OTUS

ОНЛАЙН-ОБРАЗОВАНИЕ

OTUS

Погружение в PyTorch

Динамический граф вычислений
и численные трюки

Артур Кадурин

План на сегодня

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

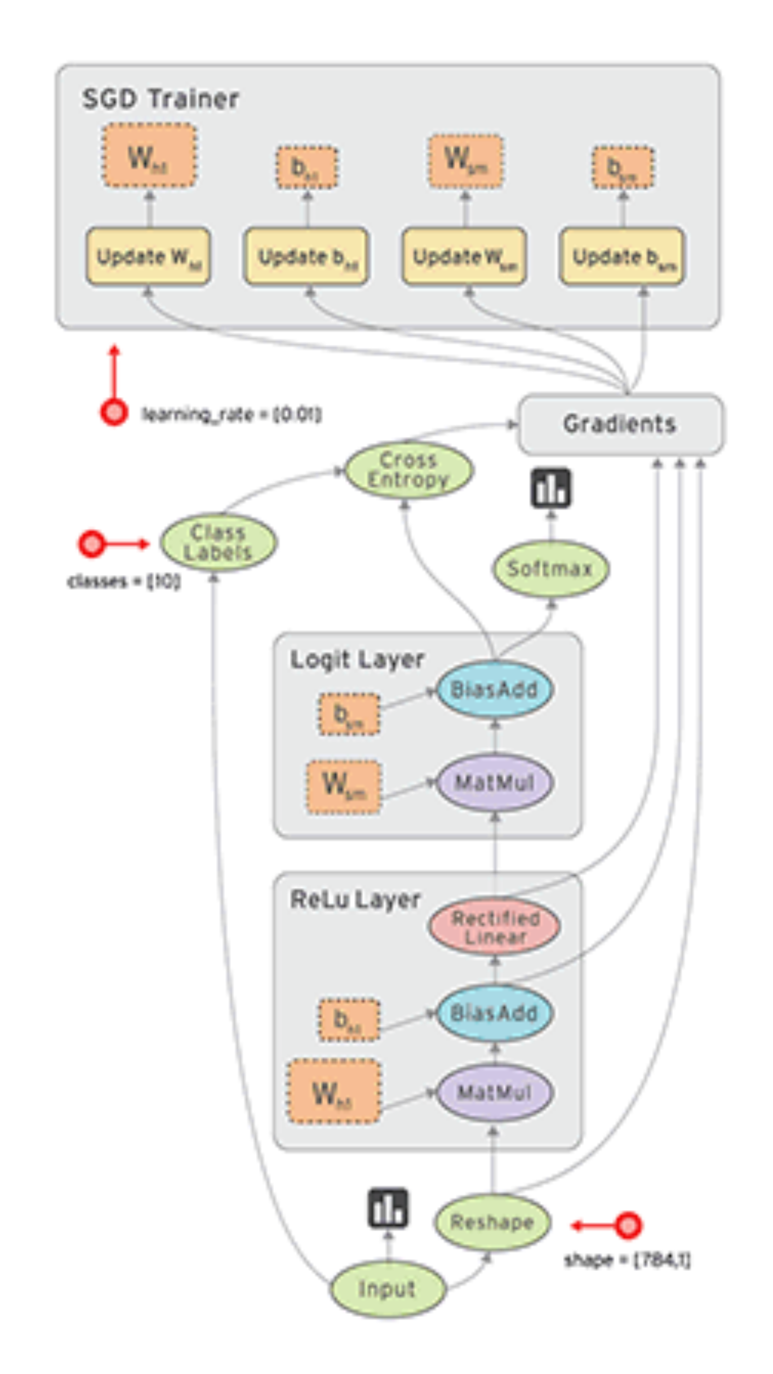
Граф вычислений в PyTorch

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 PyTorch



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.

Можете привести пример?

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.

