

ОПТИМИЗАЦИЯ ГИПЕРПАРАМЕТРОВ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ

Раднаев Ч. Б.

Научный руководитель: Гергет О. М.

Консультант: Девярых Д. В.

Национальный Исследовательский Томский политехнический университет

chingis_radnaew@mail.ru

Введение

В настоящее время актуальным является использование нейронных сетей при распознавании образов. При проектировании нейронной сети важным этапом является подбор гиперпараметров, т.е. параметров, которые не изменяются в ходе обучения нейронной сети. Нейронные сети, основанные на методах извлечения признаков, примерно могут иметь от 10 до 50 гиперпараметров, которые устанавливает экспериментатор для достижения требуемого результата. Сложность настройки этих моделей обуславливается тем, что необходимо найти компромисс между качеством распознавания и скоростью обучения. Усовершенствование методов оптимизации гиперпараметров и разработка новых методов является одной из самых острых проблем.

Целью данной работы является подбор оптимальной комбинации гиперпараметров, при которой достигается требуемая точность распознавания изображений.

Описание метода решения

Входные данные

В качестве входных данных выбрана база данных MNIST, содержащая 60 000 изображений 28*28 рукописных символов в обучающей выборке и 10 000 в тестовой выборке (рис. 1).

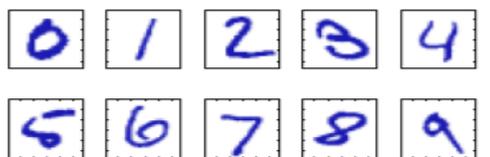


Рис. 1. Примеры входных изображений

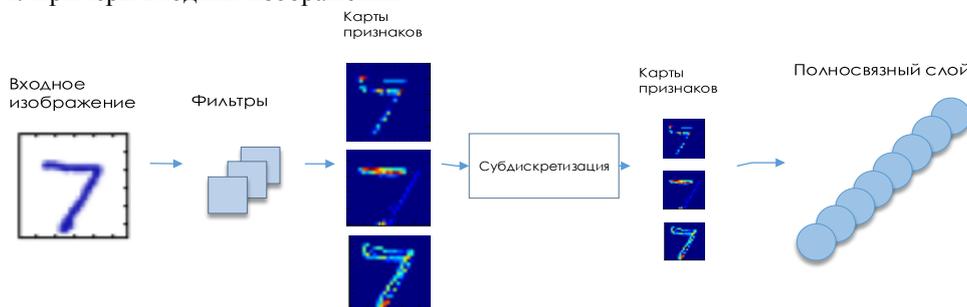


Рис. 2. Схема построения сверточной сети

Архитектура сверточной нейронной сети

Сверточные нейронные сети были предложены Я. Лекуном как специализированная архитектура для распознавания изображений [1]. Так как входное значение является изображением, то оно представляется в виде двумерной матрицы, что позволяет сохранить информацию о соседних точках. Сверточные сети характеризуются наличием чередующихся сверточных и субдискретизирующих слоев. Сверточный слой представляет собой набор фильтров (сверток) – матриц нейронов малого размера.

Выходом сверточного слоя являются карты признаков, элементы которых получаются в результате скалярного произведения фильтра с участком входного слоя размером с сам фильтр. Каждый фильтр выдает одну карту признаков. В ходе обучения параметры фильтров настраиваются так, что получаемые двумерные карты признаков характеризуют наличие во входном изображении с определенным признаком.

С применением к картам признаков дополнительных сверточных слоев выявляются признаки более высокого порядка. Далее, к картам признаков применяются слои субдискретизации (подвыборки). Слои субдискретизации применяются для пространственного уменьшения входного вектора, что ведет к инвариантности относительно масштаба и вращения. Последний полносвязный слой объединяет все карты признаков и выдает вектор выходов нейронной сети, каждое значение которой характеризует близость входного изображения к тому или иному классу [2]. На рисунке 2 изображена схема сверточной нейронной сети.

Алгоритм обучения

К многослойным нейронным сетям при обучении применяется алгоритм обратного распространения ошибки. Данный алгоритм является способом вычисления градиента активационной функции для скрытых слоев нейронной сети.

Для коррекции весовых коэффициентов использовался градиентный алгоритм Adadelta переменным параметром скорости [3].

