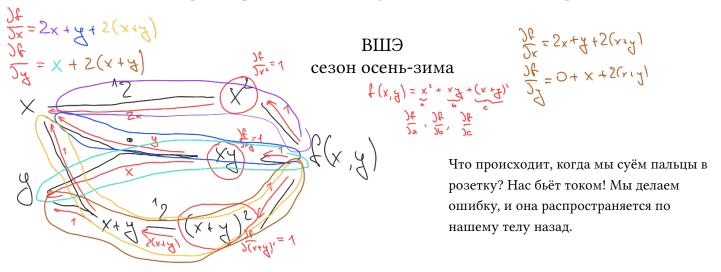
Тятя! Тятя! Наши сети заменили продавца!

Листочек с задачками №1: всего лишь функция

Решения стараюсь собрать тут:

https://github.com/FUlyankin/neural_nets_prob



Упражнение 1 (граф вычислений)

Маша вспомнила картину из кофейни Добродума и решила нарисовать у себя дома свою такую же. Она хочет изобразить для функции

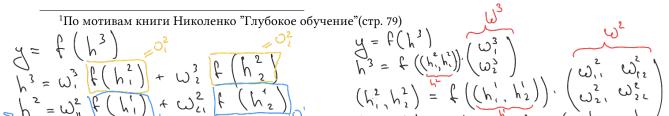
$$f(x,y) = x^2 + xy + (x+y)^2$$

граф вычислений. В кругляшах она будет записывать результаты вычислений. Каждое ребро будет обозначать элементарную операцию: плюс или умножить.

Когда картина будет нарисована, Маша хочет найти производные всех выходов из кругляшей по всем входам. Опираясь на получившийся граф Маша хочет выписать частные производные функции f по χ и по y^1 .

Упражнение 2 (придумываем backpropagation)

Маша умеет собирать нейросети. Например, у неё есть такая нейросеть:

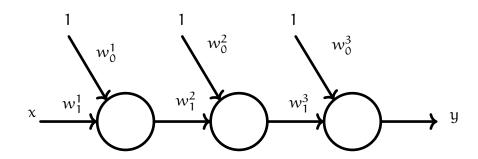


Здесь w_{ij}^k — веса для k слоя, f(t) — какая-то функция активации. Маша хочет научиться делать для такой нейронной сетки градиентный спуск:

- а. Запишите Машину нейросеть как сложную функцию. Сначала в обычной записи, а затем в матричном виде.
- б. Пусть $L(W_1,W_2,W_3)=\frac{1}{2}\cdot (y-\hat{y})^2-$ функция потерь, где W_k- веса k-го слоя. Найдите производные функции L по всем весам $W_k.$
- в. В производных постоянно повторяются одни и те же части. Постоянно искать их не очень оптимально. Выделите эти часть в прямоугольнички цветными ручками.
- г. Выпишите все производные в том виде, в котором их было бы удобно использовать для алгоритма обратного распространения ошибки, а затем, сформулируйте сам алгоритм.

Упражнение 3 (backpropagation руками)

Маша как-то раз решала задачу классифкации. С тех пор у неё в кармане завалялась нейросеть:



В качестве функции активации Маша использовала сигмоиду: $f(t) = \frac{e^t}{1+e^t}$. Как это обычно бывает, Маша обнаружила её в своих штанах после стирки и очень обрадовалась. Теперь она хочет сделать два шага стохастического градиентного спуска, используя алгоритм обратного распространения ошибки.

У неё есть два наблюдения: $x_1=1, x_2=5, y_1=1, y_2=0$. Скорость обучения $\gamma=1$. В качестве инициализации взяты нулевые веса. Сначала берётся второе наблюдение, затем первое. Помогите Маше.

Упражнение 4 (незаметный backpropagation)

Маша собрала нейросеть:

$$y = \max \left(0; X \cdot \begin{pmatrix} 1 & -1 \\ 0.5 & 0 \end{pmatrix}\right) \cdot \begin{pmatrix} 0.5 \\ 1 \end{pmatrix}$$

Теперь Маша внимательно смотрит на неё.

- а. Первый слой нашей нейросетки линейный. По какой формуле делается forward pass? Предположим, что на вход пришло наблюдение x=(1,2). Сделайте через этот слой forward pass и найдите выход из слоя.
- б. Найдите для первого слоя производную выхода по входу. При обратном движении по нейросетке, в первый слой пришёл накопленный градиент (—1, 0). Каким будет новое накопленное значение градиента, которое выплюнет из себя линейный слой? По какой формуле делается backward pass?
- в. Второй слой нейросетки функция активации, ReLU. По какой формуле делается forward pass? На вход в него поступило значение (2,-1). Сделайте через него forward pass.
- г. Найдите для второго слоя производную выхода по входу. При обратном движении по нейросетке во второй слой пришёл накопленный градиент (-1,-2). Каким будет новое накопленное значение градиента, которое выплюнет из себя ReLU? По какой формуле делается backward pass?
- д. Третий слой нейросетки линейный. По какой формуле делается forward pass? Пусть на вход поступило значение (2,0). Сделайте через него forward pass.
- е. Найдите для третьего слоя производную выхода по входу. При обратном движении по нейросетке, в третий слой пришёл накопленный градиент —2. Каким будет новое накопленное значение градиента, которое выплюнет из себя линейный слой? По какой формуле делается backward pass?
- ж. Мы решаем задачу Регрессии. В качестве функции ошибки мы используем MSE. Пусть для рассматриваемого наблюдения реальное значение у = 0. Найдите значение MSE. Чему равна производная MSE по входу (прогнозу)? Каким будет накопленное значение градиента, которое MSE выплюнет из себя в предыдущий слой нейросетки, если изначально значение градиента инициализировано единицей?
- з. Пусть скорость обучения $\gamma = 1$. Сделайте для весов нейросети шаг градиентного спуска.

Посидела Маша, посидела, и поняла, что неправильно она всё делает. В реальности перед ней не задача регрессии, а задача классификации.

- а. Маша навинтила поверх второго линейного слоя сигмоиду. Как будет для неё выглядеть forward pass? Сделайте его. Найдите для сигмоиды производную выхода по входу.
- б. В качестве функции потерь Maшa использует logloss. Как для этой функции потерь выглядит forward pass? Сделайте его. Найдите для logloss производную выхода по входу.
- в. Как будет выглядеть backward pass через logloss и сигмоиду? Прделайте его. Как изменится процедура градиентного спуска для остальной части сети?

Упражнение 5 (Тот же backpropagation)

Пусть у нас есть нейронка:

$$y = f(X \cdot W_2) \cdot W_1$$

Как для функции потерь $L(W_1,W_2)=(y-\hat{y})^2$ будет выглядеть алгоритм обратного распространения ошибки, если $f(t)=\text{ReLU}(t)=\max(0;t)$? Найдите все выходы, все промежуточные производные. Опишите правило, по которому производная будет накапливаться, а также сам шаг градиентного спуска.