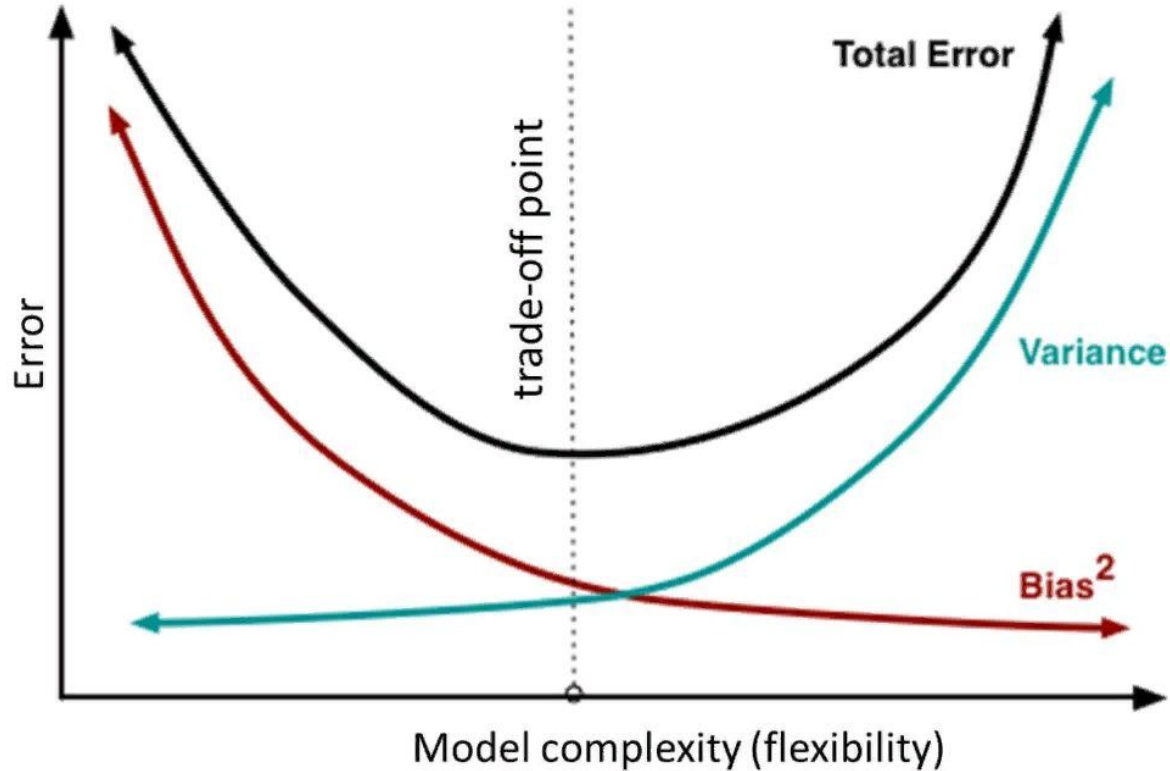


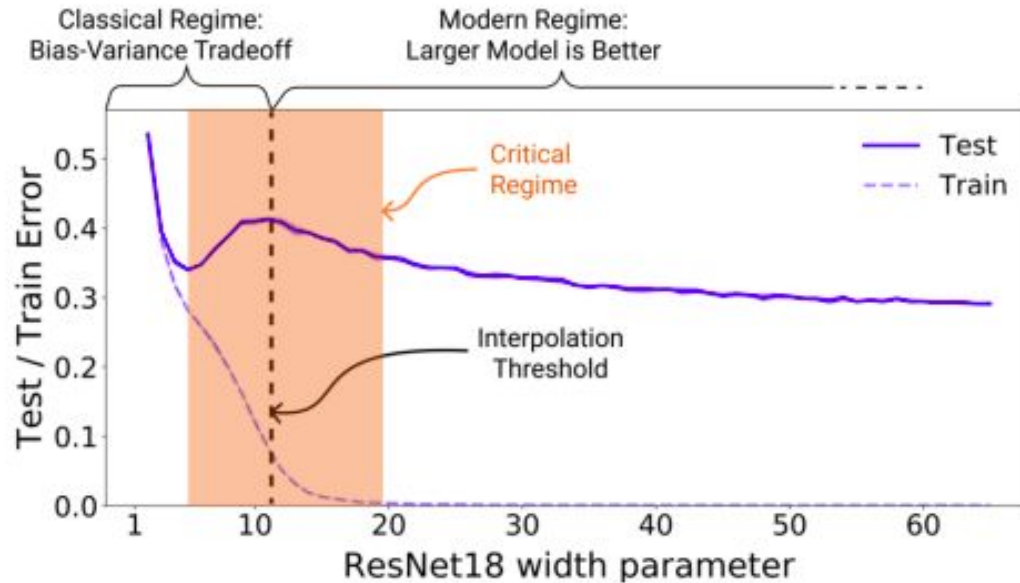
Double descent

Доклад подготовила Тяжова Наталья, студентка БПМИ211

Bias-variance trade-off



Явление двойного спуска



После прохождения порога интерполяции ошибка на тесте, вопреки ожиданиям классического машинного обучения, убывает!

Эффективная сложность модели

Эффективная сложность модели (ЕМС) процедуры обучения τ для распределения D и параметра $\varepsilon > 0$ – это

$$EMC_{D,\varepsilon}(\tau) = \max\{n | E_{S \sim D^n}[Error_S(\tau(S))] \leq \varepsilon\}$$

