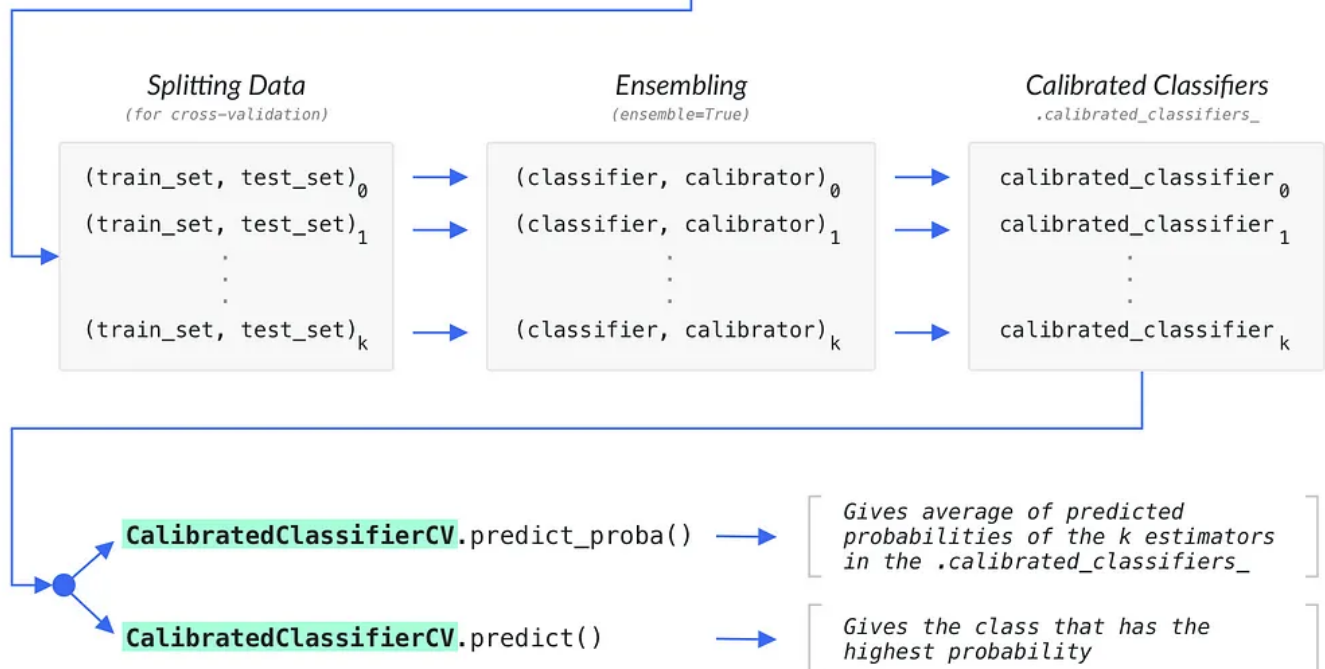
The first one is an overall picture of CalibratedClassiffierCV
(`sklearn.calibration.CalibratedClassiffierCV`):

Inside `CalibratedClassifierCV`

```
CalibratedClassifierCV(base_estimator=classifier,
                       method='sigmoid',
                       cv=k,
                       ensemble=True)
```
                                              GaussianNB(…)
                                              RandomForestClassifier(…)
                                              LinearSVC(…)
                                              etc.

```
CalibratedClassifierCV.fit(X_train, y_train)
```

Splitting Data
(for cross-validation)

$(train\_set, test\_set)_0$
$(train\_set, test\_set)_1$
.
.
.
$(train\_set, test\_set)_k$

Ensembling
(ensemble=True)

$(classifier, calibrator)_0$
$(classifier, calibrator)_1$
.
.
.
$(classifier, calibrator)_k$

Calibrated Classifiers
`.calibrated_classifiers_`

$calibrated\_classifier_0$
$calibrated\_classifier_1$
.
.
.
$calibrated\_classifier_k$

`CalibratedClassifierCV.predict_proba()` →
    *Gives average of predicted probabilities of the k estimators in the .calibrated_classifiers_*

