



ОНЛАЙН-ОБРАЗОВАНИЕ

Погружение в PyTorch

Динамический граф вычислений
и численные трюки

Артур Кадулин
Преподаватель



План на сегодня

1. **Граф вычислений**
2. Перекрестная энтропия
3. Трюки с softmax
4. Практика: модуль Module
5. Практика: первая нейросеть



Граф вычислений в PyTorch

A graph is created on the fly

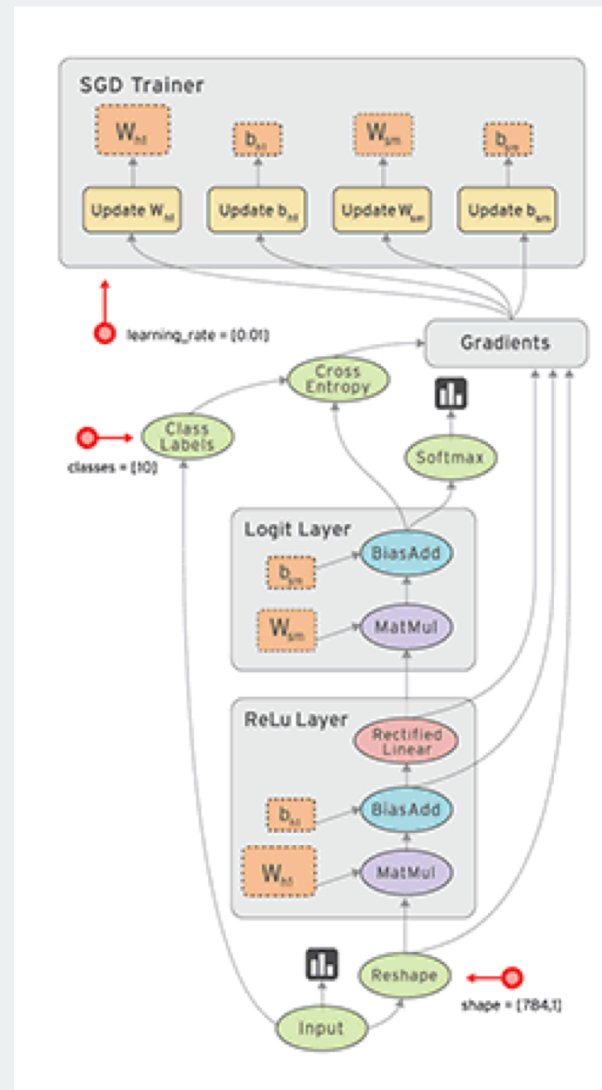
```
x = torch.randn(1, 10)
prev_h = torch.randn(1, 20)
W_h = torch.randn(20, 20)
W_x = torch.randn(20, 10)
```



Изображение с сайта <https://pytorch.org/about/>



Граф вычислений в TensorFlow



Изображение с сайта https://www.tensorflow.org/programmers_guide/graphs



Разница между графами

Динамический:

1. Память выделяется динамически
2. Объекты могут иметь произвольный размер

Статический:

1. Память выделяется при описании графа
2. Можно оптимизировать за счет компиляции



План на сегодня

1. Граф вычислений
- 2. Перекрестная энтропия**
3. Трюки с softmax
4. Практика: модуль Module
5. Практика: первая нейросеть



Информация

1. **Сведения, воспринимаемые человеком и (или) специальными устройствами как отражение фактов материального или духовного мира в процессе коммуникации (ГОСТ 7.0-99).**
2. **Знания о предметах, фактах, идеях и т. д., которыми могут обмениваться люди в рамках конкретного контекста (ISO/IEC 10746-2:1996);**



Количество информации

Количество информации — это числовая характеристика, отражающая степень неопределенности которая исчезает после получения информации.

Можете привести пример?



Количество информации

Количество информации — это числовая характеристика, отражающая степень неопределенности которая исчезает после получения информации.

До начала лотереи 1 из миллионов билетов может оказаться выигрышным. После выпадения первых нескольких чисел количество возможных выигрышных билетов уменьшается.



