

Neural Tangent Kernel

1. Как авторы Neural Tangent Kernel определяют функцию сети и функцию реализации. В чем заключаются различия между этими понятиями?
2. Определите Neural Tangent Kernel. Расшифруйте все обозначения.
3. Выпишите дифференциальное уравнение, характеризующее динамику обучения сети в пространстве функций.
4. 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. Опишите задачу регрессии картинки по координатам (вход/выход сети). Какую проблему предлагается решать с помощью признаков Фурье?
2. Напишите приближенную формулу для целевой переменной на тестовой выборке через NTK. Выведите из нее скорость убывания абсолютной ошибки на обучающей через спектр матрицы NTK.
3. Зачем нужны случайные признаки Фурье вместо равномерной сетки частот? Как меняется модель при варьировании дисперсии используемых распределений?

Transformers are RNNs: Fast Autoregressive Transformers with Linear Attention

1. Основная идея Transformers are RNNs: за счет чего происходит переход от квадратичной сложности к линейной по памяти и времени в слое Self-attention?
2. Как можно переписать вычисление слоя внимания с учетом causal masking? Как добиться линейной сложности в случае авторегрессионного трансформера?
3. Как представить авторегрессионный трансформер как рекуррентную нейронную сеть? Что будет соответствовать скрытым состояниям?

Hopfield Networks is all you need

1. Выпишите предложенную функцию энергии для сети Хопфилда с непрерывными состояниями и правило обновления состояний для нее.
2. Какие применения для слоя Хопфилда предлагают авторы статьи?
3. Какие состояния сети Хопфилда авторы называют метастабильными состояниями (metastable state), как и чем они отличаются и какие распределения представляют?