

Deep Double Descent

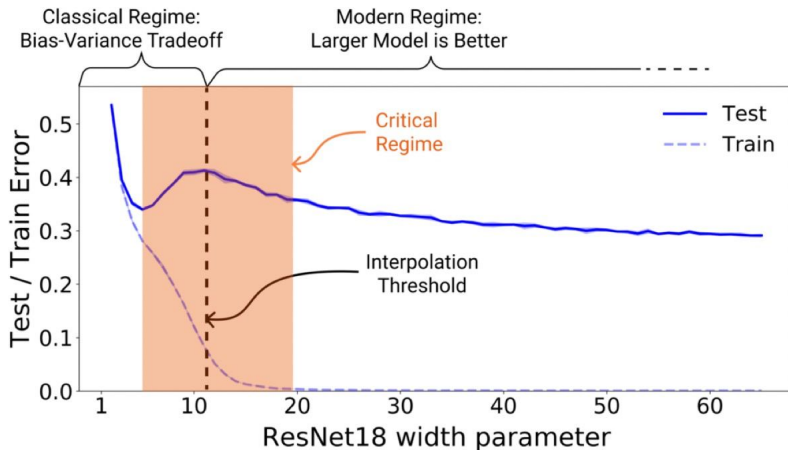
когда бóльшие модели и больше данных не улучшают качество
работы

Настя Городилова

БПМИ191

**Факультет Компьютерных Наук
НИУ ВШЭ**

7 февраля 2022 г.



Гипотеза двойного спуска

Effective model complexity (EMC)

Пусть:

\mathcal{T} - процедура обучения

$S = \{(x_1, y_1), \dots, (x_n, y_n)\}$ - набор данных

$\mathcal{T}(S)$ - классификатор

\mathcal{D} - распределение данных

$\varepsilon > 0$

Тогда:

$$EMC_{\mathcal{D}, \varepsilon}(\mathcal{T}) := \max\{n | \mathbb{E}_{S \sim \mathcal{D}^n} [Error_S(\mathcal{T}(S))] \leq \varepsilon\}$$

Гипотеза двойного спуска

Гипотеза: $\forall \mathcal{D}, \mathcal{T}, \varepsilon > 0$

Если задача: классификация n объектов из \mathcal{D}

Тогда:

■ **Недопараметризованный режим**

$EMC_{\mathcal{D}, \varepsilon}(\mathcal{T}) < n \Rightarrow \forall \text{ возмущение } \mathcal{T} : EMC \uparrow \quad Error(Test) \downarrow$

■ **Перепараметризованный режим**

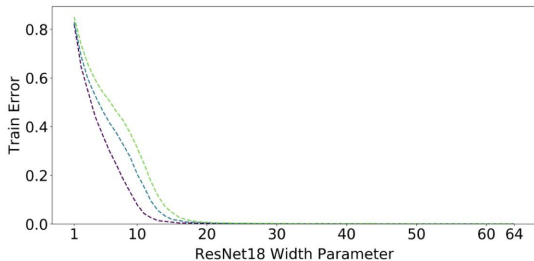
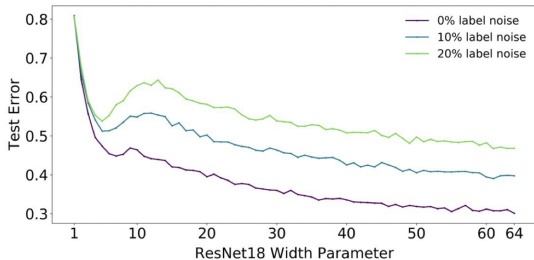
$EMC_{\mathcal{D}, \varepsilon}(\mathcal{T}) > n \Rightarrow \forall \text{ возмущение } \mathcal{T} : EMC \uparrow \quad Error(Test) \downarrow$

■ **Критический режим**

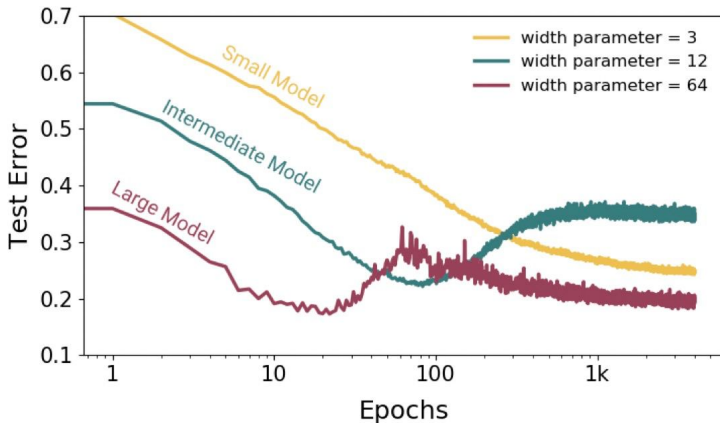
$EMC_{\mathcal{D}, \varepsilon}(\mathcal{T}) \sim n \Rightarrow \forall \text{ возмущение } \mathcal{T} : EMC \uparrow \quad Error(Test) \uparrow \downarrow$

Эвристически $\varepsilon = 0.1$

Двойной спуск, зависящий от ширины модели



Двойной спуск, зависящий от числа эпох

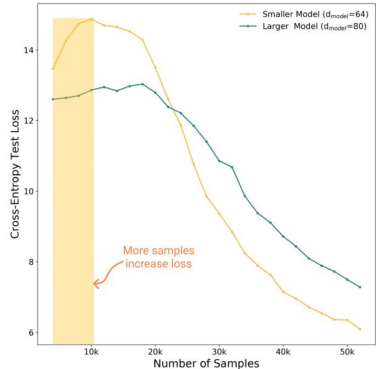
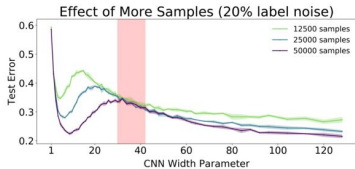
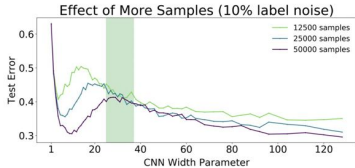


Двойной спуск, зависящий от числа трейна

Раньше: n - фиксировано, EMC - меняется

Сейчас: EMC - фиксировано, n - меняется

Двойной спуск, зависящий от числа трейна



Ссылки на источники

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