

Фамилия, имя, номер группы:

.....

Чем больше преград, тем жёстче подход.

Я не вижу берегов, мне нужен мой запретный плод.

The Hatters, Shoot Me (2021)

Это нулевой вариант мидтёрма. Он нужен для того, чтобы его формат не стал для вас сюрпризом. Работа будет состоять из трёх частей: тестовая, задачи и ответы на открытые вопросы. Списывание будет караться обнулением работы. **Разрешается использовать шпаргалку формата А4, исписанную с двух сторон. Передача шпаргалки другим людям запрещена. Удачи!**

Часть первая: тестовая

Дайте ответ на 10 тестовых вопросов. Каждый вопрос стоит 3 балла. Никакие дополнительные пояснений в этой части работы от вас не требуются. В части вопросов верными могут быть несколько ответов. За неправильные ответы вы не получаете никаких штрафов.

Вопрос 1. Какие из следующих функций активации могут привести к затуханию градиентов и параличу глубокой нейронной сети?

☐ A ReLU

☐ C Leaky ReLU

☐ E Mish

☐ B Tanh

☐ D Сигмоида

☐ F Нет верного ответа.

Вопрос 2. Какие из следующих техник могут быть использованы для аугментации данных в задаче распознавания изображений?

☐ A Зеркальное отражение

☐ C Случайные повороты

☐ E Масштабирование (зум)

☐ B Случайное обрезание

☐ D Смена цветовой гаммы

☐ F Нет верного ответа.

Вопрос 3. У нас есть полносвязная сеть с тремя слоями. На каком-то из наблюдений, на последнем слое нейросеть выплёвывает вектор $[0.3, 0.3, 0.3]$. К нему применяется некоторая функция активации. Каким будет результат её применения, если мы решаем задачу классификации на три класса?

☐ A $[0.5, 0.5, 0.5]$

☐ C $[0.2, 0.5, 0.3]$

☐ E $[0.3, 0.3, 0.3]$

☐ B $[1, 0, 0]$

☐ D $[0, 0, 0]$

☐ F Нет верного ответа.

Вопрос 4. Какой сложностью по времени обладает алгоритм обратного распространения ошибки? Под m тут имеется в виду глубина сети.

☐ A $O(m!)$

☐ C $O(m)$

☐ E $O(\ln m)$

☐ B $O(m^2)$

☐ D $O(m \ln m)$

☐ F Нет верного ответа.

Вопрос 5. Предположим, что на первом слое нейронной сети у нас есть 5 свёрток размера 7×7 с дополнением нулями (zero padding) и сдвигом равным единице (stride). На вход подаётся изображение размера $224 \times 224 \times 3$. Какого размера будет тензор, который пройдёт через этот слой.

- ☐ A $217 \times 217 \times 3$ ☐ C $220 \times 220 \times 5$ ☐ E $218 \times 218 \times 5$
☐ B $217 \times 217 \times 8$ ☐ D $224 \times 224 \times 5$ ☐ F Нет верного ответа.

Вопрос 6. Выберите все верные утверждения

- ☐ A Нормализация по батчам — это новый способ реализации дропаута.
☐ B Нормализация по батчам ускоряет обучение
☐ C В нормализации по батчам мы усредняем столбцы матрицы (признаки).
☐ D Нормализация по батчам может конфликтовать с дропаутом
☐ E Нормализация по батчам — это нелинейная трансформация данных, которая делает среднее нулевым.
☐ F Нет верного ответа.

Вопрос 7. Вы обучаете свёрточную нейронную сеть на ImageNet. Для этого вы собираетесь использовать градиентный спуск. Какие из перечисленных утверждений являются правдой? Под скоростью тут имеется в виду время в минутах.

- ☐ A Возможно, SGD по одному наблюдению сойдётся быстрее, чем SGD по мини-батчам.
☐ B Возможно, что обычный градиентный спуск сойдётся быстрее, чем Adam.
☐ C Возможно, SGD по мини-батчам сойдётся быстрее обычного градиентного спуска.
☐ D Возможно, что Adam сойдётся быстрее, чем обычный градиентный спуск.
☐ E Возможно, SGD по мини-батчам сойдётся быстрее, чем SGD по одному наблюдению.
☐ F Нет верного ответа.

Вопрос 8. Предположим, что у нас есть глубокая нейронная сеть обученная для классификации на ImageNet. Мы хотим научиться классифицировать автомобили 10 разных типов: спорткары, траки, минивэны и т.д. В нашем распоряжении есть выборка из 300 картинок. Выберите все способы обучить модель, которые дадут хорошее качество классификации.

- ☐ A Обучение свёрточной нейросети с нуля на 300 изображениях.
☐ B Заморозить в большой сети все слои и доучить для решения нашей задачи только последний.
☐ C Случайно инициализировать в большой сети все веса и обучить её для нашей задачи.
☐ D Заморозить в большой сети случайную половину слоёв, а вторую половину доучить.
☐ E Заморозить в большой сети все слои и доучить для решения нашей задачи два последних.
☐ F Нет верного ответа.

Вопрос 9. Выберите все верные утверждения о современных свёрточных архитектурах.

