



Институт интеллектуальных кибернетических систем

КАФЕДРА КИБЕРНЕТИКИ

БДЗ

по курсу "Теория нейронных сетей"

студента группы Б19-514

Каравашкин Марка

Вариант № 1

Оценка: _____

Подпись: _____

2022 г.



I. Исходные данные

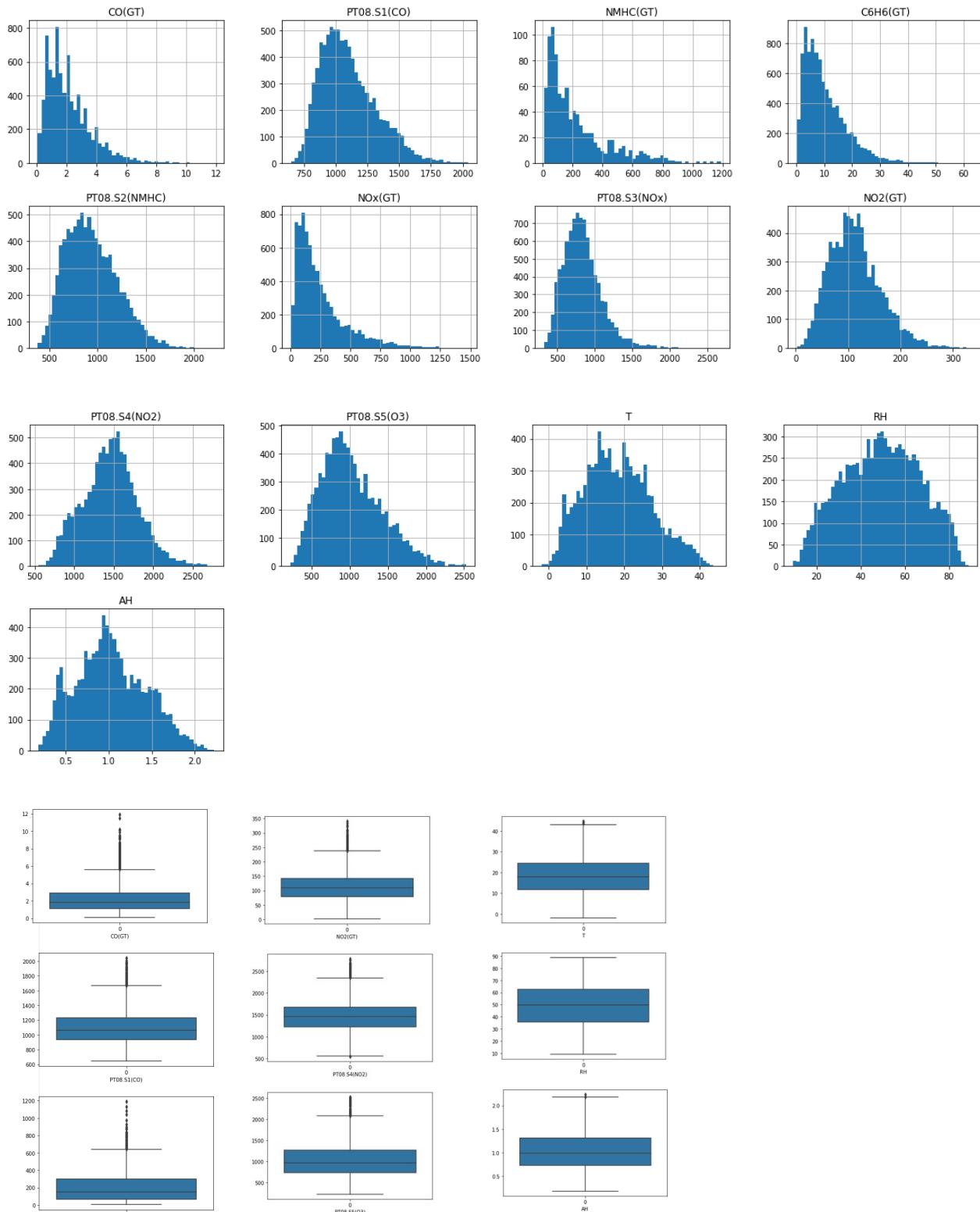
1.1. Описание исходных данных

Источник данных: <http://archive.ics.uci.edu/ml/datasets/air+quality>

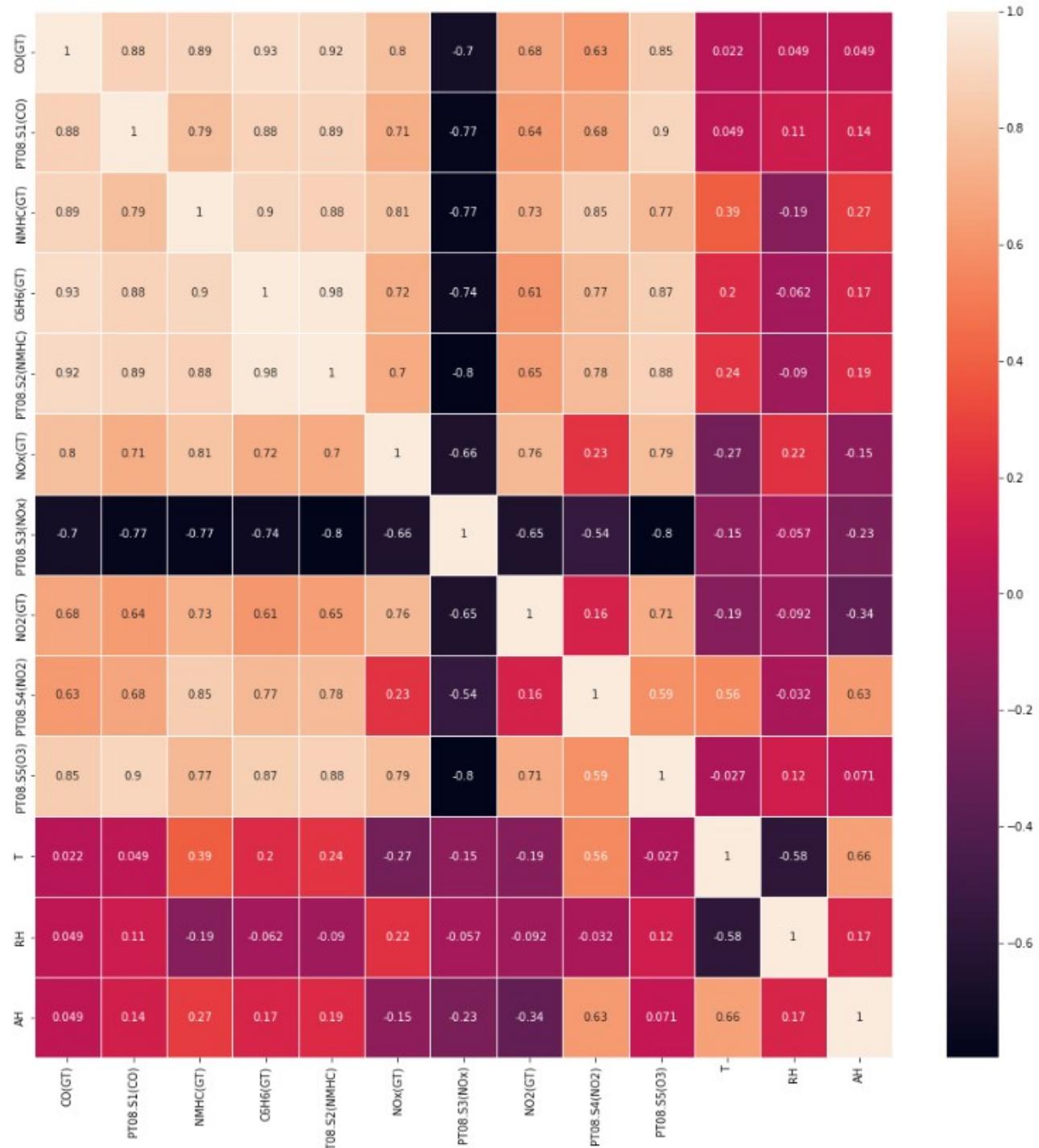
В источнике приведены данные наблюдения за вредными химическими веществами в итальянском