где $Error_S(M)$ – средняя ошибка модели M на обучающей выборке S .

Обобщённая гипотеза о двойном спуске

Если $EMC_{D,\varepsilon}(\tau) \ll n$, любое изменение процедуры τ , увеличивающее её эффективную сложность, будет уменьшать ошибку на тесте.

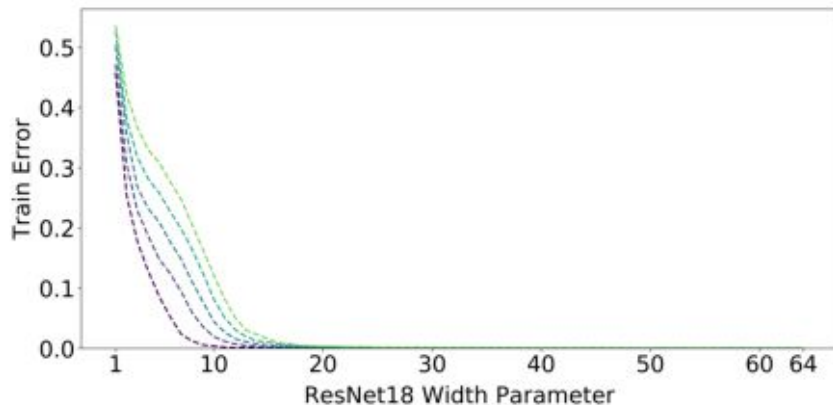
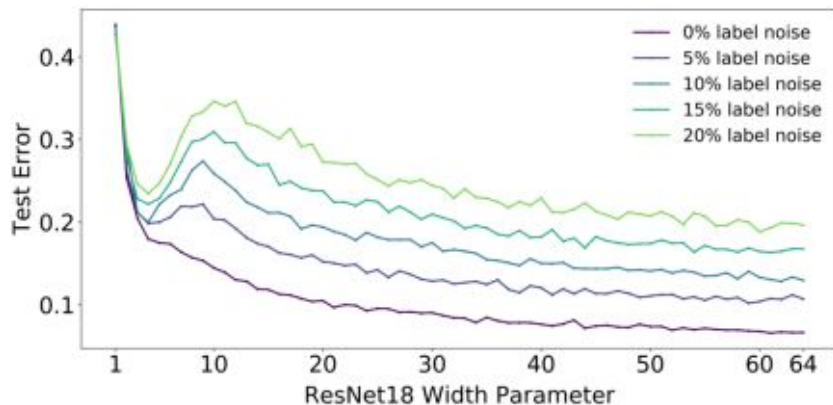
Если $EMC_{D,\varepsilon}(\tau) \gg n$, любое изменение процедуры τ , увеличивающее её эффективную сложность, будет уменьшать ошибку на тесте.

