

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

Басалаев Даниил Александрович 5030102/10201

28 апреля 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	0.0264545	0.0001157	228.5833723	0.8816056
128	adam	32	3	2000	0.001	0.0098323	2.68e-05	366.5826434	1.2999411
128	adam	32	3	3000	0.001	0.0532998	0.0017308	30.7950313	1.6693759
128	adam	32	3	4000	0.001	0.0167293	0.0005604	29.8506004	1.9675207

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

1.2 Зависимость от Количества Слоёв

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	0.0355768	0.0008272	43.0110369	1.1146262
128	adam	32	3	2000	0.001	0.0222693	0.0002391	93.1549063	1.3479249
128	adam	32	4	2000	0.001	0.0690756	0.0017236	40.075898	1.9446356
128	adam	32	5	2000	0.001	0.0267163	0.0003149	84.8317309	2.1320336

Рис. 2: Зависимость апостериорной оценки ошибки и реальной ошибки от количества слоёв 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	0.1488798	0.0067359	22.1024979	1.306756
128	adam	32	3	2000	0.001	0.0169473	0.0001197	141.5829697	1.2127318
128	adam	32	3	2000	0.0001	0.1203797	0.0012158	99.0161907	1.2006378
128	adam	32	3	2000	1e-05	4.775526	0.2313064	20.6458885	1.2254651

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	0.052682	0.0003745	140.6766474	1.4703515
128	adam	16	3	2000	0.001	0.0397322	0.0014248	27.886264	1.2525628
128	adam	32	3	2000	0.001	0.0210456	0.0001542	136.4957985	1.248086
128	adam	64	3	2000	0.001	0.0525371	0.0006832	76.9021795	1.5692601

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	0.0142475	8.98e-05	158.6081714	1.4879823
128	sgd	32	3	2000	0.001	0.1544056	0.0014825	104.1537148	0.8436258
128	msprop	32	3	2000	0.001	0.0307618	0.0008998	34.1856687	1.2681201

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	0.0224638	0.0002269	98.9884462	1.4406531
128	adam	32	3	2000	0.001	0.0410929	0.0010313	39.8458362	1.3926423
256	adam	32	3	2000	0.001	0.007158	0.0002616	27.3596757	1.3799183
512	adam	32	3	2000	0.001	0.027058	0.0009892	27.3534083	1.673435
1024	adam	32	3	2000	0.001	0.0131484	0.0004548	28.9125735	1.928153

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