городе и фиксация их значений на протяжении одного года (с 2004 по 2005 года).

1.2. Визуальный анализ исходных данных

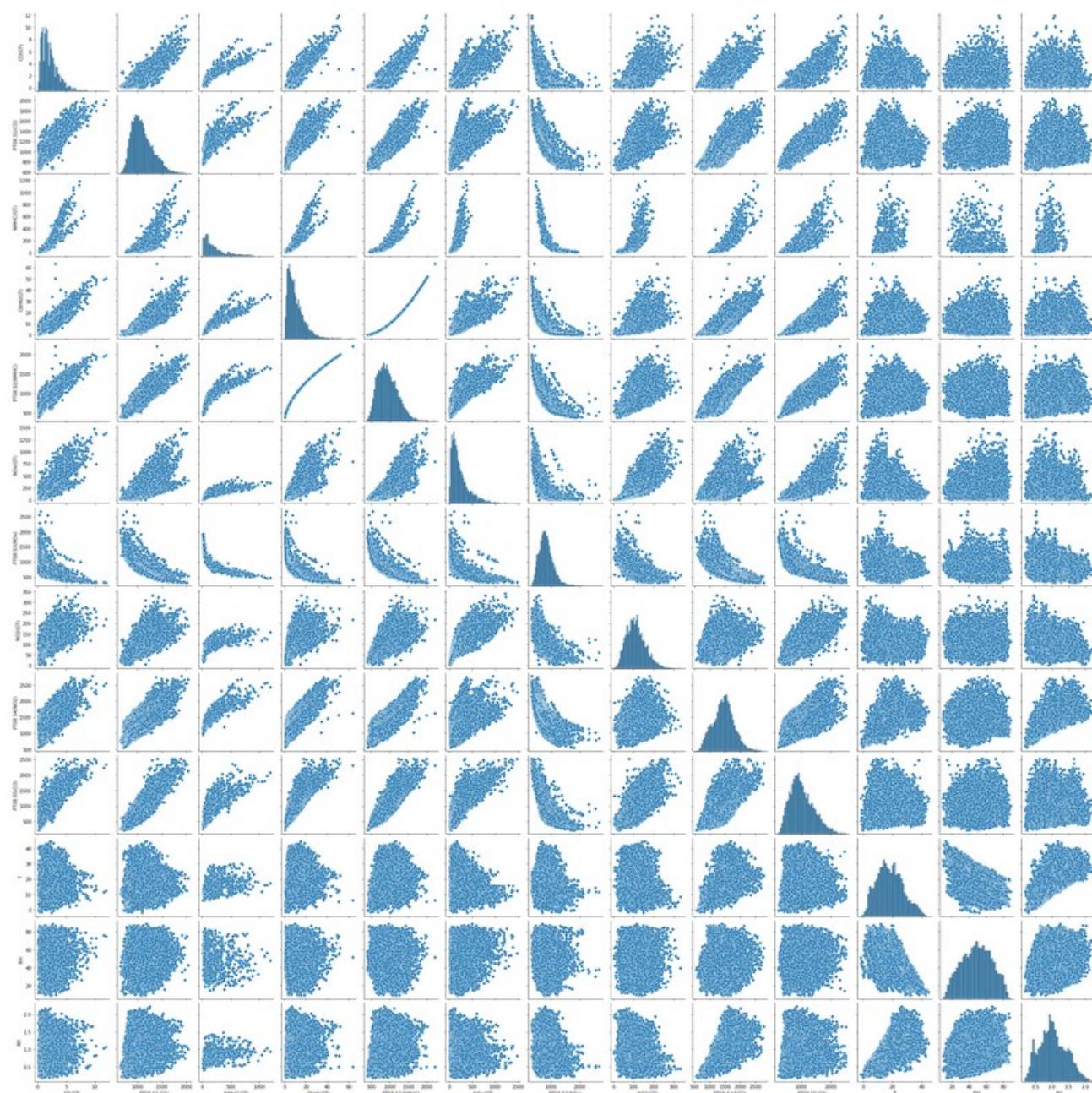
a) Гистограммы распределения и диаграммы Box-and-Whisker



б) Корреляционная матрица признаков



в) Диаграммы рассеяния



1.3. Выводы

По гистограммам можно предположить, что большинство признаков имеют нормальное распределение или распределение хи-квадрат. Имеется большое количество выбросов. Некоторые признаки сильно коррелированы между собой.

