

1. Все те же самые что и в полносвязных, эксперименты никто не запрещал
2. Обычный ГС за шаг рассчитывает градиент используя весь batch, стохастический (т.е случайный) использует 1 рандомный элемент. Часто используют mini-batch ГС, когда задаёшь размер мини-пачки и тренировочная выборка случайно дробится на эти мини-пачки
3. Это метод обучения, когда мы "поощряем" определённое поведение сети, для того, чтобы она стремилась к нужному нам результату. От обучения с учителем это отличается тем, что нет конкретных пар входные данные-ответ, есть некая среда и система вознаграждений. Модель обучается получая при каких-то действиях в среде вознаграждение. Часто такое обучение юзают чтобы научить НС играть в какую-либо игру
4. Ок
5. `input_dim` - размерность вектора входных данных, `units` - размерность вектора выходных данных
6. Это алгоритм ГС который умеет динамически настраивать для каждого признака скорость обучения, тем самым редкие данные получают увеличение  $lr$ , более частые - уменьшение  $lr$ .
7. Нормализация входных картинок путём деления на максимальное значение пикселя (т.е 255), чтобы сеть быстрее обучалась