

### Билеты:

1. Динамические графы вычислений, вычисление градиента в графах вычислений. [\[ссылка\]](#)
2. Dropout: мотивация метода, описание метода, практические аспекты (подбор темпа выключения нейронов, применение Dropout к тестовой выборке). Dropconnect. [\[статья 1\]](#), [\[статья 2\]](#)
3. Батч нормализация: covariate shift, описание метода, поведение градиента при увеличении весов в  $K$  раз, применение батч нормализации на тестовой выборке, недостатки метода (нарушение предположения независимости). Аналоги: layer normalization, weight normalization [\[статья\]](#), [\[ссылка\]](#)
4. Методы оптимизации: градиентный спуск. Выпуклые функции, оптимальное значение темпа обучения для функций из  $C^{1,1}_L$  (непрерывно дифференцируемые с Липшицевой производной) [\[ссылка\]](#)
5. Методы оптимизации: применение импульса для ускорения сходимости. Методы momentum и Nesterov Accelerated Gradient (NAG). [\[ссылка\]](#)
6. Методы оптимизации: адаптивный подбор темпа обучения. Мотивация. Описание методов AdaGrad, RMSprop, AdaDelta, ADAM. Причина использования нормировки  $(1-\beta^t)$  в методе ADAM [\[ссылка\]](#)

### Теоретический минимум:

1. Граф вычислений для batchnorm
2. Описание dropout
3. Градиентный спуск
4. AdaGrad
5. RMSprop
6. AdaDelta
7. Adam

### Домашняя работа:

1. Улучшение обучения нейронной сети для классификации рукописных цифр: реализация на PyTorch. Добавление dropout или batchnorm. Исследование методов оптимизации