

Фамилия, имя, номер группы:

.....

(Профессор Фарнсворт, наблюдая в телескоп, как Фрай и Лила убегают с темной стороны Луны):

Бог мой! Надо бы что-то предпринять... Но я уже в пижаме...

(Фарнсворт засыпает на стуле)

Футурама (1999)

Работа состоит из трёх частей: тестовая, задачи и ответы на открытые вопросы. Списывание карается обнулением работы. Ко всем одинаковым формулировкам я буду очень сильно придираюсь. 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.

Вопрос 1. Что из этого формула для шага в градиентном спуске?

- ☐ A $w_t = w_{t-1} - \eta \cdot \nabla L(w_t)$ ☐ C $w_t = w_{t-1} - \eta \cdot \nabla L(w_{t-1})$ ☐ E $w_t = w_{t-1} + \eta \cdot \nabla L(w_{t-1})$
☐ B $w_t = w_{t-1} + \eta \cdot \nabla L(w_t)$ ☐ D $w_t = w_{t-1} - \eta \cdot \nabla L(w_0)$ ☐ F Нет верного ответа.

Вопрос 2. Какие из следующих функций активации подходят для обучения нейронных сетей?

- ☐ A $f(z) = \max(0.25z, 0.75z)$ ☐ C $f(z) = 0.25z$ ☐ E $f(z) = 1 - z$
☐ B $f(z) = \min(0, z)$ ☐ D $f(z) = \begin{cases} 1, & x > 0.5 \\ -1, & x \leq 0.5 \end{cases}$ ☐ F Нет верного ответа.

Вопрос 3. Яродар берёт картинку и пририсовывает ей с каждой стороны четыре дополнительные клетки. Свёртка какого размера не изменит размер исходной картинки?

- ☐ A 3×3 ☐ C 5×5 ☐ E 7×7
☐ B 4×4 ☐ D 6×6 ☐ F Нет верного ответа.

Вопрос 4. Какие из следующих техник можно использовать, чтобы избежать переобучения?

- ☐ A Аугментация данных (Data augmentation) ☐ C Использовать Adam вместо SGD ☐ E Дропаут (Dropout)
☐ B Нормализация по батчам (Batch Normalization) ☐ D Ранняя остановка обучения (Early stopping) ☐ F Нет верного ответа.

Вопрос 5. После обучения нейронной сети, вы обнаружили, что ассигасу на тренировочной выборке равен 100%, а ассигасу на тестовой выборке равен 42%. Какие из следующих методов могут помочь сократить разницу между этими двумя метриками?

- | | | |
|---|---|---|
| <input type="checkbox"/> A Сигмоида в качестве активации | <input type="checkbox"/> C Алгоритм обратного распространения ошибки | <input type="checkbox"/> E RMSprop |
| <input type="checkbox"/> B Dropout | <input type="checkbox"/> D Softmax | <input type="checkbox"/> F Нет верного ответа. |

Вопрос 6. Выберите все верные утверждения

- | | | |
|---|---|--|
| <input type="checkbox"/> A Если мы используем сигмоиду в качестве функции активации, при обратном проходе через неё производная никогда не поменяет знак | <input type="checkbox"/> C Если мы используем Leaky ReLU в качестве функции активации, при обратном проходе через неё производная никогда не поменяет знак | <input type="checkbox"/> E В нормализации по батчам нет параметров, которые можно было бы обучить алгоритмом обратного распространения ошибки |
| <input type="checkbox"/> B Картинки нормируют на отрезок $[-1; 1]$ для более быстрой сходимости обучения | <input type="checkbox"/> D Картинки нормируют на отрезок $[0; 1]$ для более быстрой сходимости обучения | <input type="checkbox"/> F Нет верного ответа. |

Вопрос 7. Выберите все верные утверждения об инициализации Хе

- | | | |
|---|--|---|
| <input type="checkbox"/> A Такая инициализация используется только в полносвязных сетях. | <input type="checkbox"/> C Такая инициализация используется для симметричных функций активации вроде сигмоиды | <input type="checkbox"/> E Инициализация Хе корректирует параметры распределений в зависимости от входа и выхода слоя так, чтобы поддерживать дисперсию равной единице |
| <input type="checkbox"/> B Такая инициализация используется только в свёрточных сетях. | <input type="checkbox"/> D Такая инициализация используется для функций активации вроде ReLU | <input type="checkbox"/> F Нет верного ответа. |

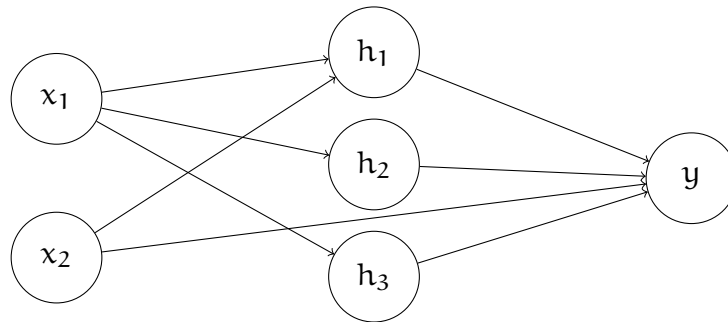
Вопрос 8. У нас есть 10 классов. Классификатор предсказывает, что объект равновероятно относится к каждому из них. Какое значение принимает logloss на этом объекте?

