

Бэггинг (Bootstrap aggregating)

Разделы: [Алгоритмы](#)

В теории машинного обучения — метод построения ансамбля моделей, в котором обучение базовых моделей производится параллельно. При этом каждая модель обучается на отдельной выборке, сформированной из исходного набора данных с помощью алгоритма бутстрапа. Выход ансамбля определяется путем усреднения выходов базовых моделей.

Метод позволяет улучшить точность и устойчивость работы алгоритмов машинного обучения, уменьшить дисперсию ошибки и уменьшить эффект переобучения. Хотя изначально метод был разработан для классификаторов на основе деревьев решений, он может использоваться для любых видов моделей.

Из исходного набора данных D , содержащего n примеров, методом равномерного случайного сэмплинга с возвратом формируется m выборок из k примеров. Поскольку используется сэмплинг с возвратом, некоторые примеры могут попасть в выборки по несколько раз. Затем на полученных выборках обучаются базовые модели с использованием одинаковых алгоритмов, и сделанные ими предсказания усредняются.

Среди преимуществ бэггинга можно отметить значительное увеличение точности предсказания ансамбля относительно базовых классификаторов (порядка 10-40%). Оно достигается за счет уменьшения разброса предсказаний базовых моделей при усреднении.

К недостаткам метода можно отнести слабую математическую обоснованность улучшения точности предсказаний, недетерминированность результата из-за случайного формирования выборок и относительно сложность интерпретации результатов.

Метод был предложен Лео Брейманом в 1994 году для улучшения точности классификаторов на основе деревьев решений.