II. Предобработка данных

2.1. Очистка данных

a) Обнаружение и устранение дубликатов

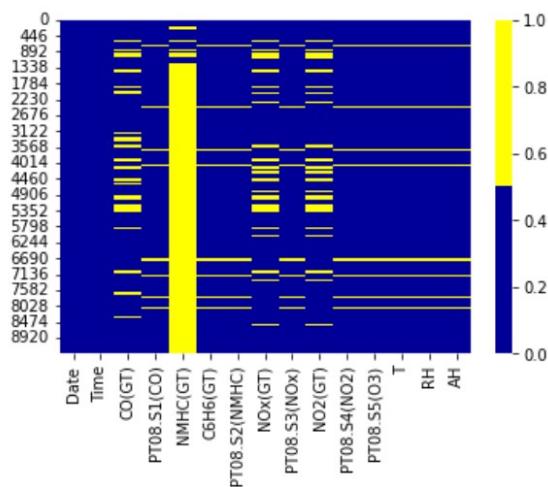
В наборе данных дубликатов нет.

б) Обнаружение и устранение выбросов

Выбросы были обнаружены на диаграммах *Box-and-Whisker*. Были устраниены все выбросы, значения которых больше чем $Q3 + 1.75 * (Q3 - Q1)$, где $Q1$ и $Q3$ – первый и четвертый выборочный квартиль соответственно. Значения в этих выбросах были изменены на случайное число в диапазоне ($Q1; Q3$).