- | | | |
|---|--|---|
| <input type="checkbox"/> A $-\log 10$ | <input type="checkbox"/> C $-\log 0.1$ | <input type="checkbox"/> E $-10 \log 10$ |
| <input type="checkbox"/> B $-0.1 \log 1$ | <input type="checkbox"/> D $-10 \log 0.1$ | <input type="checkbox"/> F Нет верного ответа. |

Вопрос 9. К изображению размера $W \times H$ применили D фильтров. Получилось новое изображение размера $W \times H \times D$. К нему мы хотим применить свёртку размера 1×1 . Сколько параметров нам надо будет обучить в этой свёртке?

- | | | |
|---|---------------------------------------|---|
| <input type="checkbox"/> A 1 | <input type="checkbox"/> C D | <input type="checkbox"/> E H |
| <input type="checkbox"/> B $W \cdot H$ | <input type="checkbox"/> D W | <input type="checkbox"/> F Нет верного ответа. |

Вопрос 10. Выберите функции, которые могли бы описать части вычислительного графа, представленного ниже. Рёбрам соответствуют какие-то веса, а вершинам любые алгебраические операции и применения основных элементарных функций (экспонента, косинус и тп).



☐ A $h_1 = x_1^2 + x_2$

☐ C $h_3 = 2x_1 + 3x_2$

☐ E $h_3 = 2x_1$

☐ B $h_2 = e^{x_1}$

☐ D $y = 2h_1 + 3h_2 + h_3$

☐ 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 две функции активации: сигмоиду и гиперболический тангенс

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}.$$

1. Как взаимосвязаны $\sigma(z)$ и $\sigma'(z)$? Как взаимосвязаны $\tanh(z)$ и $\tanh'(z)$?
2. Выпишите уравнения для шага обратного распространения для обеих функций активации.
3. Что такое паралич нейронной сети? Как сигмоида способствует ему? Способствует ли параличу гиперболический тангенс?
4. Объясните, почему использование гиперболического тангенса вместо сигмоиды делает оптимизацию нейронной сети проще. Какие проблемы использование тангенса не исправляет?

Вопрос 12. Предположим, что у вас есть цветная картинка размера $10 \times 10 \times 3$. Вы хотите собрать сетку из двух свёрточных слоёв с ядром размера 3×3 с 10 и 20 фильтрами. Сколько параметров нужно будет оценить в этих двух слоях? Не забудьте про константу!

Вопрос 13. Мы обучаем нейронную сеть из N свёрточных слоёв размера $k \times k$ с параметром сдвига (stride) равным единице. Дополнение картинки (padding) не делается. Нумерация слоёв начинается с единицы. Выпишите формулу, по которой можно вычислить размер поля обзора (receptive field) сети.

Вопрос 14. Решите следующую задачу матричной оптимизации. Убедитесь, что найденное решение действительно является минимумом.

$$f(x) = x^T A x - x^T b + c \rightarrow \min_x$$

Вопрос 15. Предположим, что softmax принимает на вход вектор z_1, \dots, z_k и возвращает вектор p_1, \dots, p_k . Мы можем записать функцию в виде двух уравнений

$$r = \sum_{j=1}^k e^{z_j} \quad y_i = \frac{e^{z_i}}{r}.$$

1. Пусть $k = 3$. Нарисуйте граф вычислений, на котором описывается взаимосвязь между величинами $z_1, z_2, z_3, p_1, p_2, p_3$ и r .
2. Опишите алгоритм обратного распространения ошибки для softmax-слоя. Выпишите формулу, по которой обновляется градиент при проходе через softmax.

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. Есть теорема, которая говорит, что двухслойной нейронной сетью можно приблизить любую непрерывную функцию. Почему люди не ограничиваются двумя слоями и учат глубокие нейронные сети?

Вопрос 17. Можно ли утверждать, что оптимизация градиентным спуском гарантирует нахождение глобального оптимума весов глубокой нейронной сети? Почему?

Вопрос 18. Объясните, почему инициализировать в нейронной сети веса константами — плохая идея?

Вопрос 19. Представим себе, что вы реализовали глубокую нейросеть и отправили её учиться. Прошло несколько дней. По логам обучения вы поняли, что нейросеть не хочет учиться. Её качество оказалось лучше случайного, но сильно хуже, чем ожидается в вашей задаче. Какие именно действия вы предпримите, чтобы понять, где именно у вас возникли проблемы?

Вопрос 20. Однажды на следующий день после жёсткой вечеринки в середине недели ваш коллега написал такую архитектуру нейросети. Она решает задачу классификации RGB-изображения размера 100×100 на 10 классов. Каждое изображение принадлежит только к одному классу. Какие проблемы есть у этой нейросети?

```
model = Sequential()
model.add(InputLayer([100, 100, 3]))

model.add(Conv2D(filters=512, kernel_size=(3, 3), kernel_initializer="glorot_uniform"))
model.add(Activation('relu'))
model.add(MaxPool2D(pool_size=(2, 2)))

model.add(Conv2D(filters=128, kernel_size=(3, 3), kernel_initializer="glorot_uniform"))
model.add(Activation('relu'))

model.add(Conv2D(filters=32, kernel_size=(3, 3), kernel_initializer="glorot_uniform"))
model.add(Conv2D(filters=32, kernel_size=(1, 1), kernel_initializer="glorot_uniform"))
model.add(MaxPool2D(pool_size=(10, 10)))

model.add(Flatten()) # convert 3d tensor to a vector of features
model.add(Dense(64))
model.add(Dropout(rate=1))
model.add(Dense(10))
model.add(Activation('sigmoid'))
model.add(Dropout(rate=0.5))
```

Вопрос 21. Объясните мем

