

Глубокое обучение и вообще

Ульянкин Филипп и Соловей Влад

19 марта 2020 г.

Посиделка 11: Пора вспомнить все!

1. Нейронные ячейки - наши кубики
2. Хаки специфичные под нейронки
3. Технические хаки

А что такое нейросеть? Что мы можем назвать нейросетью в текущем понимании индустрии?

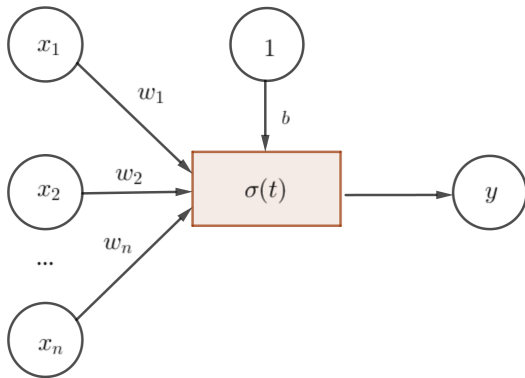
А какие нейронные ячейки помнишь ты?

Dense

Наша база - полносвязная ячейка. Чем она характеризуется? Сколько слоев нам надо на практике?

Dense

Наша база - полносвязная ячейка. Чем она характеризуется? Сколько слоев нам надо на практике?



Функции активации

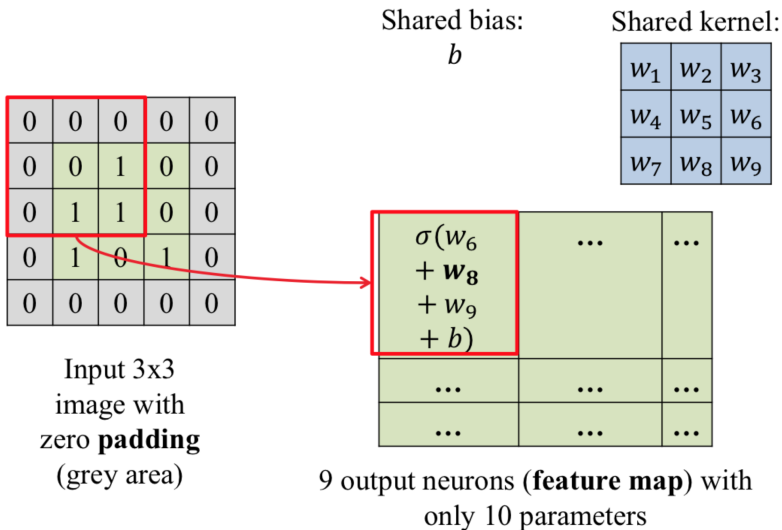
Мы говорим, про функции активации. Вообще какие они бывают и какие они могут быть? На что влияет их выбор?

Функции активации

| Название функции | Формула $f(x)$ | Производная $f'(x)$ |
|---------------------------------|---|---|
| Логистический сигмоид σ | $\frac{1}{1+e^{-x}}$ | $f(x)(1-f(x))$ |
| Гиперболический тангенс \tanh | $\frac{e^x - e^{-x}}{e^x + e^{-x}}$ | $1 - f^2(x)$ |
| SoftSign | $\frac{x}{1+ x }$ | $\frac{1}{(1+ x)^2}$ |
| Ступенька (функция Хевисайда) | $\begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$ | 0 |
| SoftPlus | $\log(1 + e^x)$ | $\frac{1}{1+e^{-x}}$ |
| ReLU | $\begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases}$ | $\begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$ |
| Leaky ReLU, Parameterized ReLU | $\begin{cases} ax, & x < 0 \\ x, & x \geq 0 \end{cases}$ | $\begin{cases} a, & x < 0 \\ 1, & x \geq 0 \end{cases}$ |
| ELU | $\begin{cases} \alpha(e^x - 1), & x < 0 \\ x, & x \geq 0 \end{cases}$ | $\begin{cases} f(x) + \alpha, & x < 0 \\ 1, & x \geq 0 \end{cases}$ |