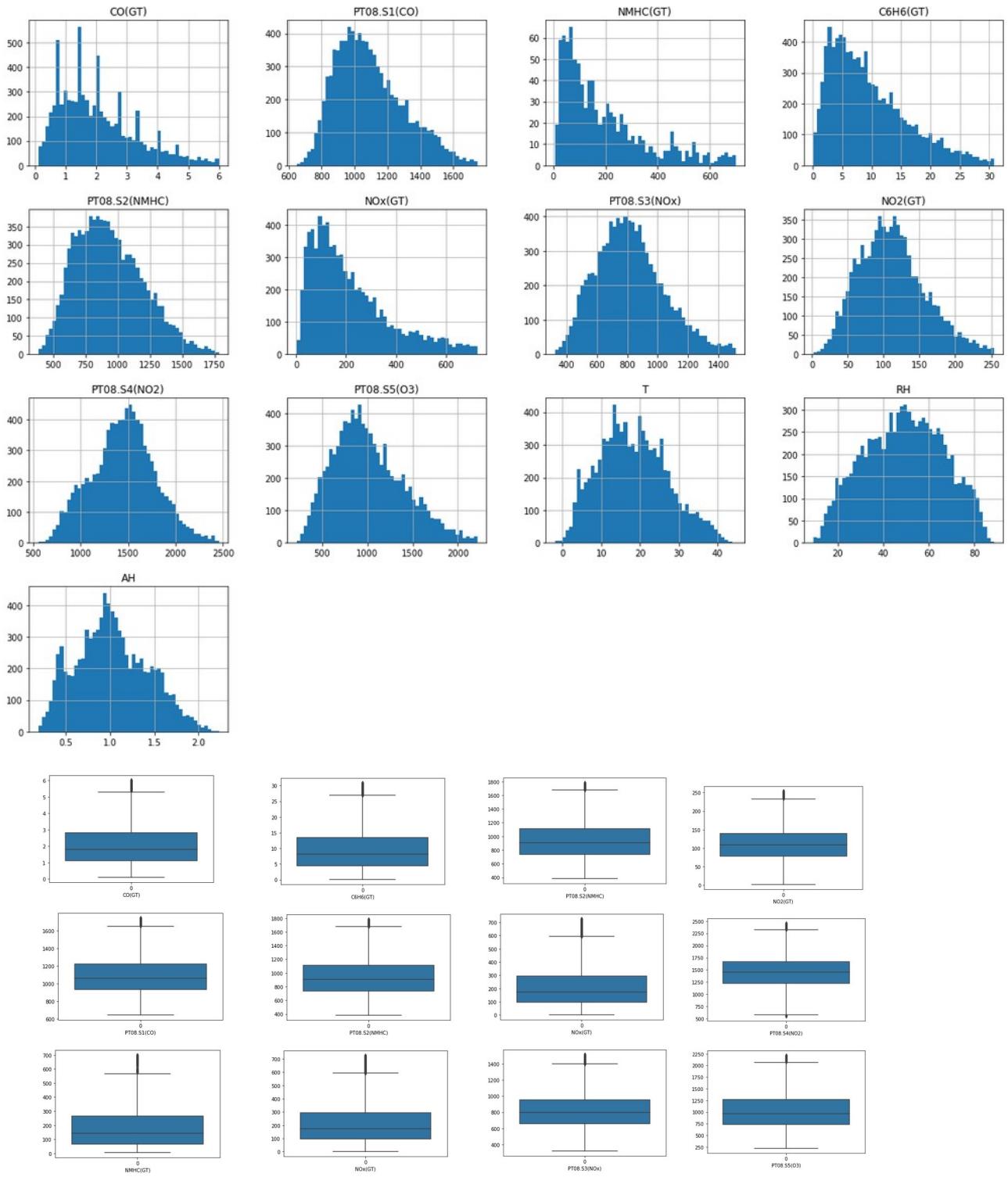
в) Пропущенные значения

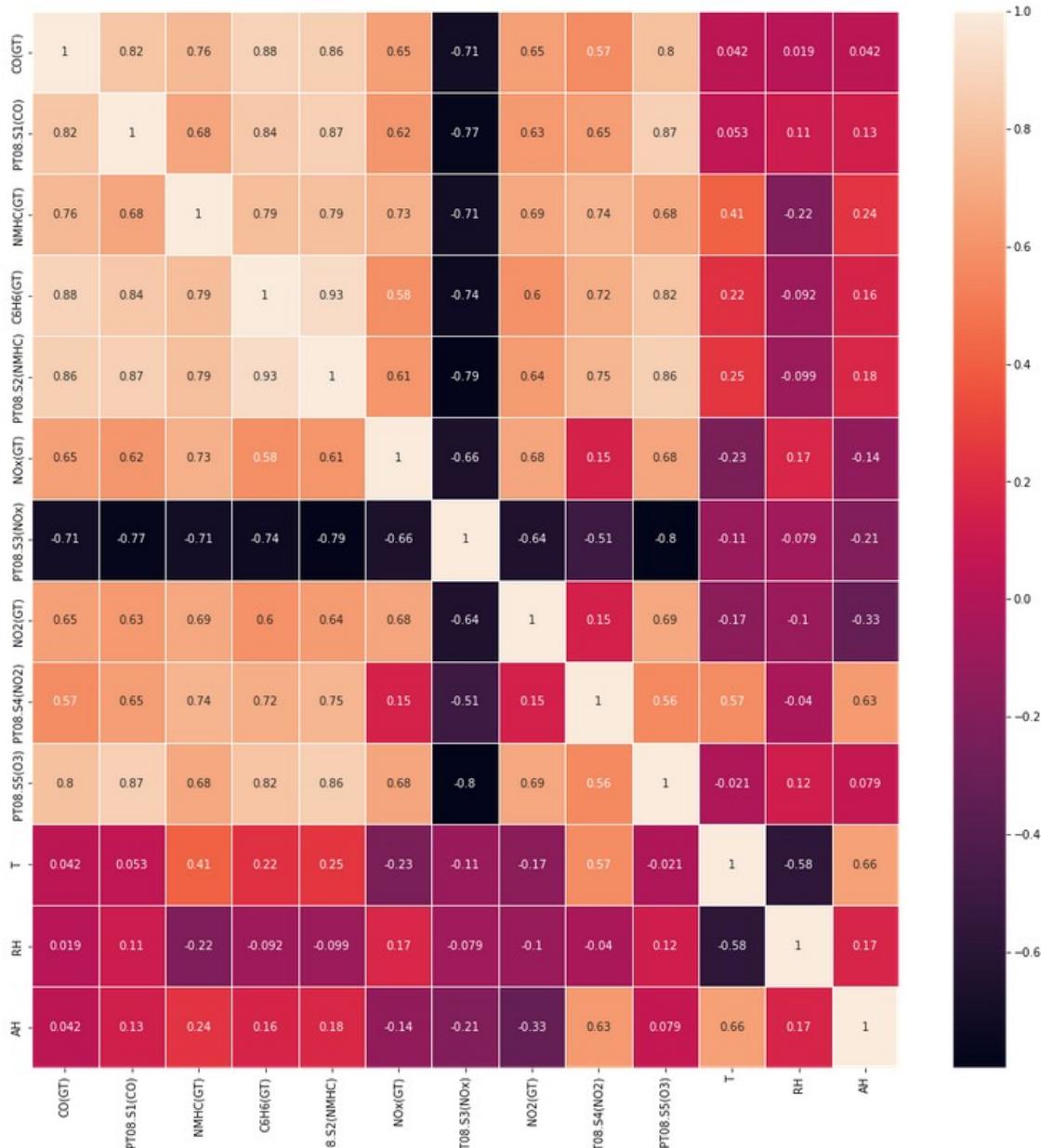
Желтым цветом показаны пропущенные значения.



Пропущенные значения были также устраниены по методу из п. б).

г) Визуальный анализ очищенных данных





д) Выводы

Стало меньше выбросов, были удалены редактированы пропущенные значения с сохранением первоначальных распределений.

2.2. Преобразование данных

Преобразований данных не проводилось

III. Формирование признаков

3.1. Сокращение числа признаков

Из таблицы пропущенных значений в п. 2.1, в) видно, что признак NMHC(GT) в основном состоит из пропусков, поэтому скорее всего не имеет статистической значимости. В связи с этим данный признак был исключен из общего числа признаков

3.2. Конструирование новых признаков

Конструирование новых признаков не проводилось.

3.3. Выводы

Был удален из рассмотрения признак NMHC(GT).

IV. Построение и исследование нейросетевых моделей

4.1. Параметры архитектуры и обучения многослойной нейронной сети

Параметр	Значение
Функция потерь	MSE
Число входов сети	11
Число выходов сети	1
Число скрытых слоев сети*	2
Число и AX нейронов 1-го скрытого слоя*	15
Число и AX нейронов 2-го скрытого слоя*	10
Число и AX нейронов 3-го скрытого слоя*	-
AX нейронов выходного слоя	linear
Кросс-валидация	Holdout (60/30/10)
Объёмы обучающей / валидационной / тестовой выборок	5613/2807/936
Режим обучения*	batch
Метод инициализации весов	метод Хавьера
Критерий останова	Ошибка на валидационной выборке увеличивается на протяжении 10 эпох
Ранний останов	да

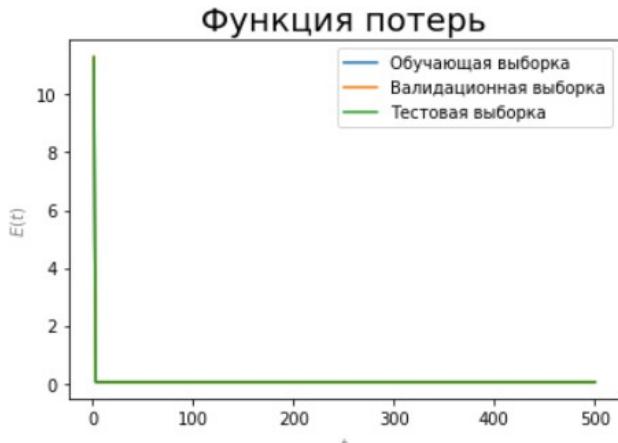
4.2. Исследование простого градиентного метода обучения

a) Исследование влияния параметра скорости обучения на качество обучения

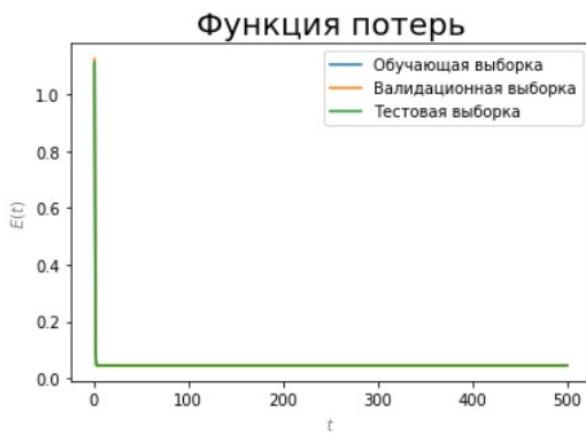
Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) при различных значениях параметра скорости обучения α (значения указать в таблице ниже).

