

Imagenet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness

1. В чем заключается основная гипотеза авторов “Imagenet-trained CNNs are biased towards texture”? Какие эксперименты они поставили чтобы её проверить? Какие результаты получили?
2. Какое решение предложили авторы “Imagenet-trained CNNs are biased towards texture” чтобы бороться с найденным эффектом? Какие свойства этого решения исследовали авторы?

Deep Double Descent: Where Bigger Models and More Data Hurt

1. Определите метрику *EMC*, предложенную в рамках исследования гипотезы двойного спуска. Что она должна отражать? От чего она зависит и почему её может быть трудно вычислить на практике?
2. Какие три режима обучения выделяет гипотеза двойного спуска? Как она связывает эти режимы обучения с метрикой *EMC*?

Multiplying Matrices Without Multiplying

1. Как работают “традиционные” подходы к ускорению матричного умножения? Как изменится асимптотическая сложность алгоритма умножения матриц при применении такого метода (предполагая, что умножение на вспомогательные матрицы не ускорено и производится по определению).
2. Опишите алгоритм Product Quantization для ускоренного умножения матриц. Какие величины он вычисляет заранее? Как вычисляет результат для приходящего вектора \mathbf{b} ?

Dataset Distillation

1. Что такое дистилляция данных? Какую задачу оптимизации предлагают решать авторы для дистилляции данных? Выпишите постановку для одного шага градиентного спуска и поясните обозначения.
2. Как отличались результаты дистилляции данных в случае фиксированной и случайной инициализации модели? В каком случае было выше качество? Что можно сказать об интерпретируемости данных?

The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks

1. В чем заключается гипотеза лотерейного билета?
2. Опишите алгоритм разреживания сетей, предложенный на основе гипотезы лотерейного билета.

Does Knowledge Distillation Really Work?

1. В чем заключается процедура дистилляции нейронной сети? Для чего она нужна? Какие метрики авторы “Does Knowledge Distillation Really Work?” используются для оценки качества дистилляции?
2. Насколько согласованы оказываются дистиллированные сети в экспериментах “Does Knowledge Distillation Really Work?”? Как ведет себя согласованность учителя и студента на обучающих данных по мере увеличения обучающей выборки?