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

Количество информации

OTUS

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

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

$$\mathrm{I} = -\log_2 p$$

Количество информации часто измеряют в битах. Наблюдая событие, вероятность наступления которого равна p мы узнаем $-\log_2 p$ бит информации

Информационная энтропия

Информационная энтропия — мера неопределенности или непредсказуемости некоторой системы.

$$\mathrm{H}(P) = -\sum_i p_i \log p_i$$

Информационная энтропия

OTUS

Информационная энтропия — мера неопределенности или непредсказуемости некоторой системы.

$$\mathrm{H}(P) = -\sum_i p_i \log p_i$$

По сути, это среднее количество информации которую мы получаем при наблюдении одного события. Если все события равновероятны и их N то энтропия равна...?

Информационная энтропия

Информационная энтропия — мера неопределенности или непредсказуемости некоторой системы.

$$\mathrm{H}(P) = -\sum_i p_i \log p_i$$

По сути, это среднее количество информации которую мы получаем при наблюдении одного события. Если все события равновероятны и их N то энтропия равна $\log_2 N$.

Перекрестная энтропия

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

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

Если мы хотим закодировать числа от 1 до 3 с помощью броска монетки, одним из возможных способов будет такая схема:

OO = 1, OP = 1, PO = 2 PP = 3

Перекрестная энтропия

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

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

Если мы хотим закодировать числа от 1 до 3 с помощью броска монетки, одним из возможных способов будет такая схема:

OO = 1, OP = 1, PO = 2 PP = 3

Если бы у нас была трехгранная монетка, то энтропия системы была бы равна $\log_2 \mathbf{3}$. А в случае с двумя бросками обычной монетки?

Логистическая регрессия

Логистическая регрессия — это модель применяющаяся для предсказания вероятности наступления события в зависимости от значений набора признаков.

$$\mathbb{P}\{y = 1|\boldsymbol{x}\} = \sigma(\boldsymbol{\theta}^T \boldsymbol{x}) \qquad \sigma(\boldsymbol{z}) = \frac{1}{1 + e^{-\boldsymbol{z}}}$$

$$\mathbb{P}\{y|\boldsymbol{x}\} = \sigma(\boldsymbol{\theta}^T \boldsymbol{x})^y \left(1 - \sigma(\boldsymbol{\theta}^T \boldsymbol{x})\right)^{1-y}$$

Функция правдоподобия

Если $\mathbb{P}\{y|\boldsymbol{x}\}$ это вероятность того что мы угадали исход в одном из экспериментов, то произведение вероятностей по всем экспериментам — это «правдоподобность» нашей модели на конкретной выборке, или вероятность корректности модели.

$$\mathbb{P}\{y|\boldsymbol{x}\} = \sigma(\boldsymbol{\theta}^T\boldsymbol{x})^y\left(1 - \sigma(\boldsymbol{\theta}^T\boldsymbol{x})\right)^{1-y}$$

$$L(\boldsymbol{x}, y|\boldsymbol{\theta}) = \prod_{i=0}^{N} \mathbb{P}\left\{y = y^{(i)}\middle|\boldsymbol{x} = \boldsymbol{x}^{(i)}\right\}$$

Функция правдоподобия

Если $\mathbb{P}\{y|\boldsymbol{x}\}$ это вероятность того что мы угадали исход в одном из экспериментов, то произведение вероятностей по всем экспериментам — это «правдоподобность» нашей модели на конкретной выборке, или вероятность корректности модели.

$$\mathcal{L}(\boldsymbol{x}, y|\boldsymbol{\theta}) = -\frac{1}{N}\log L(\boldsymbol{x}, y|\boldsymbol{\theta}) = -\frac{1}{N}\sum_i \left[y^{(i)}\log p^{(i)} + \left(1 - y^{(i)}\right)\log\left(1 - p^{(i)}\right)\right]$$

А логарифм правдоподобия взятый со знаком минус — это количество информации которую мы получаем проверяя корректность модели. И, неожиданно, это и есть перекрестная энтропия между нашими предсказаниями и реальными данными!

Перекрестная энтропия

Пример: Классификация рукописных цифр.

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

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

Артур Кадурин, Insilico Medicine

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.

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

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

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

Какие могут быть проблемы?

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

Трюк первый

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

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

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

Почему это хорошо?

Трюк второй

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

$$\log softmax(\boldsymbol{z}) = \log \frac{e^{\boldsymbol{z}}}{\sum_k e^{z_k}}$$

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

Какие могут быть проблемы?

Трюк второй

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

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

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**

OTUS

Спасибо
за внимание!