

Методы регуляризации для сокращения переобучения нейронных сетей

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Дубров И.В.

Телеш И.А. – канд. геогр. наук, доцент

Увеличение количества данных обучения является одним из способов сокращения переобучения. Есть ли другие способы уменьшить степень переобучения? Один из возможных подходов – уменьшить размер нейронной сети. Однако большие сети обладают потенциалом быть более мощными, чем небольшие сети, и поэтому этот вариант можно отбросить [1].

К счастью, существуют и другие методы, которые могут сократить переобучение, даже когда есть фиксированная сеть и фиксированные данные обучения. Они известны как методы регуляризации. Следует рассмотреть один из наиболее часто используемых методов регуляризации, метод регуляризации $L2$. Идея регуляризации $L2$ заключается в добавлении дополнительного параметра к функции стоимости, называемого параметром регуляризации. Вот регуляризованная перекрестная энтропия:

$$C = -\frac{1}{n} \sum_{x_j} [y_j \ln a_j^I + (1 - y_j) \ln(1 - a_j^I)] + \frac{\lambda}{2n} \sum_w w^2 \quad (1)$$

где $\lambda > 0$ – параметр регуляризации;

n – размер обучающего множества.

Первый параметр – это просто обычное выражение для перекрестной энтропии. Но также добавлен второй параметр, а именно сумма квадратов всех весов в сети. Это масштабируется с коэффициентом $\lambda/2n$.

Конечно, можно регуляризовать и другие функции стоимости, такие как квадратичная стоимость. Это можно сделать аналогичным образом, но в обоих случаях можно написать регуляризованную функцию стоимости как:

$$C = C_0 + \frac{\lambda}{2n} \sum_w w^2 \quad (2)$$

где C_0 – исходная нерегуляризованная функция стоимости.

Интуитивно, эффект регуляризации состоит в том, чтобы сделать так, чтобы сеть предпочитала учиться малым весам, при прочих равных условиях. Большие веса допустимы только в том случае, если они значительно улучшают первую часть функции стоимости. Другими словами, регуляризация может рассматриваться как способ компромисса между нахождением небольших весов и минимизацией первоначальной функции стоимости. Относительная важность двух элементов компромисса зависит от значения λ : когда λ мало, предпочитаем минимизировать оригинальную функцию стоимости, но когда λ велико, предпочитаем малые веса [2].

Если функция стоимости нерегулярна, то длина вектора веса, вероятно, будет расти, при прочих равных условиях. Со временем это может привести к тому, что весовой вектор будет очень большим. Это может привести к тому, что веса застрянут, указывая более или менее в одном и том же направлении, так как изменения, вызванные градиентным спуском, производят лишь малые изменения. Это явление затрудняет алгоритму правильное обучение весов, и, следовательно, препятствует нахождению хороших минимумов функции стоимости.

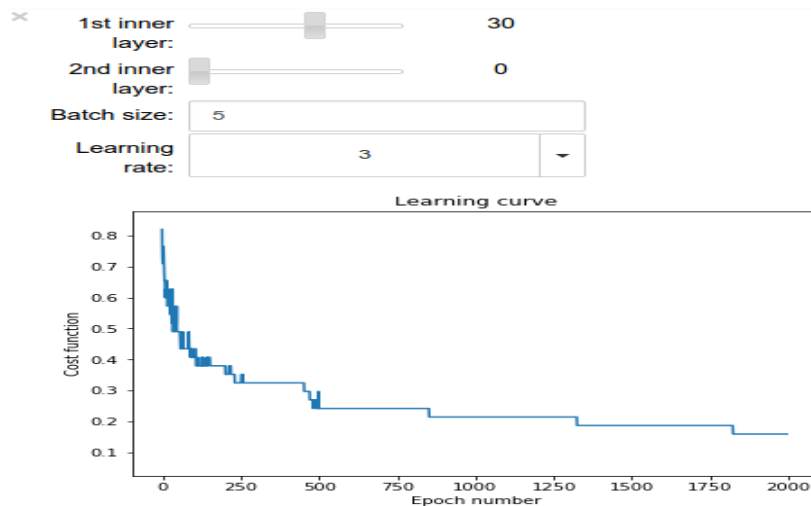


Рисунок 1 – Пример обучения нейронной сети без регуляризации

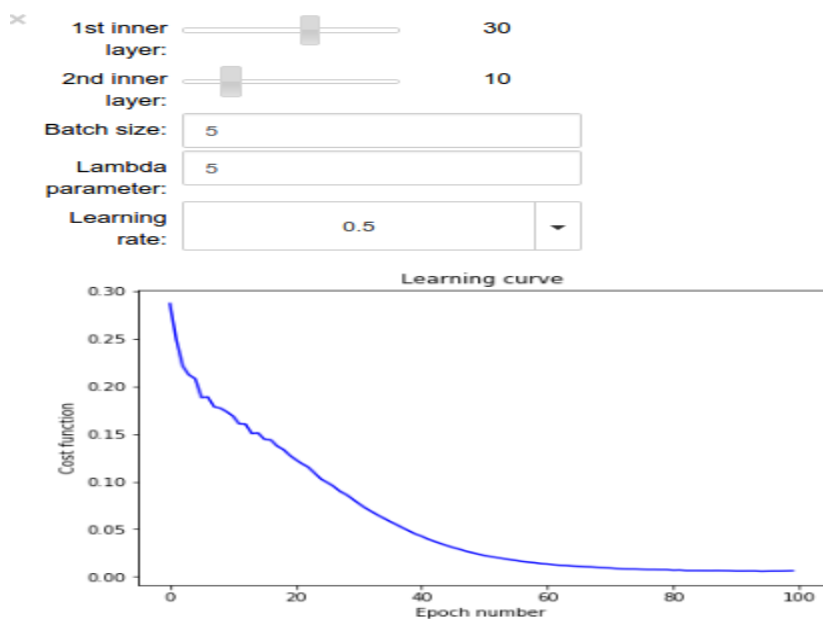


Рисунок 2 – Пример обучения нейронной сети с регуляризацией L2

При $\eta = 0.5$ и $\lambda = 5$ стоимость плавно снижается до конечной эпохи. Из приведенного графика можно заметить хорошую скорость обучения нейронной сети в сравнении с прошлыми экспериментами (уменьшилось количество эпох обучения). Также уже не наблюдается того колебательного поведения функции, которое было, скорее всего, связано со слишком большими значениями «learning rate», что приводило к постоянному выходу из зон минимума.

В новых экспериментах удалось достичь 100% распознавания на обучающих примерах и при этом для этого, в среднем, требовалось около 100 эпох при различных параметрах настройки сети.

Такие хорошие результаты, конечно, были получены из-за малого количества обучающих примеров, но зато можно увидеть реальные изменения в обучении сети, которые были достигнуты за счет изменения функции стоимости и добавления регуляризации.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Галушкин, А.И. Нейронные сети: основы теории. / А.И. Галушкин. – М.: РиС, 2015. – 496 с.
2. Круглов, В.В. Искусственные нейронные сети. Теория и практика: моногр. / В.В. Круглов, В.В. Борисов. – М.: Горячая линия – Телеком; Издание 2-е, стер., 2002. – 382 с.