

## Вопрос 1

Верно

Баллов: 1,00 из 1,00

Проверим утверждение про затухание градиента на практике. В [документации pytorch](#) можно найти следующие функции активации (самые популярные мы подсветили жирным шрифтом.):

ELU, Hardtanh, **LeakyReLU**, LogSigmoid, PReLU, **ReLU**, ReLU6, RReLU, SELU, CELU, **Sigmoid**, Softplus, Softshrink, Softsign, **Tanh**, Tanhshrink, Hardshrink.

Вам предстоит найти активацию, которая приводит к наименьшему затуханию градиента.

Для проверки мы сконструируем SimpleNet, которая будет иметь внутри 3 fc-слоя, по 1 нейрону в каждом без bias'ов. Веса этих нейронов мы проинициализируем единицами. На вход в эту сеть будем подавать числа из нормального распределения. Сделаем 200 запусков (NUMBER\_OF\_EXPERIMENTS) для честного сравнения и посчитаем среднее значение градиента в первом слое. Найдите такую функцию, которая будет давать максимальные значения градиента в первом слое. Все функции активации нужно инициализировать с аргументами по умолчанию (пустыми скобками).

Для примера:

| Тест                                | Ввод | Результат |
|-------------------------------------|------|-----------|
| print(round(np.mean(fc1_grads), 5)) | 11   | 0.73722   |

Ответ: (штрафной режим: 0 %)

Сброс ответа

```
1 import torch
2 import numpy as np
3
4 seed = int(input())
5 np.random.seed(seed)
6 torch.manual_seed(seed)
7
8 NUMBER_OF_EXPERIMENTS = 200
9
10 class SimpleNet(torch.nn.Module):
11     def __init__(self, activation):
12         super().__init__()
13
14         self.activation = activation
15         self.fc1 = torch.nn.Linear(1, 1, bias=False) # one neuron without bias
16         self.fc1.weight.data.fill_(1.) # init weight with 1
17         self.fc2 = torch.nn.Linear(1, 1, bias=False)
18         self.fc2.weight.data.fill_(1.)
19         self.fc3 = torch.nn.Linear(1, 1, bias=False)
20         self.fc3.weight.data.fill_(1.)
21
22     def forward(self, x):
23         x = self.activation(self.fc1(x))
24         x = self.activation(self.fc2(x))
25         x = self.activation(self.fc3(x))
26         return x
27
28     def get_fc1_grad_abs_value(self):
29         return torch.abs(self.fc1.weight.grad)
30
31 def get_fc1_grad_abs_value(net, x):
32     output = net.forward(x)
33     output.backward() # no loss function. Pretending that we want to minimize output
34                     # In our case output is scalar, so we can calculate backward
35     fc1_grad = net.get_fc1_grad_abs_value().item()
36     net.zero_grad()
37     return fc1_grad
38
39 activation = torch.nn.Hardshrink()
40
41 net = SimpleNet(activation=activation)
42
43 fc1_grads = []
44 for x in torch.randn((NUMBER_OF_EXPERIMENTS, 1)):
45     fc1_grads.append(get_fc1_grad_abs_value(net, x))
46
47 # Проверка осуществляется автоматически, вызовом функции:
48 # print(round(np.mean(fc1_grads), 5))
```

49 | # (раскомментируйте, если решаете задачу локально)  
50

Прошли все тесты! ✓

Верно

Баллы за эту попытку: 1,00/1,00.

## Вопрос 2

Неверно

Баллов: 0,00 из 1,00

Попробуйте решить следующую задачу пользуясь только бумагой и калькулятором:

Допустим, у нас есть нейросеть, состоящая из 4-х полносвязных слоев, в каждом из которых по одному нейрону. Для простоты будем считать, что bias-ы у нейронов отсутствуют, а все веса равны 1.

После каждого слоя мы поставим активации. В первом случае это будут  $\tanh$ , во втором -  $ReLU$ . Не будем добавлять никакую loss-функцию. Тогда нашу сеть можно будет записать в виде функции:

$$f = a_4(w_4 \cdot a_3(w_3 \cdot a_2(w_2 \cdot a_1(w_1 \cdot x))))$$

Где  $a$  - это либо  $\tanh$  либо  $ReLU$ .  $w_i$  - это одно число.

Пусть на вход подали  $x = 100$

Зная, что  $\tanh'(x) = 1 - \tanh^2(x)$ , рассчитайте градиенты весов сети:  $[f'_{w_1}, f'_{w_2}, f'_{w_3}, f'_{w_4}]$  для случая  $a = \tanh$  и для случая  $a = ReLU$ . Результат округлите до 3-го знака.

Формат ответа:

[1234.123, 123.12, 123.1, 123.0], [0.0, 0.1, 0.12, 0.123]

Правда ли, что для активаций гиперболическим тангенсом, градиенты затухают быстрее?

Ответ: (штрафной режим: 0 %)

1 [[1.0, 1.0, 1.0, 1.0], [0.0, 0.0, 0.0, 0.0]]

Ваш код не прошел один или несколько скрытых тестов.

Ваш код должен пройти все тесты, чтобы заработать какие-либо оценки. Пробуйте снова.

Неверно

Баллы за эту попытку: 0,00/1,00.