

## Ответы к л/р 4

### 1. В чём разница между методами Adam и Nadam?

Метод Adam это по сути RMSProp с моментом, а Nadam — RMSProp с моментом Нестерова. Момент Нестерова является простым изменением обычного момента, они помогают ускорить процесс обучения. У момента Нестерова градиент не вычисляется от текущей позиции  $\theta_t$  в *параметрическом пространстве*, а от позиции  $\theta_{intermediate} = \theta_t + \mu v_t$ . Это помогает, поскольку, не смотря на то, что градиент всегда направлен в верном направлении, момент может быть направлен неверно. Если момент направлен в неверном направлении, или проскакивает, градиент все равно может "вернуться" и исправить это в том же шаге обновления.

### 2. Что такое полностью свёрточная сеть?

Устройство свёрточной нейронной сети состоит в чередовании слоёв: полносвязных, свёрточных, субдискретизирующих. Подобная каскадная модель завершается, как правило, полносвязными нейронными слоями. А в полностью свёрточной сети полносвязный нейронный слой преобразуется в свёртку, так называемый цифровой фильтр. Для этого пересчитываются веса полносвязного слоя в матрицу свёртки. При этом для эквивалентного свёрточного слоя необходимо установить размерность фильтра равной размерности карты признаков, а количество фильтров должно равняться количеству нейронов данного слоя.