Указание: обучение каждый раз начинать из одной и той же начальной точки.

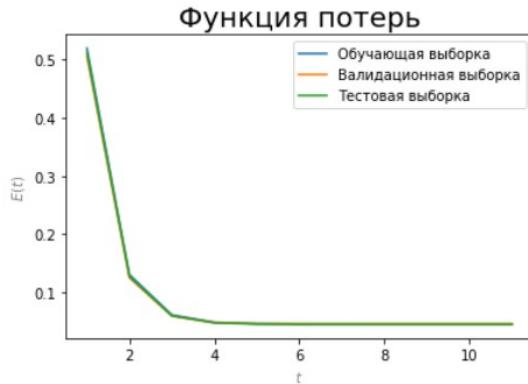
$$\alpha = 0.5$$



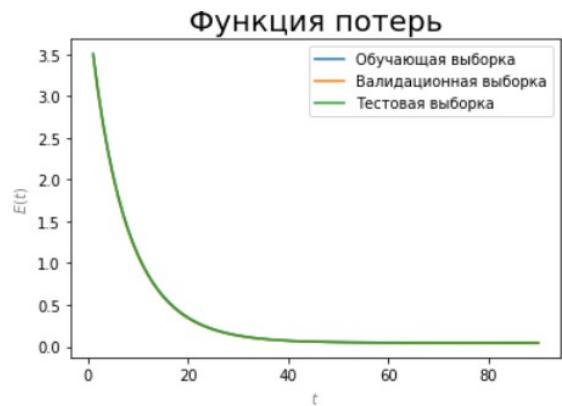
$$\alpha = 0.25$$



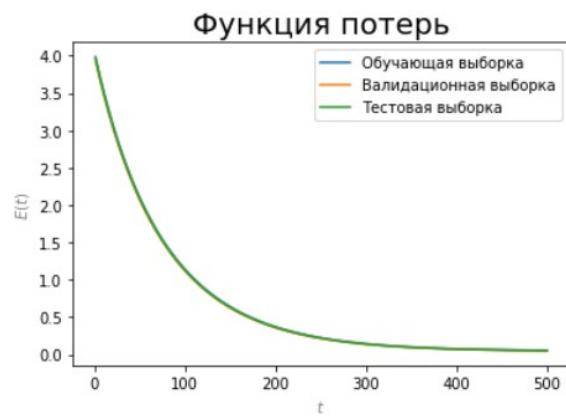
$$\alpha = 0.1$$



$\alpha = 0.01$



$\alpha = 0.001$



б) Заполнить таблицу по результатам обучения

№ п/п	Скорость обучения, α	Число эпох обучения	Ошибка на обучающей выборке, $E_{обуч}$	Ошибка на тестовой выборке, $E_{тест}$
1	0.5	500	0.046	0.046
2	0.25	500	0.046	0.046
3	0.1	11	0.046	0.046
4	0.01	90	0.046	0.046
5	0.001	500	0.055	0.054

Указание: все ошибки указываются для обученной сети.

в) Выводы

Параметр альфа ощутимо влияет на обучаемость сети. Лучшим результатом является $\lambda = 0.01$.

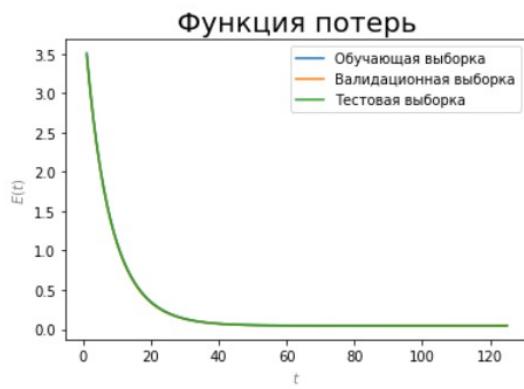
4.3. Исследование методов GDM и NAG

a) Исследование влияния параметра момента на качество обучения

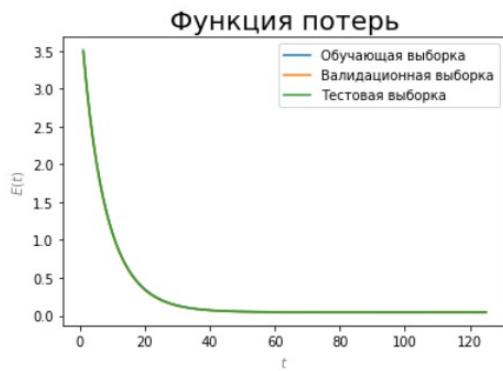
Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) при различных значениях параметра момента μ (указать в таблице ниже).

Указание: параметр скорости обучения выбрать наилучшим по результатам исследований п. 4.2.