- | | | |
|---|--|---|
| <input type="checkbox"/> A ResNet это обычная свёрточная сетка, но увеличенная в тысячу раз. | <input type="checkbox"/> C Ключевой элемент ResNet это skip-connection, который добавляется чтобы исправлять ошибки предыдущих слоёв. | <input type="checkbox"/> E Сделать одну свёртку размера 5×5 предпочтительнее, чем сделать друг за другом две свёртки размера 3×3 . |
| <input type="checkbox"/> B VGG и ResNet показывают сопоставимые результаты на ImageNet. | <input type="checkbox"/> D В современных архитектурах отказались от идеи свёр- | <input type="checkbox"/> F Нет верного ответа. |

Вопрос 10. Выберите все хорошие практики, используемые при организации DL-экспериментов.

- | | | |
|---|---|--|
| <input type="checkbox"/> A Делать по одному изменению за раз | <input type="checkbox"/> C Тестировать на одном батче | <input type="checkbox"/> E Никогда не использовать нормализацию по батчам |
| <input type="checkbox"/> B Логировать все эксперименты | <input type="checkbox"/> D Использовать проверенные временем архитектуры | <input type="checkbox"/> F Нет верного ответа. |

Часть вторая: задачи

Все ответы должны быть обоснованы. Решения должны быть прописаны для каждого пункта. Рисунки должны быть чёткими и понятными. Все линии должны быть подписаны. За решение каждой задачи можно получить 8 баллов.

Вопрос 11. Рассмотрим следующие функции активации: SoftPlus и ReLU:

$$\text{ReLU}(z) = \max(0, z) \quad \text{SoftPlus}(z) = \ln(1 + e^z).$$

1. В современных нейронных сетях SoftPlus в отличие от ReLU практически не используется. Перечислите основные преимущества ReLU перед SoftPlus.
2. Выпишите уравнения для шага обратного распространения для обеих функций активации.
3. Как взаимосвязаны SoftPlus и сигмоида? Может ли привести использование SoftPlus к параличу нейронной сети?
4. ReLU потенциально может занулить выход из слоя нейроной сети. Как на практике решают эту проблему?

Вопрос 12. У Мирона есть картинка размера 3×4 и свёртка размера 3×3

0	1	0	2
1	0	1	0
0	2	0	0

Картинка

0	1	0
1	1	2
0	2	0

Свёртка

1. Пусть используется дополнение нулями (zero padding). Найдите результат применения свёртки к исходной картинке.
2. Пусть при свёртке не используется никаких дополнений. Найдите результат применения свёртки к исходной картинке.

Вопрос 13. Свёрточная сеть состоит из N слоёв. Каждый слой состоит из свёртки с ядром размера 3×3 , сигмоиды и max-pooling размера 2×2 . Входное изображение имеет размер 512×512 .

1. Какого размера будет выход первого слоя?
2. Какого размера будет выход второго слоя?
3. Каким должно быть N , чтобы на выходе получилось изображение размера 1×1 ? Каким при этом N будет поле обзора (receptive field)?

Вопрос 14. Пусть A, X — матрицы размера $n \times n$. Для следующей задачи оптимизации найдите ее множество решений и оптимальное значение целевой функции

$$\operatorname{tr}(A^T X) - \ln \det(X) \rightarrow \min_X.$$

Почему получившееся значение будет точкой минимума?

Подсказка: подумайте как выглядит $A^T X$ и будет ли эта функция выпуклой вверх или вниз. Вторую производную искать не требуется.

Вопрос 15. В коробке на кухне завалялось три персептрона, у каждого два входа с константой и пороговая функция активации:

$$f(h) = \begin{cases} 1, & h \geq 0 \\ 0, & h < 0. \end{cases}$$

Реализуйте с помощью них функцию:

$$y = \begin{cases} 1, & \text{если } x_2 \geq |x_1 - 3| + 2; \\ 0, & \text{иначе} \end{cases}.$$

Изобразите соответствующую нейросеть в виде картинки.

Часть третья: открытые вопросы

Эта часть состоит из открытых вопросов. На них необходимо дать краткие, но ёмкие ответы. За ответ на каждый вопрос можно получить 5 баллов.

Вопрос 16. Объясните что такое свёртка 1×1 , где она применяется и зачем.

Вопрос 17. Почему сверточные слои чаще используются для обработки изображений, чем полностью связанные слои?

Вопрос 18. Для чего нужен метод инерции (momentum)? Как он работает? В чём его основное отличие от метода инерции с поправкой Нестерова? Запишите формулы.

Вопрос 19. Что такое fine-tuning нейронных сетей? Почему качество при таком способе обучения, как правило, выше, чем при обучении из случайной инициализации?

Вопрос 20. Винни-Пух учит нейросеть отличать правильный мёд от неправильного. Он использует две архитектуры. Как думаете какая из них покажет лучший результат и почему?

Dense(64, 'relu')
Dense(32, 'relu')
Dense(16, 'relu')
Dense(1, 'sigmoid')

Dense(128, 'relu')
Dense(10, 'relu')
Dense(64, 'relu')
Dense(1, 'sigmoid')

Вопрос 21. Предположим, что у нас есть n наблюдений. Мы хотим обучить глубокую нейронную сеть. В процессе мы хотим подобрать скорость обучения, размер батча и способ регуляризации, которые дадут нам лучшую модель. Подробно опишите стратегию, с помощью которой вы будете это делать.