

1. Какое преимущество дают сверточные нейронные сети над обычными сетями прямого распространения?

Использование единого весового ядра освобождает сеть от создания и корректировки слишком большого количества весов. Этот подход хорошо зарекомендовал себя в плане эффективности и оптимизации в области распознавания образов (например, изображений). Такая структура сети позволяет ей при обучении обобщать поступающую информацию, а не анализировать данных поэлементно. Отсюда следует гибкость сверточных сетей к «сдвигам образов», т. е., например, в классификации изображений сверточные сети будут менее чувствительны к их поворотам или сдвигам.

2. Для чего нужна регуляризация весов?

В процессе обучения какие-то веса могут начать приобретать слишком большие или слишком маленькие значения. Например, какой-то один вес значительно возрастет и в результате сильно увеличит сумму взвешенных входных параметров, поступающую в функцию активации в соответствующем нейроне. Подобное поведение модели при обучении может приводить к сильной чувствительности ее к небольшим изменениям (колебаниям, шумам) входных данных. Алгоритмы регуляризации направлены, в основном, на корректировку весов в таких ситуациях, т. е. на приведение значений весов модели на определенных слоях к одним числовым промежуткам.

3. Как могут сказаться на качестве обучения не нормализованные данные?

Разные входные параметры принадлежат, как правило, к разным промежуткам, однако при последующем их использовании сетью для получения ответа на задачу серьезные различия в значениях входных данных могут привести к некорректным результатам, например, это может привести к сильной чувствительности модели к незначительному изменению определенных входных параметров. Сеть рано или поздно научится работать с такими данными, но на это уйдет слишком много времени обучения. Поэтому принято нормализовывать входные параметры перед использованием, т. е. приводить их к единому числовому промежутку.

4. На странице 6 Вы утверждаете, что при 6 блоках у Вас наиболее низкое отклонение. Как Вы выяснили, что это наиболее низкое отклонение и ниже быть не может?

Модель исследовалась при различных значениях K . Целью была сборка, при которой средняя оценка отклонений прогнозов стоимости домов будет наименьшей, и такая, при которой время обучения модели будет наиболее оптимальным. Показатели уменьшались с ростом K , однако данная тенденция изменилась в обратную сторону после $K = 6$ (оценка на 6-ти блоках сильнее уменьшилась по отношению к 5-ти, чем, например, при 5-ти по

отношению к 4-м и т. д.). В отчете для наглядности приведены результаты для $K = 7$ и 8. Далее тесты не проводились, т. к. время обучения таких моделей стало слишком велико для целесообразности их использования.

5. Как можно улучшить перекрестную проверку по k блокам?

Необходимо перемешать примеры рабочей выборки (на которой осуществляется обучение и тестирование модели).

6. Почему анализируется mae , а не mse ?

MSE основывается на вычислении среднего квадрата разности между вычисленным и целевым значением, а MAE – на той же разности, но просто по модулю. Следовательно, MSE будет более чувствительно к скачкам (выбросам), что на больших отклонениях неблагоприятно скажется на результате вычислений. Другими словами, среднеквадратическое отклонение придает относительно большой вес большим ошибкам.

7. Почему Вы установили `batch_size` равный 1?

Такое значение наиболее целесообразно на данной (небольшой) обучающей выборке. Значение `batch_size = 1` было приведено в листинге работы и не являлось объектом исследования, поэтому я не включил это в отчет. Однако протестировал модель с разными значениями `batch_size` (кол-во эпох – 50): при единице модель дает наименьшую оценку отклонений прогнозов (около 2.25). При значении 5 оценка увеличилась до 2.3, при 10-ти – примерно 2.32, при 30-ти – 2.4. Следовательно `batch_size = 1` – остается наиболее оптимальным выбором.