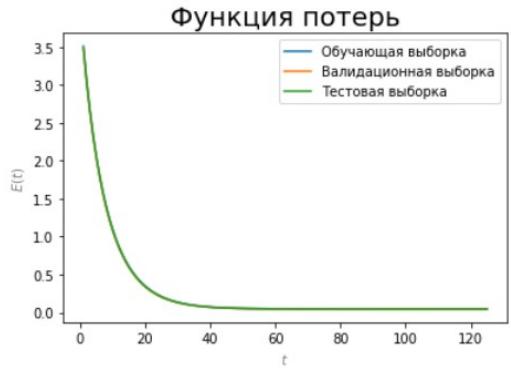
$\mu = 0.001$, nesterov = false



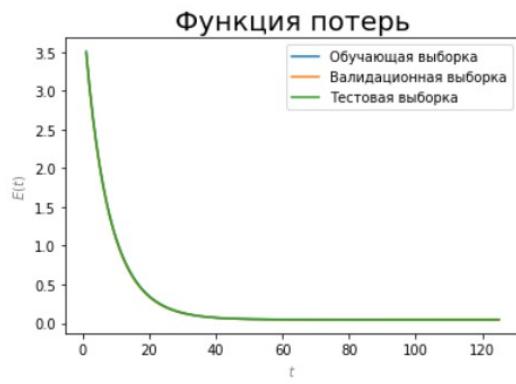
$\mu = 0.001$, nesterov = true



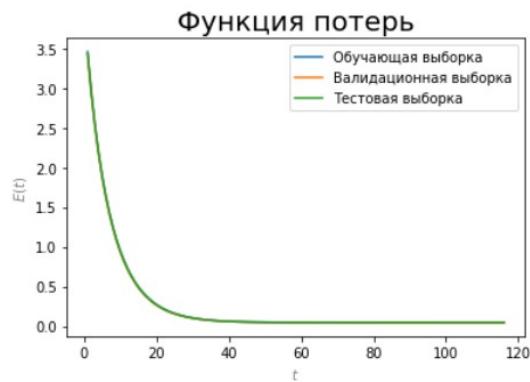
$\mu = 0.01$, nesterov = false



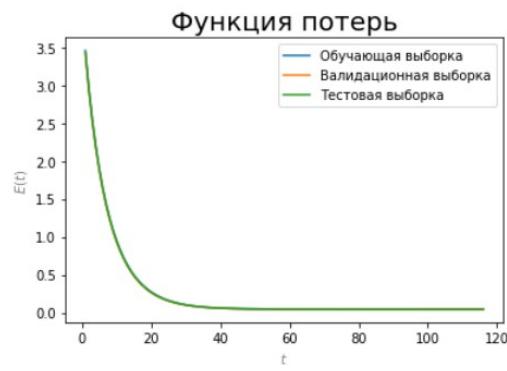
$\mu = 0.01$, nesterov = true



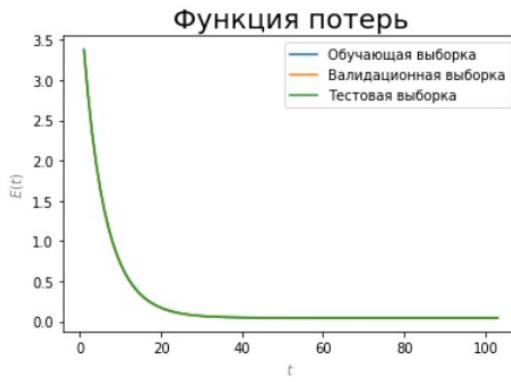
$\mu = 0.1$, nesterov = false



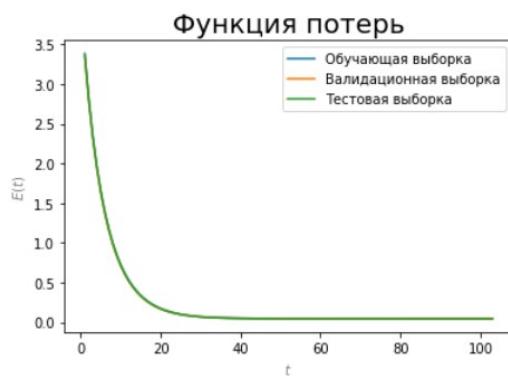
$\mu = 0.1$, nesterov = true



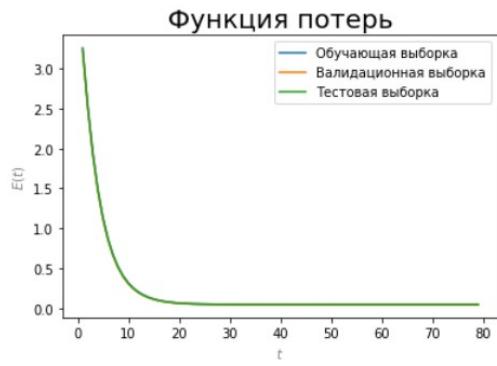
$\mu = 0.25$, nesterov = false



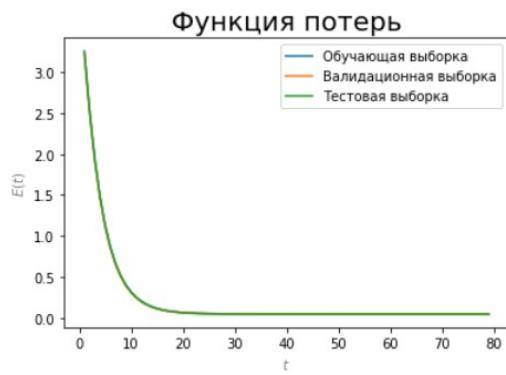
$\mu = 0.25$, nesterov = true



$\mu = 0.5$, nesterov = false



$\mu = 0.5$, nesterov = true



б) Заполнить таблицу по результатам обучения

№ п/п	Метод	Момент, μ	Число эпох обучения	Ошибка на обучающей выборке, $E_{обуч}$	Ошибка на тестовой выборке, $E_{тест}$
1	GDM	0.001	125	0.046	0.046
2	NAG		125	0.046	0.046
3	GDM	0.01	124	0.046	0.046
4	NAG		124	0.046	0.046
5	GDM	0.1	116	0.046	0.046
6	NAG		116	0.046	0.046
7	GDM	0.25	103	0.046	0.046
8	NAG		101	0.046	0.046

9	GDM	0.5	76	0.046	0.046
10	NAG		79	0.046	0.046

в) Выводы

Чем больше момент, тем меньше эпох требуется модели, чтобы прийти в оптимальную точку. Лучшим значением является $\mu = 0.5$.

4.4. Исследование методов наискорейшего спуска и сопряжённых градиентов

a) Сравнение кривых обучения

Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) для: 1) метода наискорейшего спуска; 2) метода Флетчера-Ривса; 3) метода Полака-Райбера.

б) Заполнить таблицу по результатам обучения

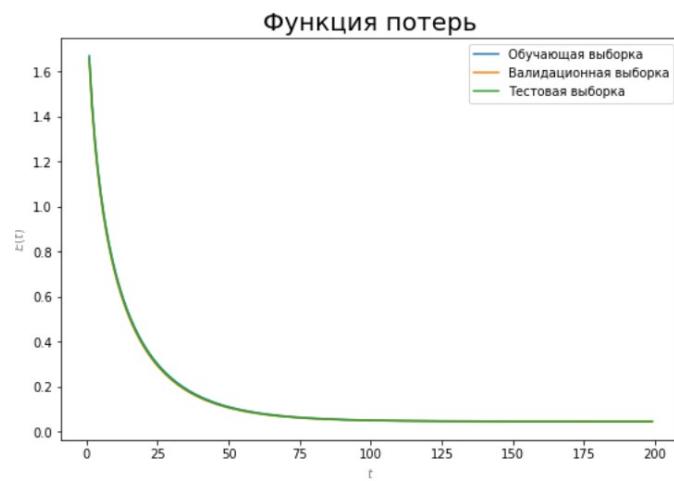
№ п/п	Метод	Число эпох обучения	Ошибка на обучающей выборке, $E_{обуч}$	Ошибка на тестовой выборке, $E_{тест}$
1	SGD			
2	Fletcher-Reeves			
3	Polak-Ribiere			

в) Выводы

4.5. Исследование метода AdaGrad

a) Кривые обучения

Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения).

