

В каких случаях при ансамблировании лучше использовать моду, а не среднее арифметическое?

Если рассмотреть задачу классификации с двумя классами $\{0, 1\}$ и три модели, каждая из которых ошибается с вероятностью p , то в предположении, что их ответы – независимые случайные величины, получаем, что комитет большинства этих трёх алгоритмов ошибается с вероятностью $p^2(3-2p)$. Как видно на рис. 1, это выражение может быть существенно меньше p (при $p=0.1$ почти в два раза), т.е. использование такого ансамбля уменьшает ошибку базовых алгоритмов.

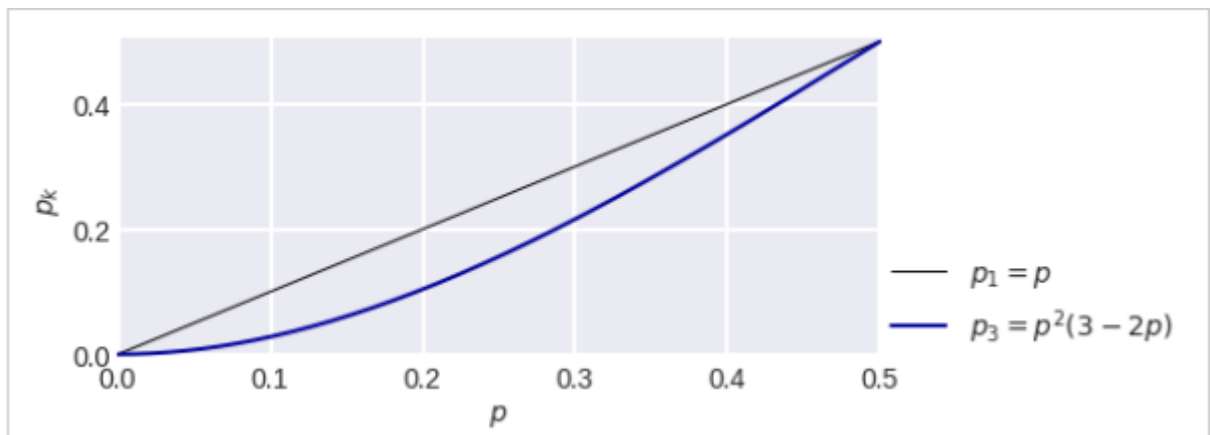


Рисунок 1 – График вероятности ошибки

Для чего используется EarlyStopping?

EarlyStopping – это callback, который останавливает обучение, когда отслеживаемый показатель перестает улучшаться. С помощью него, например, можно избежать переобучение сети.

Для чего может добавляться шум к весам в ходе обучения?

Добавление шума к весам используется для регуляризации моделей (модификация алгоритма обучения, предпринятая с целью уменьшить его ошибку обобщения, не уменьшая ошибку обучения). В основном применяется

в рекуррентных нейронных сетях. Применение шума к весам можно также интерпретировать как эквивалент (при определенных допущениях) более традиционной формы регуляризации, поощряющей устойчивость обучаемой функции.

Почему выбран именно такой набор моделей для ансамбля?

Для получения надежного ансамбля сетей стоит использовать хорошие и разные модели. Для работы с текстом обычно используют рекуррентные нейронные сети, поэтому одна из моделей была рекуррентная. Чтобы использовать разные архитектуры (использование одинаковых может привести к незначительному улучшению), поэтому была взята рекуррентная сверточная сеть.