

Neural Tangent Kernel

1. Как связан градиент нейронной сети в точке инициализации с ядровым трюком (и Neural Tangent Kernel)?
2. Каким дифференциальным уравнением описывается изменение предсказания сети $y(w)$ по мере обучения для квадратичной ошибки (градиентный поток, gradient flow)? Как выглядит решение этого дифференциального уравнения?
3. Какие гиперпараметры полносвязной сети определяют Neural Tangent Kernel для бесконечно широкой сети после инициализации? Как меняется NTK бесконечно широкой сети по мере обучения (в рамках предположений статьи)?

Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains

1. Какую задачу рассматривали авторы работы о признаках Фурье (Fourier features)? Какие варианты признаков были ими рассмотрены и какие из них показали лучшие результаты?
2. Какой возникающий при обучении полносвязных сетей эффект называется Spectral bias?
3. В чем особенность спектрального смещения при обучении нейронной сети?
4. Какой гиперпараметр для признаков Фурье авторы подбирают при обучении? Как он влияет на поведение модели?

Transformers are RNNs

1. Какова асимптотическая сложность (по памяти и времени вычисления) обычного трансформера в зависимости от длины последовательности токенов n ? Какова асимптотическая сложность gnn в такой-же ситуации? Какова асимптотическая сложность трансформера из статьи?
2. Выписать формулу attention стандартного softmax трансформера
3. Выписать формулу attention произвольного трансформера через функцию похожести sim.
4. Выписать формулу attention произвольного трансформера, через функцию ϕ добавляющую новые признаки соответствующие ядру трансформера.

Hopfield Layer is All You Need

1. Что из себя представляет классическая сеть Хопфилда, какую задачу она решает и каким образом?
2. Какие проблемы существуют у классической сети Хопфилда и как их решает современная версия?
3. Какую аналогию проводят авторы между правилом обновления состояний непрерывной сети Хопфилда и механизмом внимания?