`CalibratedClassifierCV.predict()` →
    Gives the class that has the highest probability
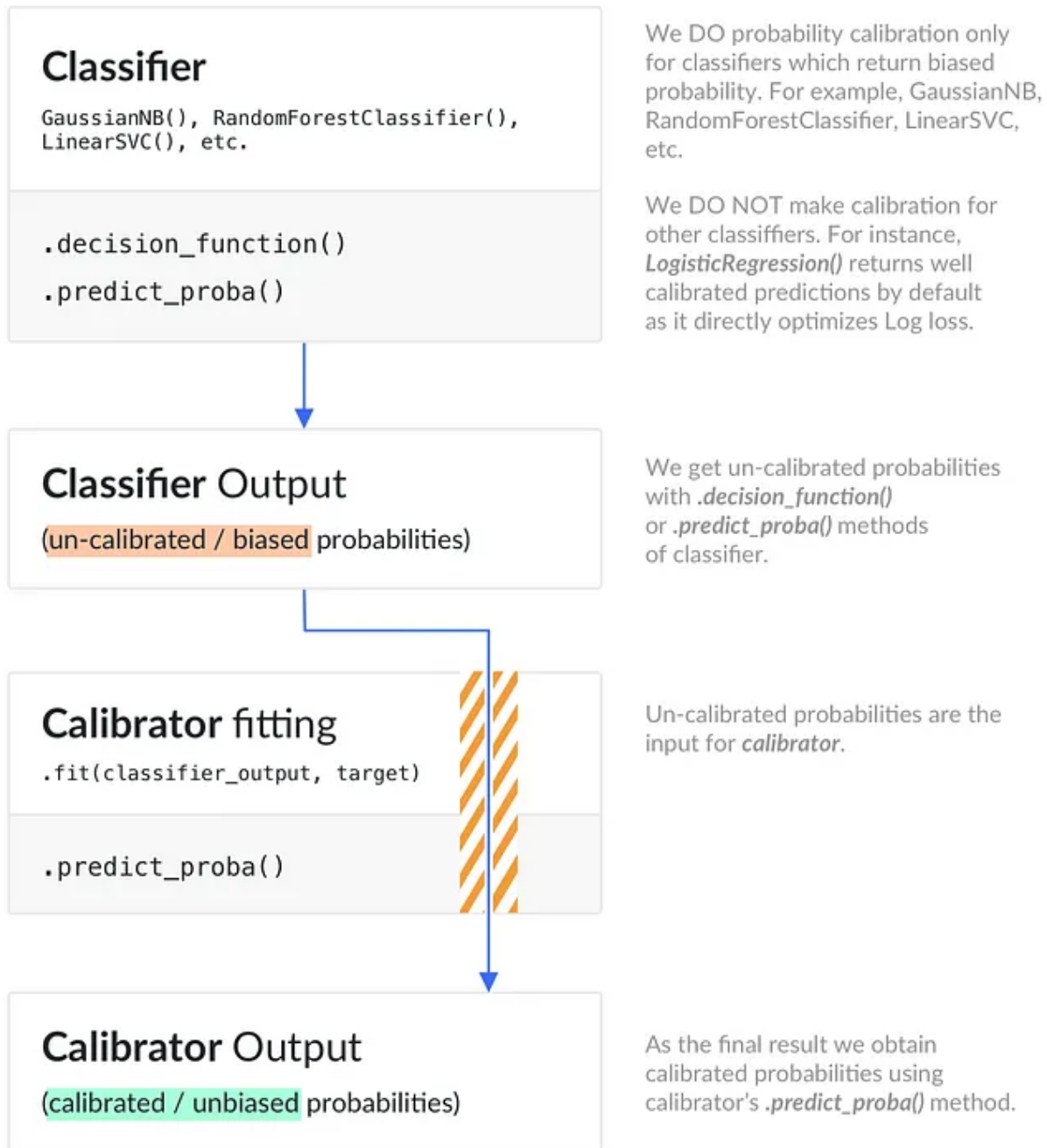
© Alex Kataev, https://medium.com/@alexkataev

Note: In case of `ensemble=False`, unbiased predictions are obtained via `cross_val_predict`.

And a little more detail on how ensemble of classifier and calibrator works:

Inside `(classifier, calibrator)`

Classifier

GaussianNB(), RandomForestClassifier(),
LinearSVC(), etc.

.decision_function()

.predict_proba()

We DO probability calibration only for classifiers which return biased probability. For example, GaussianNB, RandomForestClassifier, LinearSVC, etc.

We DO NOT make calibration for other classiffiers. For instance, *LogisticRegression()* returns well calibrated predictions by default as it directly optimizes Log loss.

Classifier Output

(un-calibrated / biased probabilities)

We get un-calibrated probabilities with *.decision_function()* or *.predict_proba()* methods of classifier.

Calibrator fitting

.fit(classifier_output, target)

.predict_proba()

Un-calibrated probabilities are the input for *calibrator*.

Calibrator Output

(calibrated / unbiased probabilities)

As the final result we obtain calibrated probabilities using calibrator's *.predict_proba()* method.

That's it. I also recommend you to take a look at official guide on probability calibration: https://scikit-learn.org/stable/modules/calibration.html — it's

really worth it.

Hope that was helpful. Press `thumbs up` 👍 if it was.

Thank you for your time! 🎉

Follow me for more interesting topics 😉👍

Scikit Learn   Sklearn   Probability Calibration

More from the list: "ML"

Curated by Himanshu Birla

👤 Kyosuke… in Towards Dat…	👤 Mattia Ci… in Analytics Vi…	👤 Jason Yo…
Probability Calibration for Imbalanced Dataset	**How Probability Calibration Works**	**Why Calibrat the Series on**
✨ · 8 min read · Oct 20, 2019	✨ · 6 min read · May 28, 2020	7 min read · Oc

View list

Written by Alex Kataev

Following

Software Engineering & Data Science Enthusiast

More from Alex Kataev

Alex Kataev in Geek Culture

SVM Classification with sklearn.svm.SVC: How To Plot A...