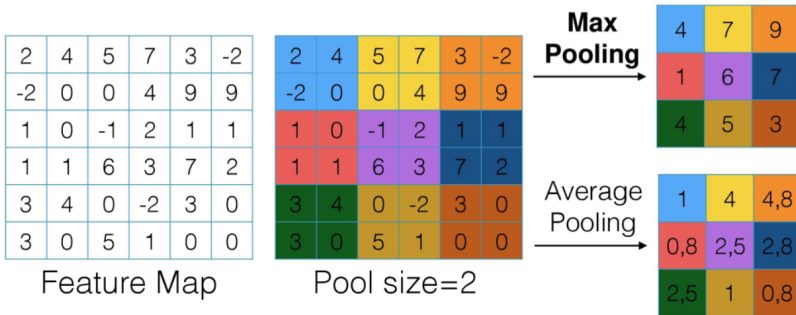
Сверточная часть

А что у нас тут указано, какие еще есть характеристики сверток?

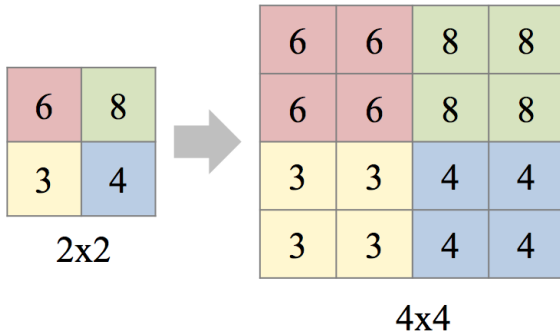


Пулинги

Когда и для чего используем каждый из пулингов?

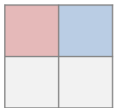


Анпулинг

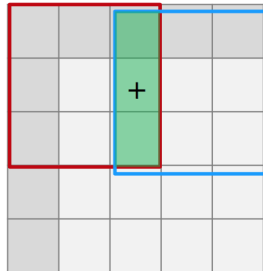


ConvTranspose

Input: 2x2



Input gives
weight for
filter



Stride: 2

Output: 4x4

Обучение и градиентный спуск

Градиентные спуски

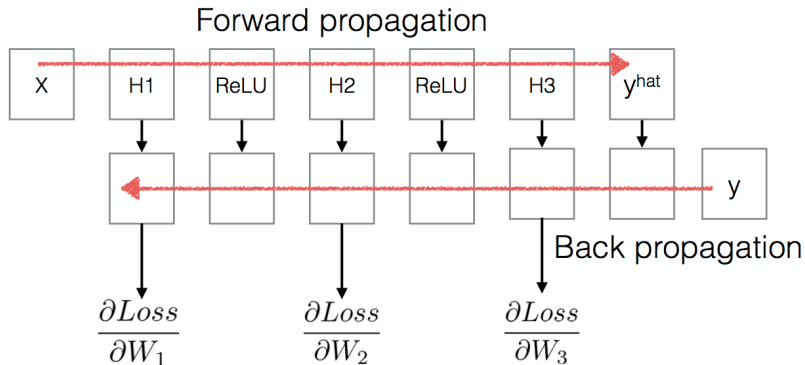
Какой главный гиперпараметр во всех градиентных спусках, которые мы используем?

Вспоминаем особенности кажлого градиентного спуска - ничто не будет забыто! (week2)

1. SGD
2. Momentum SGD
3. RMSprop
4. Adam

Градиентные спуски

В чем прелесть обратного распространения ошибки? Чем он глобально лучше прямого?



и еще эвристик

WEEK4

Архитектуры

Архитектуры

Мы рассмотрели не так много архитектур, но ключевые:

1. Inception
2. Resnet
3. YOLO
4. Autoencoder
5. GAN
6. Unet
7. W2V(к этому классу можно отнести все подобные эмбединги)

Что еще?

Наша мощь

Что мы еще умеем?

1. Задавать модели последовательно или нет
2. Использовать разные метрики и каллбеки
3. Доставать разные кусочки моделей и применять их
4. Поучились чуток дебажить по кускам наш код

Что научимся

И что нам еще осталось

1. Работать с последовательностями
2. Обсудить seq2seq модели
3. Несколько нейронных ячеек - attention, RNN,LSTM,GRU
4. BERT и иже с ним модность
5. Поговорим немного об обучении с подкреплением