Количество информации

Количество информации — это числовая характеристика, отражающая степень неопределенности которая исчезает после получения информации.

$$I = \log_2 N = -\log_2 \frac{1}{N} = -\log_2 p$$



Информационная энтропия

Информационная энтропия — мера неопределенности или непредсказуемости некоторой системы.

$$H(P) = - \sum_i p_i \log p_i$$



Перекрестная энтропия

Перекрестная энтропия — среднее количество информации в системе Q необходимое для опознания события из системы P.

$$H(P, Q) = - \sum_i p_i \log q_i$$



Перекрестная энтропия

Пример: Классификация рукописных цифр.

$p =$



$q = [0.03, 0.01, 0.14, 0.20, 0.09, 0.35, 0.13, 0.03, 0.01, 0.01]$



Перекрестная энтропия

Пример: Классификация рукописных цифр.

$$p = [0.00, 0.00, 0.00, 0.00, 0.00, 1.00, 0.00, 0.00, 0.00, 0.00]$$



$$q = [0.03, 0.01, 0.14, 0.20, 0.09, 0.35, 0.13, 0.03, 0.01, 0.01]$$



Перекрестная энтропия

Пример: Классификация рукописных цифр.

$$p = [0.00, 0.00, 0.00, 0.00, 0.00, 1.00, 0.00, 0.00, 0.00, 0.00]$$



$$H(P, Q) = - \sum_i p_i \log q_i = -1.00 * \log 0.35$$

$$q = [0.03, 0.01, 0.14, 0.20, 0.09, 0.35, 0.13, 0.03, 0.01, 0.01]$$



Относительная энтропия

Дивергенция Кульбака-Лейблера или относительная энтропия — это величина потерь информации при переходе от одной системы к другой.

$$\begin{aligned} D_{KL}(P||Q) &= H(P, Q) - H(P) = \\ &= \sum_i p_i \log p_i - \sum_i p_i \log q_i = \sum_i p_i \log \frac{p_i}{q_i} \end{aligned}$$



Относительная энтропия

Дивергенция Кульбака-Лейблера или относительная энтропия — это величина потерь информации при переходе от одной системы к другой.

Когда мы минимизируем кросс-энтропию $H(P, Q)$ по Q — $H(P)$ константа! Поэтому мы одновременно минимизируем и расстояние Кульбака-Лейблера.



План на сегодня

1. Граф вычислений
2. Перекрестная энтропия
- 3. Трюки с softmax**
4. Практика: модуль Module
5. Практика: первая нейросеть



Функция softmax

$$\text{softmax}(\mathbf{z}) = \frac{e^z}{\sum_k e^{z_k}}$$

Какие могут быть проблемы?



Функция softmax

$$\text{softmax}(\mathbf{z}) = \frac{e^z}{\sum_k e^{z_k}}$$

Какие могут быть проблемы?

При относительно небольших абсолютных значениях z e^z может оказаться слишком большим или слишком маленьким.



Первый трюк

$$\text{softmax}(\mathbf{z}) = \frac{e^{\mathbf{z}}}{\sum_k e^{z_k}}$$

$$\text{softmax}(\mathbf{z} - c) = \frac{e^{\mathbf{z} - c}}{\sum_k e^{z_k - c}} = \frac{e^{\mathbf{z}} / e^c}{\sum_k e^{z_k} / e^c} = \frac{e^{\mathbf{z}}}{\sum_k e^{z_k}}$$

При изменении всего вектора на одну и ту же константу значение функции не меняется

Почему это хорошо?



Второй трюк

$$\text{softmax}(\mathbf{z}) = \frac{e^z}{\sum_k e^{z_k}}$$

При вычислении кросс-энтропии мы считаем логарифм от выходов сети.

Какие могут быть проблемы?



Второй трюк

$$\text{softmax}(\mathbf{z}) = \frac{e^z}{\sum_k e^{z_k}}$$

$$\log \text{softmax}(\mathbf{z}) = \log \frac{e^z}{\sum_k e^{z_k}} = \mathbf{z} - \log \sum_k e^{z_k}$$

При вычислении кросс-энтропии мы считаем логарифм от выходов сети.

Какие могут быть проблемы?



План на сегодня

1. Граф вычислений
2. Перекрестная энтропия
3. Трюки с softmax
- 4. Практика: модуль Module**
- 5. Практика: первая нейросеть**





Спасибо
за внимание!