

Коллоквиум письменный, 2 вопроса, час времени.

1. Построение градиентного спуска с помощью ряда Тейлора.
2. Стохастический градиентный спуск. Версия с инерцией.
3. Построение метода Ньютона с помощью ряда Тейлора.
4. Покоординатный спуск.
5. Бустинг для задачи регрессии, связь с градиентом функции потерь.
6. Бустинг для классификации. Алгоритм AdaBoost (без доказательства формулы для весов отдельных деревьев) и используемая верхняя оценка пороговой функции.
7. Градиентный бустинг для произвольной функции потерь.
8. Нейронная сеть без скрытых слоев, вывод формулы производной функции потерь по одному из параметров.
9. Нейронная сеть с одним скрытым слоем, вывод формулы производной функции потерь по одному из параметров первого слоя. Идея алгоритма back-propagation.
10. Прямой и обратный проход для MLP с линейной активацией (матричные формулы) для одного примера
11. Прямой и обратный проход для MLP с линейной активацией (матричные формулы) для мини-батча (без производной по  $X$ )
12. Инициализация весов в нейронных сетях. Можно ли нулями и почему. Из каких соображений лучше выбирать масштаб случайных весов.
13. Сверточный слой для изображений. Примеры сверток. Как работает back-propagation для сверточного слоя.
14. Что такое receptive field (поле обзора нейрона). Объясните какими способами его можно увеличить в нейронной сети (два способа). Как работает back-propagation для слоя max pooling.
15. Батч нормализация и дропаут
16. Устройство простой рекуррентной сети (две формулы с объяснением мотивации). Производная скрытого состояния  $s_2$  по  $W$ . Применение для задачи language modeling.
17. Производная скрытого состояния  $s_3$  по  $W$ . Общий вид производной для  $s_k$ .
18. Проблема затухающих или взрывающихся градиентов (показать на примере производной для  $s_k$  по  $W$ ). Что делать в случае затухания градиентов и что делать в случае взрывающихся градиентов?
19. Transfer learning и fine-tuning. Когда нужно использовать.
20. Варианты реализации un-pooling слоев (по соседям, максимальные индексы, transpose convolution) и мотивация
21. Примеры рекуррентных архитектур для разных задач

Примеры задач:

1. Посчитать для простой (3x3) картинки операцию свертки.
2. Взять производную сложной функции (например,  $d[(xy + e^{\{xy\}})^y] / dy$ )