$EMC_{D,\varepsilon}(\tau) \approx n$, любое изменение процедуры τ , увеличивающее её эффективную сложность, будет уменьшать или увеличивать ошибку на тесте.

Недостатки гипотезы

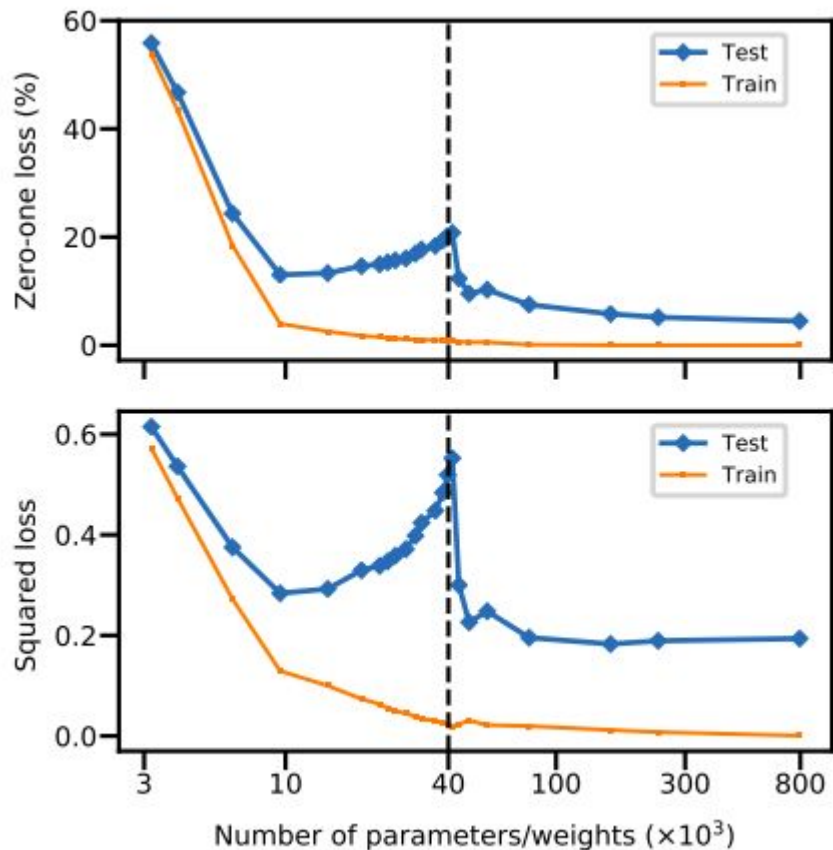
- Как выбрать ϵ ?
- Что значит значительно меньше или значительно больше?
- Мы слабо представляем, отчего и как именно зависит ширина критического интервала, на который и опирается вся наша гипотеза

