

Отчет об экспериментах сравнения апостериорной оценки и реальной ошибки решения, полученного с помощью Deep Ritz

Басалаев Даниил Александрович 5030102/10201

3 мая 2025 г.

Аннотация

В этом отчете представлены результаты экспериментов с вариационной нейронной сетью для решения уравнения:

$$\begin{aligned} -u'' &= f(x) \\ f(x) &= (\pi^2) \sin(\pi x) \\ u(0) &= u(1) = 0 \end{aligned}$$

соответственно решение ищем в интервале $[0, 1]$. Обозначим $U(x)$ - точное решение (или его аппроксимация, но в этой работе я использую точное решение), $V(x)$ - решение, полученное нейросетью. Для нахождения апостериорной оценки использовал следующую формулу:

$$\|U' - V'\|^2 = \int_a^b (U'(x) - V'(x))^2 dx \approx \frac{b-a}{N} \sum_{i=1}^N (U'(x_i) - V'(x_i))^2$$

$$C_\Omega = \frac{b-a}{\pi}$$

$$\|U'' + f(x)\|^2 = \int_a^b (U''(x) + f(x))^2 dx \approx \frac{b-a}{N} \sum_{i=1}^N (U''(x_i) + f(x_i))^2$$

$$M^2 = (1 + \beta) \|U' - V'\|^2 + \left(1 + \frac{1}{\beta}\right) C_\Omega^2 \|U'' + f(x)\|^2$$

Ниже приведены таблицы экспериментов, в которых меняется один параметр.

1 Результаты Экспериментов

1.1 Зависимость от Количества Итераций

batch size	Optimizer	Neurons	Layers	Iterations	Learning Rate	$\ M\ ^2$	$\ U'-V'\ ^2$	$\ M\ ^2 / \ U'-V'\ ^2$	Training Time
128	adam	32	3	1000	0.001	0.0288319	0.014416	2.0000011	0.7992196
128	adam	32	3	2000	0.001	0.0390249	0.0195124	2.0000016	1.1996627
128	adam	32	3	3000	0.001	0.0267927	0.0133963	2.0000016	1.607137
128	adam	32	3	4000	0.001	0.0621919	0.031096	1.9999983	1.95872

Рис. 1: Зависимость апостериорной оценки ошибки и реальной ошибки от количества итераций обучения.

1.2 Зависимость от Количества Слой

batch size	Optimizer	Neurons	Layers	Iterations	Learning Rate	$\ M\ ^2$	$\ U-V\ ^2$	$\ M\ ^2/\ U-V\ ^2$	Training Time
128	adam	32	2	2000	0.001	0.0104334	0.0052167	2.0	0.9911716
128	adam	32	3	2000	0.001	0.0274618	0.0137309	2.000002	1.1831977
128	adam	32	4	2000	0.001	0.0165289	0.0082645	2.000003	1.4395688
128	adam	32	5	2000	0.001	0.0244715	0.0122357	1.9999984	1.6878664

Рис. 2: Зависимость апостериорной оценки ошибки и реальной ошибки от количества слоев в нейронной сети.

1.3 Зависимость от Скорости Обучения

batch size	Optimizer	Neurons	Layers	Iterations	Learning Rate	$\ M\ ^2$	$\ U-V\ ^2$	$\ M\ ^2/\ U-V\ ^2$	Training Time
128	adam	32	3	2000	0.01	0.0978493	0.0489246	2.000009	1.2079573
128	adam	32	3	2000	0.001	0.0558039	0.0279019	2.0000016	1.1849289
128	adam	32	3	2000	0.0001	0.1381064	0.0690532	2.000009	1.1794379
128	adam	32	3	2000	1e-05	3.2754166	1.6377081	2.000003	1.1891747

Рис. 3: Зависимость апостериорной оценки ошибки и реальной ошибки от скорости обучения.

1.4 Зависимость от Количества Нейронов

batch size	Optimizer	Neurons	Layers	Iterations	Learning Rate	$\ M\ ^2$	$\ U-V\ ^2$	$\ M\ ^2/\ U-V\ ^2$	Training Time
128	adam	10	3	2000	0.001	0.0237491	0.0118745	2.0000015	1.3183661
128	adam	16	3	2000	0.001	0.0208345	0.0104173	1.9999979	1.1513414
128	adam	32	3	2000	0.001	0.0315887	0.0157943	1.9999986	1.1933982
128	adam	64	3	2000	0.001	0.0266398	0.0133199	2.0000017	1.3904302

Рис. 4: Зависимость апостериорной оценки ошибки и реальной ошибки от количества нейронов в слое.

1.5 Сравнение Оптимизаторов

batch size	Optimizer	Neurons	Layers	Iterations	Learning Rate	$\ M\ ^2$	$\ U-V\ ^2$	$\ M\ ^2/\ U-V\ ^2$	Training Time
128	adam	32	3	2000	0.001	0.0276072	0.0138036	1.9999979	1.3500774
128	sgd	32	3	2000	0.001	0.1315062	0.0657531	2.0000008	0.7959518
128	rmsprop	32	3	2000	0.001	0.0139145	0.0069572	2.0000025	1.0211978

Рис. 5: Сравнение различных оптимизаторов (Adam, SGD, RMSprop) по апостериорной оценке ошибки и реальной ошибке.

1.6 Сравнение по количеству обучающим точкам

batch size	Optimizer	Neurons	Layers	Iterations	Learning Rate	$\ M\ ^2$	$\ U-V\ ^2$	$\ M\ ^2/\ U-V\ ^2$	Training Time
64	adam	32	3	2000	0.001	0.0809995	0.0404998	1.9999989	1.1954935
128	adam	32	3	2000	0.001	0.0400514	0.0200257	1.999999	1.1713238
256	adam	32	3	2000	0.001	0.008901	0.0044505	1.9999984	1.3083191
512	adam	32	3	2000	0.001	0.023314	0.011657	1.9999981	1.4157133
1024	adam	32	3	2000	0.001	0.0098121	0.0049061	2.0000044	1.8438294

Рис. 6: Сравнение различных размеров тренировочной последовательности по апостериорной оценке ошибки и реальной ошибке.

2 Выводы

Основываясь на проведенных экспериментах, можно сделать следующие выводы:

- Апостериорная оценка слишком слабая (отличие минимум в 30 раз)
- При улучшение модели (более хорошей подборкой гиперпараметров) реально ошибка уменьшается сильнее апостериорной оценки (различие в 300 раз!)
- Мне показалось странным, что апостериорная оценка в основном зависит от нормы разности производных приближённого и точного решения, а в то время как сравниваем с нормой разности просто решения и приближённого решения. Я не очень понял, как можно дать оценку ошибки функции, основывая на производную...
- Есть риски, что я неправильно применил формулу, поэтому указал все формулы, которые я использую, в самом начале документа