Оценка адекватности модели

Проверка адекватности модели проводилась с помощью кроссвалидации.

Также для оценки вероятности распознавания использовались свойства функции *softmax* (формула 1) в качестве активационной функции выходного слоя, что позволяет интерпретировать выходные значения сети как вероятности прогноза того или иного класса [4].

$$S_i = \frac{e^{\beta O_i}}{e^{\alpha + \sum_k E^{\beta O_k}}}, \quad (1)$$

где O_i – активационный потенциал i -го нейрона, а S_i – показатель *Softmax* для класса i .

Член, включающий α , представляет собой уровень активации искусственной $N+1$ -й категории, означающий «ни один из вышеперечисленных».

Проведение вычислительного эксперимента Программная реализация моделей

В данной работе рассматривались следующие архитектуры нейронных сетей: многослойный перцептрон и сверточная нейронная сеть. Схема рассматриваемых архитектур приведена в таблице 1.

Таблица 1. Архитектуры рассматриваемых нейронных сетей

№ слоя/ архитектура		1	2	3	4	5	6	7	8
Тип слоя	Сеть с одним сверточным слоем	Vx	CC	CD	PC	PC	V		
	Сеть с двумя сверточными слоями	Vx	CC	CD	CC	CD	PC	PC	V
	Многослойный перцептрон	Vx	PC	PC	V				

Сокращения: *Vx* – входной слой, *CC* – сверточный слой, *CD* – субдискретизационный слой, *PC* – полносвязный слой, *V* – выходной слой.

Результаты вычислений

Была написана программа на языке *Python* с использованием библиотеки моделирования нейронных сетей *Keras* и *Theano*, с помощью которой проведены обучение и тестирование нейронных сетей для каждой из вариаций значений

гиперпараметров. Для каждой комбинации гиперпараметров проведена кроссвалидация по обучающей выборке. Число эпох обучения для всех случаев равен 5.

Итоговые результаты с наилучшими показателями в своем классе по всем рассмотренным архитектурам были отобраны в результате проведения многочисленных вычислительных экспериментов (см. таблицу 2).

Таблица 2. Итоговые результаты вычислений

Архитектура	Гиперпараметры	Процент распознанных примеров, %	Число весовых коэффициентов
Многослойный перцептрон	PC(800),PC(200),PC(10)	8.31	790,210
Сеть с одним сверточным слоем	CC(3*3,64),CD(2*2),PC(128),PC(64),PC(10)	6.15	1,394,122
Сеть с двумя сверточными слоями	CC(3*3,16),CD(2*2),C(3*3,32),CD(2*2),PC(128),PC(64),PC(10)	3.30	83,066

Сокращения: $PC(n)$ – полносвязный слой с n нейронами, $CC(n*m,k)$ – сверточный слой с k фильтрами размером $n*m$, $CD(n,m)$ – субдискретизационный слой с ядром $n*m$.

Заключение

- Наилучшая точность распознавания достигается при размере фильтра 3*3.
- С увеличением количества фильтров растет точность распознавания, но в то же время повышается количество настраиваемых параметров. Наилучшим для сети с одним сверточным слоем является число фильтров 32, а для двух сверточных слоев – 8 и 32.
- При использовании сети с двумя сверточными слоями используется меньшее число настраиваемых параметров, чем при использовании однослойной при достижении точности одного порядка.
- Полносвязная сеть уступает в точности сетям со сверточными слоями, и для достижения

точности одинакового порядка требуется использование большого числа параметров.

Список используемых источников

1. Lecun Y. Convolutional Neural Networks for Images, Speech, and Time-Series [Статья]/ Lecun Y., Bengio Y. // The Handbook of Brain Theory and Neural Networks. – Boston : MIT Press, 1995.
2. Lecun Y. Proc. International Symposium on Circuits and Systems (ISCAS'10), IEEE [Конференция]/ Y. Lecun, K. Kavukcuoglu, C. F. Farabet // Convolutional Networks and Applications in Vision. – 2010.
3. D. Zeiler. Adadelta: an adaptive learning rate method. / D. Zeiler, D. Matthew – CoRR, 2012.
4. Bridle J. S. Probabilistic interpretation of feedforward classification network outputs with relationships to statistical pattern recognition [Журнал]. – [б.м.] : Springer-Verlag, 1989.