

Оценка параметров регуляризации эластичных сетей с помощью максимизации обоснованности

И.А. Дубнов^{1,2}, А.Б. Мерков¹¹Институт системного анализа РАН (ФИЦ ИУ РАН - ИСА РАН)²Московский физико-технический институт (государственный университет)

В работе [1] предложена эластичная сеть – параметризованная вероятностная модель для классификации и регрессии с регуляризацией, являющейся линейной комбинацией L_1 -нормы и L_2 -нормы вектора параметров:

$$J(X, Y, w) + \lambda |w| + \frac{\mu}{2} \|w\|_2^2 \rightarrow \min_w. \quad (1).$$

Обучение с такой регуляризацией позволяет получать разреженную модель, т.е. независимую от многих признаков, с большей точностью классификации, чем при обучении только с L_1 -регуляризацией. Варьируя параметры регуляризации λ и μ , можно балансировать между разреженностью и качеством обученной модели.

В данной работе предлагается итеративный метод оценки параметров регуляризации эластичных сетей с помощью максимизации обоснованности. Применение максимизации обоснованности широко известно для оценки одного параметра регуляризации, например в случае гребневой регрессии ([2]).

Предлагается итеративное преобразование параметров регуляризации

$$\lambda_{k+1} \leftarrow F(\lambda_k, \mu_k, w), \mu_{k+1} \leftarrow G(\lambda_k, \mu_k, w), \quad (2)$$

зависящее от значений параметров λ и μ на текущем шаге и от вектора параметров модели w – решения минимизационной задачи (1) при текущих λ и μ . Таким образом, двухуровневый процесс обучения эластичной сети состоит из повторения следующих шагов: минимизация (1) при фиксированных параметрах регуляризации, и итеративное изменение параметров λ и μ преобразованием (2).

Описанный метод протестирован на примере задачи распознавания рукописных символов из базы MNIST ([3]). В первой серии экспериментов значения параметров регуляризации устанавливались из заданной сетки значений. Полученные результаты отражают возможность эластичной сети получать разреженную модель при больших значениях λ , а также улучшение точности классификации при удачном подборе μ .

Во второй серии экспериментов эластичная сеть обучается с использованием предлагаемого метода оценки параметров. Эксперименты показывают, что итеративное преобразование зависит от начальных значений параметров λ_0 и μ_0 . Это закономерный результат, т.к. задача максимизации обоснованности для эластичной сети не является

выпуклой по λ и μ , а итеративное преобразование (2) является аналогом градиентного спуска, достигающего в некоторых случаях локального максимума.

Сравнение двух серий экспериментов показывает, что метод максимизации обоснованности для эластичной сети позволяет получить значения параметров регуляризации за меньшее количество итераций, чем требуется для поиска по сетке значений. Более того, начиная итерации с разных углов сетки значений, можно получить модели с разным соотношением разреженности и точности классификации.

Предложенный метод является альтернативой алгоритмам поиска по сетке значений при подборе параметров регуляризации.

Литература

1. *Hui Zou and Trevor Hastie*. Regularization and Variable Selection via the Elastic Net // *Journal of the Royal Statistical Society B*. – 2005. – N 67. – P. 301–320.
2. *Bishop M.C.* Pattern Recognition and Machine Learning. – М.: Springer, 2006 – 740 p.
3. *Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner*. Gradient-based learning applied to document recognition // *Proceedings of the IEEE* – 1998. – V 11, N 86. – P. 2278-2324.