

Для чего и в каких случаях полезны различные варианты усреднения для метрик качества классификации: **micro, macro, weighted**?

В случае когда требуется оценить вариант предсказания с несколькими классами применяют F_score с параметрами micro, macro, weighted.

https://scikit-learn.org/stable/modules/generated/sklearn.metrics.precision_recall_fscore_support.html

Так функция `sklearn.metrics.precision_recall_fscore_support`, вычисляет значения precision ($tp / (tp + fp)$), recall ($tp / (tp + fn)$), F-measure для каждого класса.

Параметр

pos_labelstr or int, 1 by default определяет тип классификации бинарная (1 класс по умолчанию) или многоклассовая N

Если `average='binary'` предсказание выполняется для бинарных данных. Если классов несколько данный параметр игнорируется. Задавая `labels=[pos_label]` и `average != 'binary'` получим метрику только для заданного класса.

average = string, [None (default), 'binary', 'micro', 'macro', 'samples', 'weighted']

Если `None` вернет метрики для каждого класса, иначе выполнит определенный тип осреднения над данными:

'binary': Выводит только результат для заданного класса `pos_label`. Используется для оценки бинарных данных - if targets (`y_{true,pred}`) are binary.

'micro':

Вычисляет метрики для всех классов глобально, подсчитывает количества TP, FN, FP

Calculate metrics globally by counting the total true positives, false negatives and false positives.

'macro':

Вычисляет метрики для каждого класса и находит их НЕвзвешенное среднее, не учитывая разбаланс классов.

Calculate metrics for each label, and find their unweighted mean. This does not take label imbalance into account.

'weighted':

Вычисляет метрики для каждого класса и находит их взвешенное среднее, учитывая количество TP для каждого класса. Альтернатива 'macro' если есть разбаланс классов, можно получить F-score за рамками precision и recall.

Calculate metrics for each label, and find their average weighted by support (the number of true instances for each label). This alters 'macro' to account for label imbalance; it can result in an F-score that is not between precision and recall.