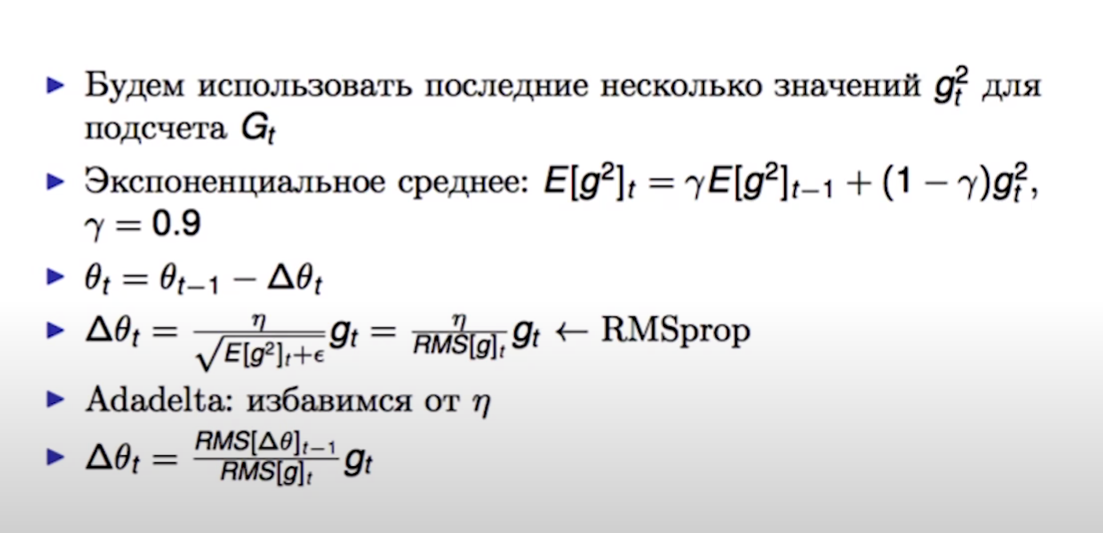
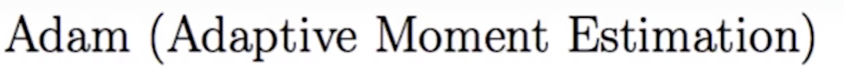


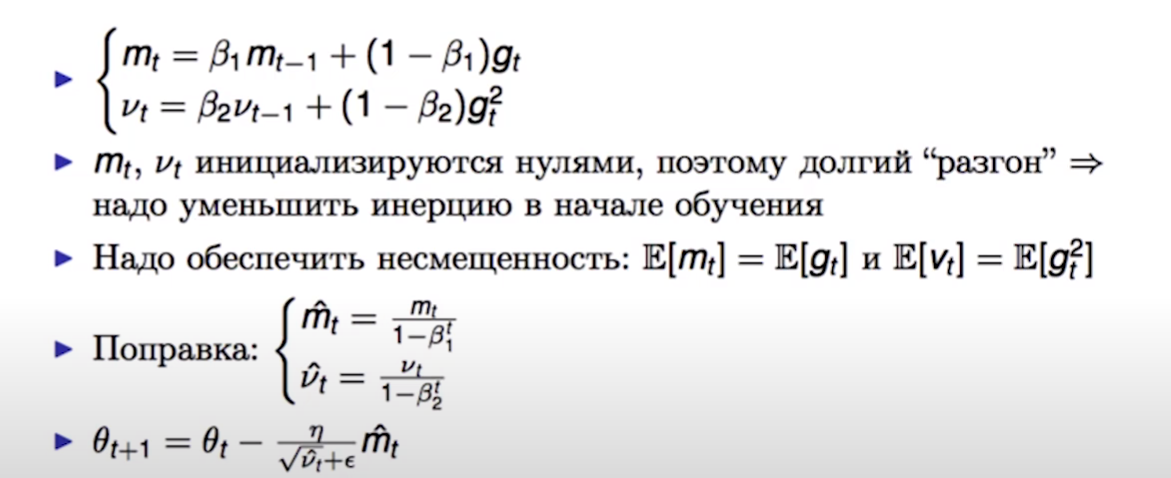
Меняем lr таким образом, что если суммарно много раз произошло обновление, то темп обучения снижается

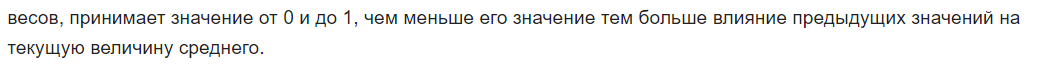
RMSprop и Adadelta



Эта – усредненное изменение тета. Мы считаем сколько за последнюю итерацию в среднем сместились по тета и говорим, что это примерно равно нашему темпу обучения







Оценка среднего градиента (эксп скольз ср)

Оценка квадрата градиента (эксп скольз ср)

Бета – скорость затухания