Model-wise double descent



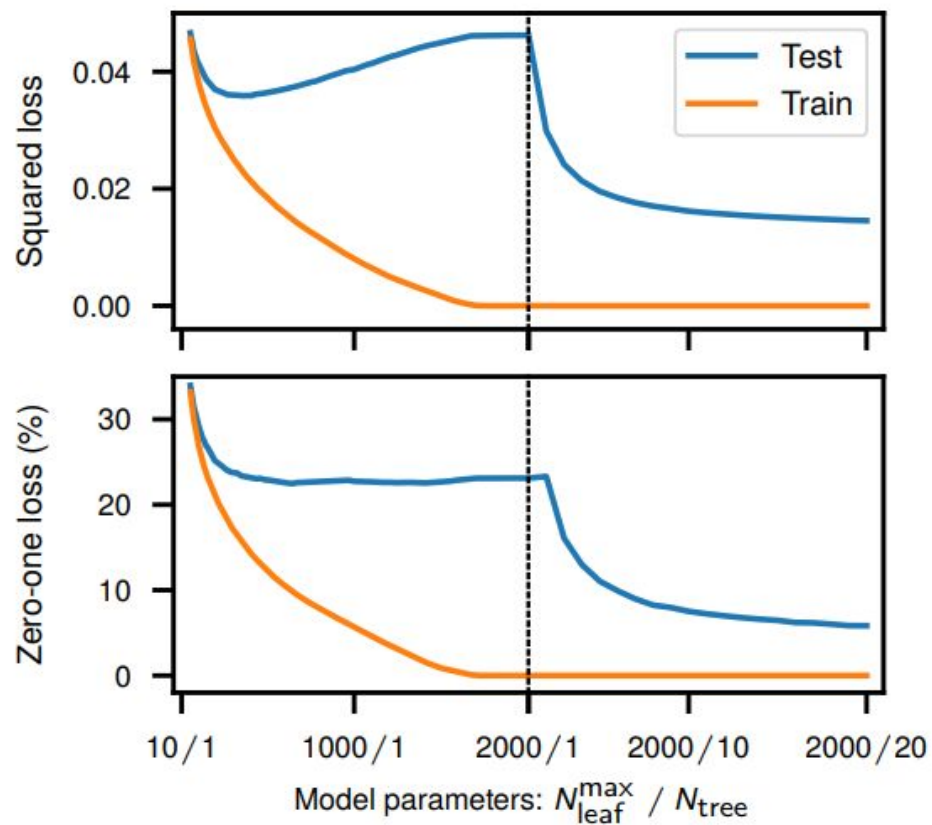
- Пик ошибки на тесте зависит от порога интерполяции
- Добавление шума сдвигает вправо порог интерполяции, а следовательно, и пик ошибки на тесте