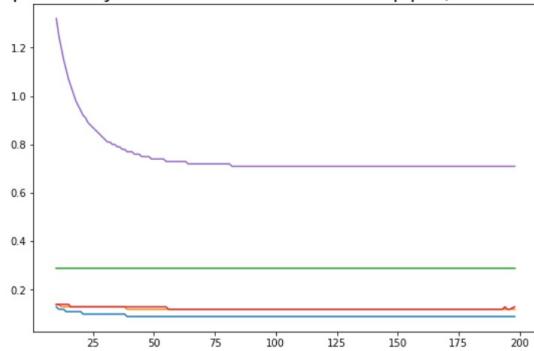


Указание: базовую скорость обучения выбрать наилучшей по результатам исследований п. 4.2.

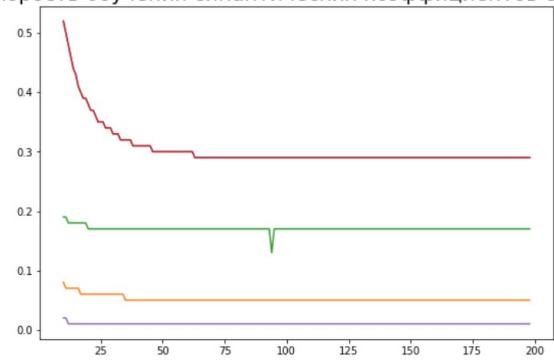
б) Исследование динамики скорости обучения

Построить графики зависимости скорости обучения отдельных синаптических коэффициентов сети (выбрать произвольно из разных слоёв сети) от времени обучения, а также графики зависимости минимальной, максимальной и средней (по всем настраиваемым параметрам сети) скорости обучения от времени обучения.

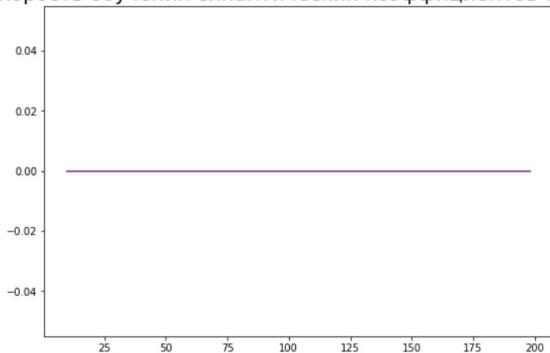
Скорость обучения синаптических коэффициентов слоя 1



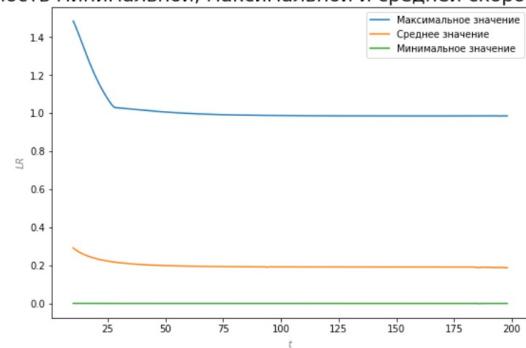
Скорость обучения синаптических коэффициентов слоя 2



Скорость обучения синаптических коэффициентов слоя 3



Зависимость минимальной, максимальной и средней скорости обучения



в) Заполнить таблицу по результатам обучения

№ п/п	Метод	Число эпох обучения	Ошибка на обучающей выборке, $E_{обуч}$	Ошибка на тестовой выборке, $E_{тест}$
1	AdaGrad	199	0.046	0.046

г) Выводы

AdaGrad позволяет коэффициентам достигнуть точки минимума за достаточно малое число эпох (даже с ранней остановкой).

