

1. Что такое Стэкинг?

Стекинг (Stacked Generalization или Stacking) — один из самых популярных способов ансамблирования алгоритмов, т.е. использования нескольких алгоритмов для решения одной задачи машинного обучения. Пожалуй, он замечателен уже тем, что постоянно переизобретается новыми любителями анализа данных. Это вполне естественно, его идея лежит на поверхности. Известно, что если обучить несколько разных алгоритмов, то в задаче регрессии их среднее, а в задаче классификации — голосование по большинству, часто превосходят по качеству все эти алгоритмы. Возникает вопрос: почему, собственно, использовать для ансамблирования такие простые операции как усреднение или голосование? Можно же ансамблирование доверить очередному алгоритму (т.н. «метаалгоритму») машинного обучения.

2. Что такое блэндинг?

Простейшая схема стекинга — блэндинг (Blending): обучающую выборку делят на две части. На первой обучают базовые алгоритмы. Затем получают их ответы на второй части и на тестовой выборке. Понятно, что ответ каждого алгоритма можно рассматривать как новый признак (т.н. «метапризнак»). На метапризнаках второй части обучения настраивают метаалгоритм. Затем запускают его на метапризнаках теста и получают ответ.