

1. Сформулируйте теорему Нестерова.
2. В чем главный недостаток алгоритма AdaGrad? Предложите один из способов решения.
3. Как подбирается адаптивный learning rate в алгоритме AdaDelta?
4. Опишите алгоритм Adam.
5. Напишите идею стратегии Curriculum Learning и приведите пример использования ее на какой-нибудь задаче.