So, if you're tired of digging up tons of useless links, wishing to find a detailed explanation o...

6 min read · Aug 19, 2021

92   2

Alex Kataev

📈 UCI Bank Marketing Dataset. Part 1. Exploratory Data Analysis...

21 min read · Dec 20, 2021

7

Alex Kataev

Alex Kataev

🐍 Magic Python: Mutable vs Immutable and How To Copy…

How a Classification Tree Algorithm Splits Feature By Groups

What is the difference between immutable and mutable objects and what is the differen…

Based on an assignment #3 by ODS (Open Data Science Community)…

9 min read · Mar 20, 2020

2 min read · Nov 27, 2020

👏 3

👏 1

See all from Alex Kataev

Recommended from Medium
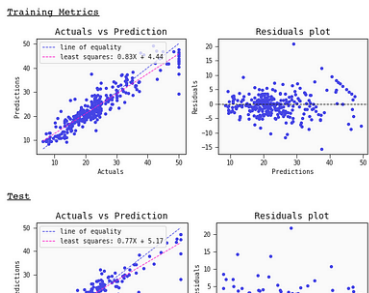
Casper Skern Wilstrup

Symbolic Regression: a Simple and Friendly Introduction

Symbolic Regression is like a treasure hunt for the perfect mathematical equation to...

3 min read · May 5

👏 18   💬 1   🔖   •••

Preethi Prakash

Mastering Machine Learning Predictions in Scikit-Learn

Understanding the predict, predict_proba, predict_log_proba, and decision_function...
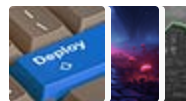
3 min read · May 13

👏 2   💬   🔖   •••

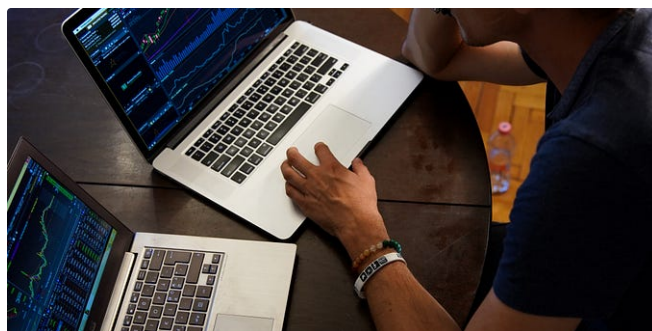## Lists

Practical Guides to Machine Learning
10 stories · 544 saves

Predictive Modeling w/ Python
20 stories · 473 saves

Natural Language Processing
689 stories · 304 saves

Abel K Widodo

Exploring Heuristics Analytics with Python

In the era of big data and complex problem-solving, businesses and researchers are...

6 min read · Jul 28

Saurav Agrawal

Hyperparameter Tuning of KNN Classifier

K-Nearest Neighbor Classifier is a machine learning algorithm used for classification an...
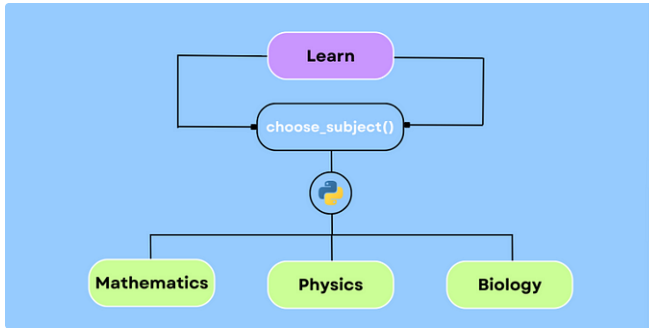
3 min read · Jun 5

Isha Choudhary

Übermensch

All About "Class Method" in Python
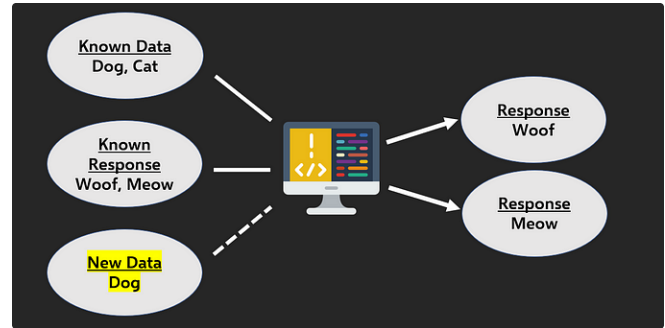
OOPs in Python — Extension

Linear Regression and Logistic Regression - What is different?

What is a regression model in Machine Learning?

5 min read · Sep 15

4 min read · May 22

See more recommendations