Полносвязная нейронная сеть

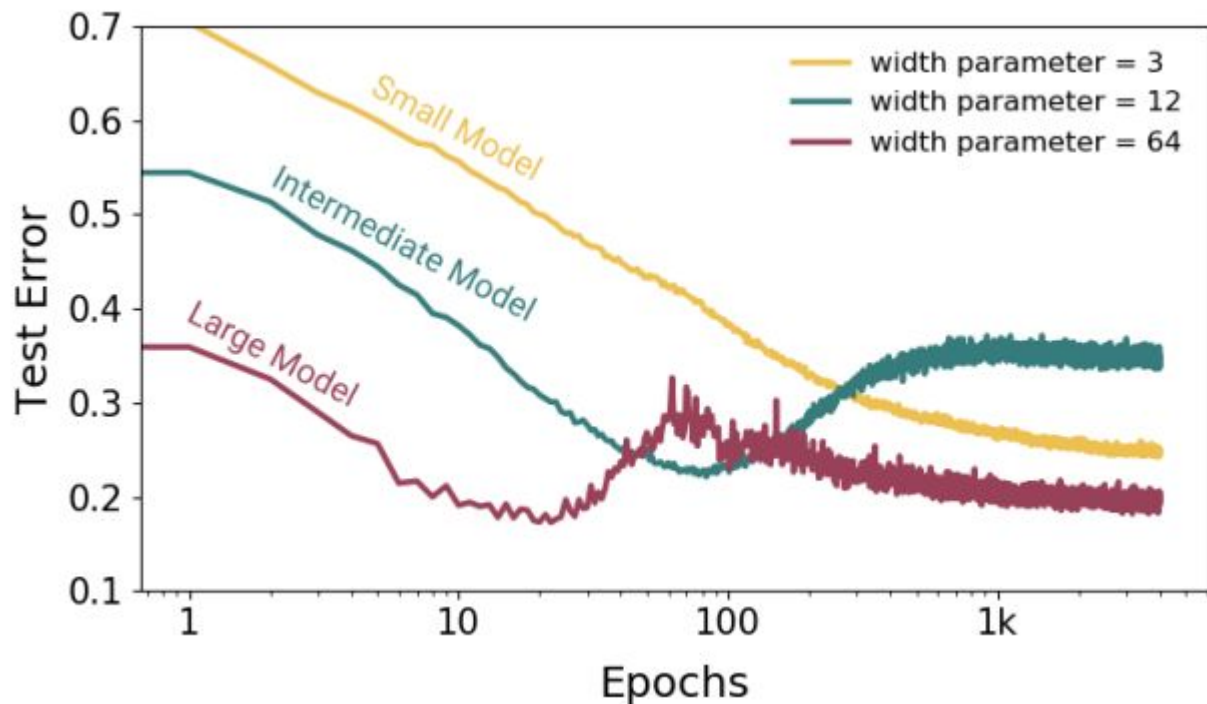


После преодоления порога интерполяции
ошибка на тесте достигает своего
минимума

Случайный лес

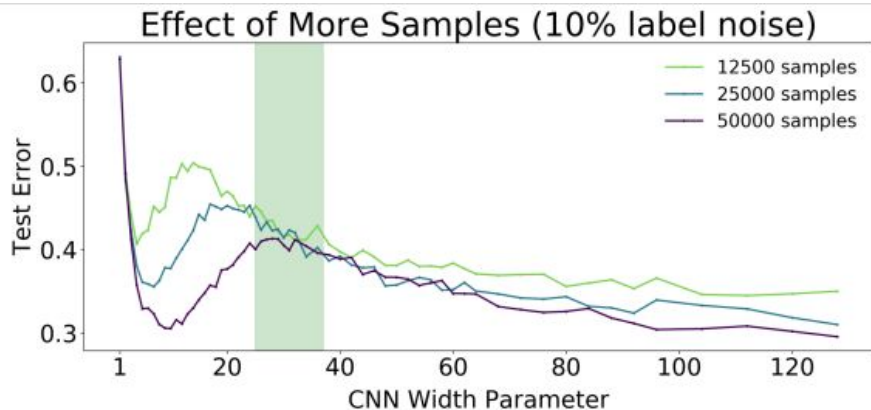


Epoch-wise double descent

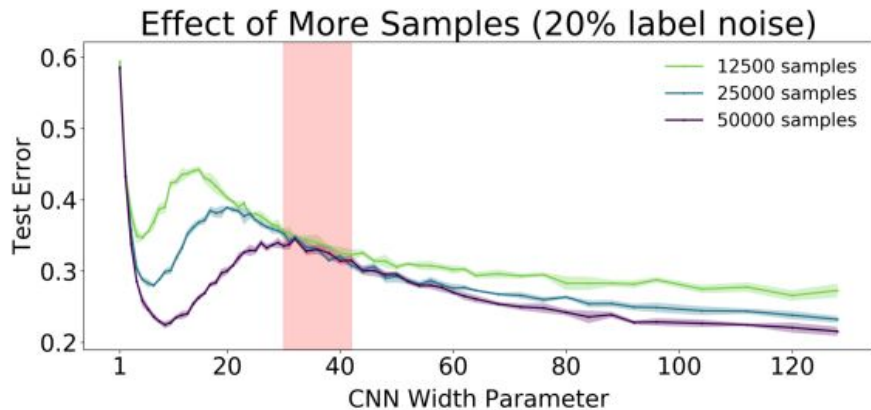


- Маленькие модели остаются в недообученном режиме
- Для средних моделей наблюдается классическая ситуация с U-образной ошибкой на тесте
- Большие модели переходят в режим двойного спуска

