

Нейронные сети в машинном обучении

Майер Алексей

Содержание

Лекция 1. Бэкроп, оптимизаторы, PyTorch	2
Лектор(-ы)	2
План курса	2
Лекции и семинары	2
Домашки и система оценивания	2
Причём тут нейроны?	2
История	2
Чуть-чуть матеши	2
Какая аналогия с Нейронами?	3
Как это можно представить?	3
Виды функций активаций	3
Почему сети?	3
Теория. Теорема Цыбенко.	3
Как оценить качество имеющегося решения?	3
Лекция 2.	3

Лекция 1. Бэкроп, оптимизаторы, PyTorch

Лектор(-ы)

Батраков Юрий @BatrakovYuri

Евдокимов Егор @ea_evdokimov

План курса

Задумывался чисто про глубокое обучение. Но пришлось добавить ещё и нейронные сети.

Лекции и семинары

1. Бэкроп, оптимизаторы, PyTorch
- 2.

Домашки и система оценивания

1. Базовые функции pytorch, классификация/регрессия, методы регуляризации (15)
2. Сегментация. Ускорение модели. (10)
3. NLP. Классификация текстов. Переводчик. NER. (15)
4. GAN + VAE. (15)
5. Representation learning. (10)
6. RecSys. ALS. (10)

+ Экзамен (40)

Итог: Сумма баллов (**максимум 115**) делённая на 10. Округление математическое.

Сумма баллов за домашки

Причём тут нейроны?

История

Нейроны – штука, которая накапливает сигнал, при активации передаёт дальше. Нейронов очень много, (>80 млрд.)

Как это смоделировать?

Изначально хотели для использования машинного обучения, кодировать слова числами (просто прогнать словарь), но это оказалось неэффективно.

Давайте моделировать нейрон! Используем для этого:

Чуть-чуть матеши

(x_1, x_2, \dots, x_n) – векторное представление объекта, n – количество признаков.

(w_1, w_2, \dots, w_n) – коэффициенты

b – смещение (bias) («байас» по русски)

Легче представлять, что векторы $(1, x_1, \dots, x_n)$ – для удобства умножения f – какая-то функция (функция активации).

Какая функция? Например больше – меньше. Если больше какой-то константы, то активируется (пороговая функция активации).

Какая f функция? Нелинейная, кусочно дифференцируемая.

$$a = f\left(\sum_{j=1}^n w_{i,j}x_j + b\right)$$

——— Какая аналогия с Нейронами? ———

——— Как это можно представить? ———

Теперь у нас m выходов. Тогда выходы можно представить как матричное произведение матрицы весов на вектор.

$$W = \begin{pmatrix} W_{0,1} & \dots & W_{0,m} \\ \dots & \dots & \dots \\ W_{d,1} & \dots & W_{d,m} \end{pmatrix} \in \mathbb{R}^{d+1}$$

x

——— Виды функций активаций ———

ACTIVATION FUNCTION	PLOT	EQUATION	DERIATIVE	RANGE	ЗАЧЕМ
Линейная		$f(x) = x$	$f'(x) = 1$	$(-\infty, \infty)$	
Бинарный шаг					Для классификации и удобства
Гиперболический тангенс					Для

——— Почему сети? ———

В сетях не один, а гораздо больше слоёв.

——— Теория. Теорема Цыбенко. ———

Формальное изложение с википедии.

——— Как оценить качество имеющегося решения? ———

Вводится несколько функций. Считаем ошибку. Вставить слайды из лекции.

——— Лекция 2. ———

Два линейных слоя подряд это плохо, так как их можно объединить в один слой (умножения на 2 матрицы последовательно можно переписать в одну матрицу). Но иногда это может быть полезно, чтобы сузить размер промежуточной матрицы, чтобы она быстрее передавалась по информационному каналу. Бред, но не всегда автор – дурак.