Эффект увеличения размера выборки

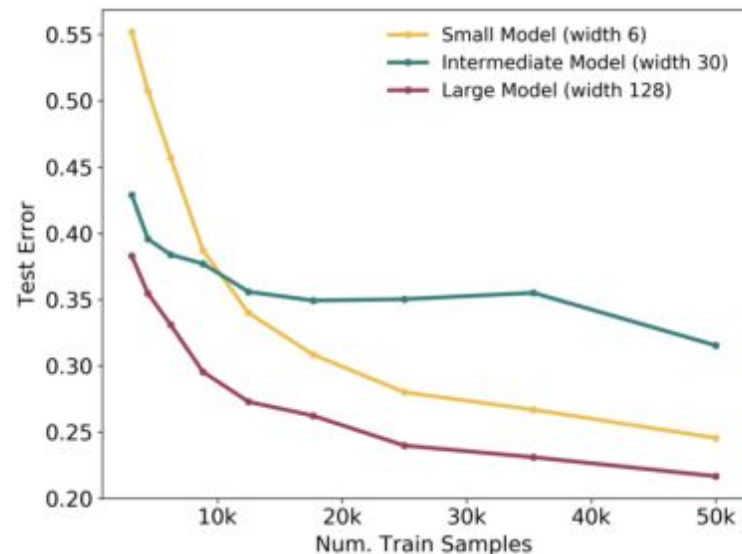
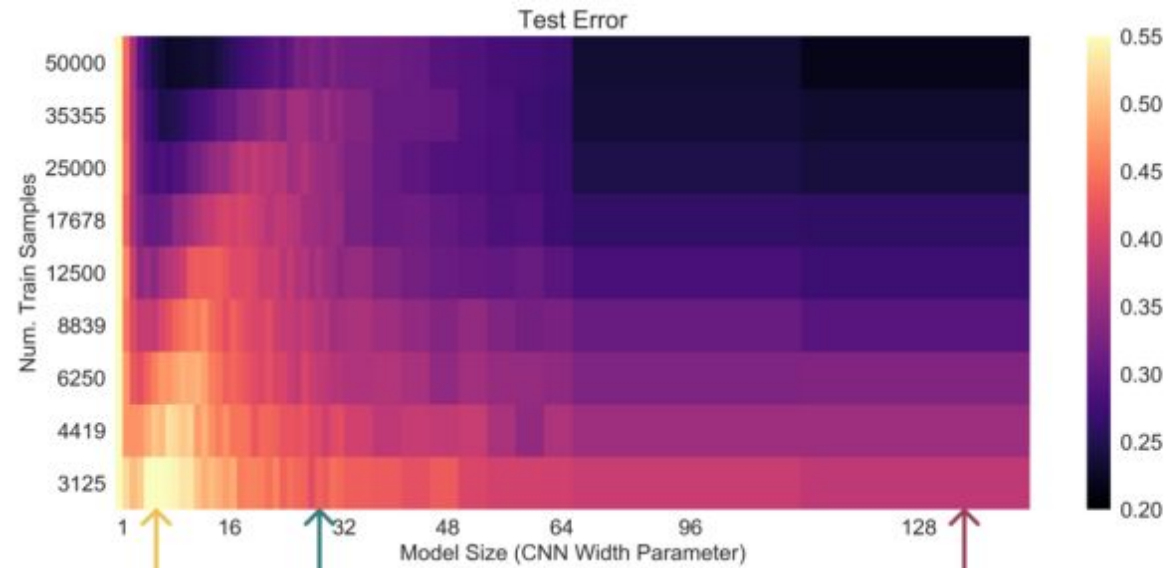


Существует диапазон, где обучение на 2х большем размере выборки не улучшает ошибку на тесте.



Есть диапазон, где обучение на 4х большем размере выборки не улучшает ошибку на тесте.

Эффект увеличения размера выборки



Где и когда возникает эффект двойного спуска?

- В совершенно разных моделях: от алгоритмов классического машинного обучения по типу рандомного леса до самых различных типов нейросетей: свёрточных, полносвязных, трансформеров.
- Зашумление и мисспецификация моделей.
- Эффективная сложность модели гораздо выше размера данных.

Каковы причины возникновения эффекта двойного спуска?

- В классе слишком сложных моделей существует больше моделей, сохраняющих околонулевую ошибку на трейне, и методы оптимизации могут найти ту, которая даёт низкую ошибку на тесте.
- Наличие медленно обучающихся, но информативных признаков в данных.

Заключение

Модели и процедуры обучения демонстрируют нетипичное поведение, когда их эффективная сложность сопоставима с размером данных: усложнение модели может привести как к уменьшению, так и к увеличению ошибки на тесте.

Этот феномен устойчив к выбору набора данных, архитектуры и процедур обучения.

Наконец, двойной спуск может привести к режиму, при котором обучение на большем количестве данных приводит к ухудшению ошибки на тесте.

ИСТОЧНИКИ

Deep double descent: where bigger models and more data hurt

<https://arxiv.org/pdf/1912.02292.pdf>

Reconciling modern machine learning practice and the bias-variance trade-off

<https://arxiv.org/pdf/1812.11118.pdf>

When and how epochwise double descent happens

<https://arxiv.org/pdf/2108.12006.pdf>