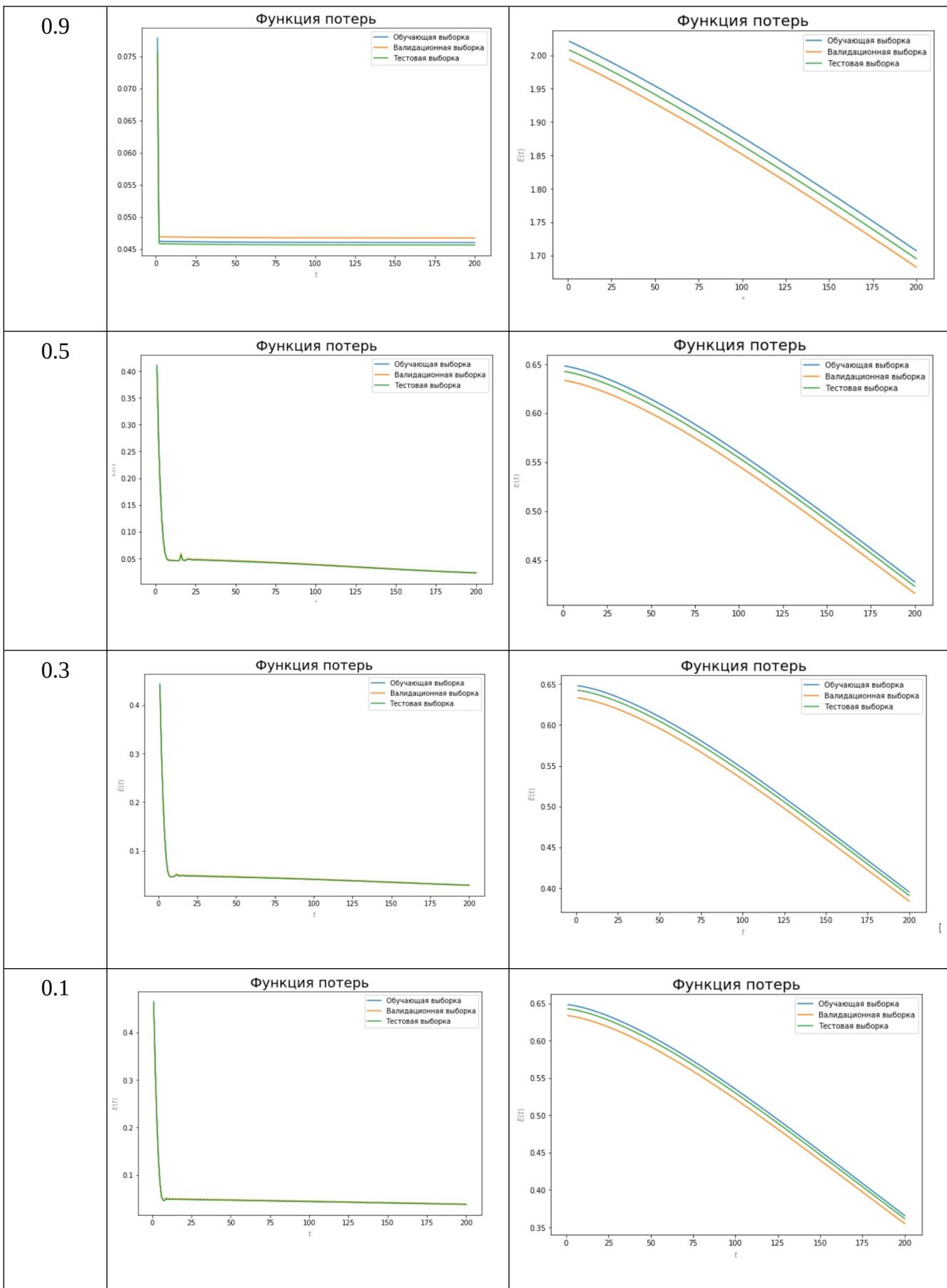
4.6. Исследование методов RMSProp и AdaDelta

а) Сравнение кривых обучения

Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) при различных значениях параметра сглаживания ρ (значения указать в таблице ниже) для методов RMSProp и AdaDelta.

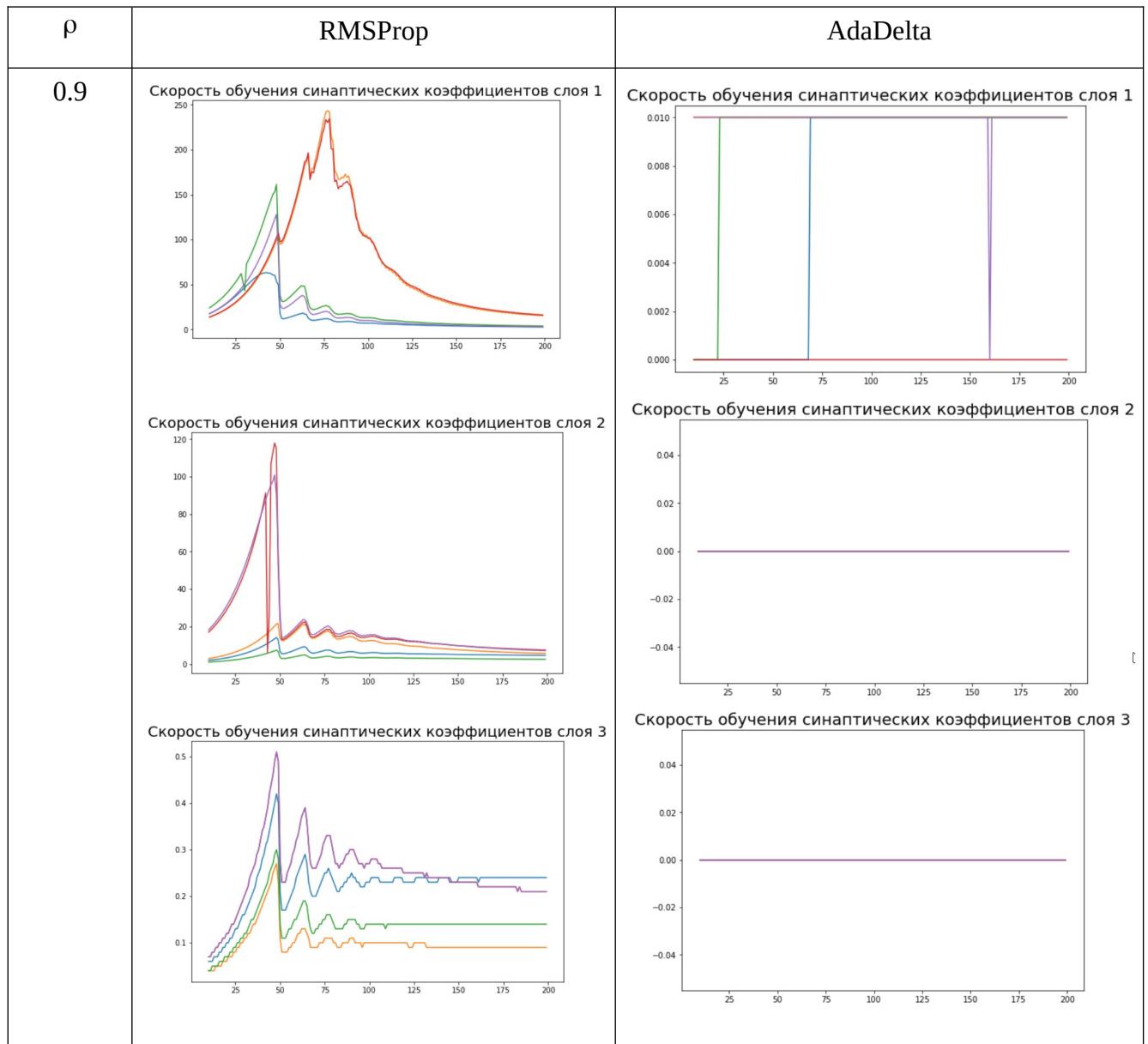
Указание: базовую скорость обучения в методе RMSProp выбрать наилучшей по результатам исследований п. 4.2.

ρ	RMSProp	AdaDelta

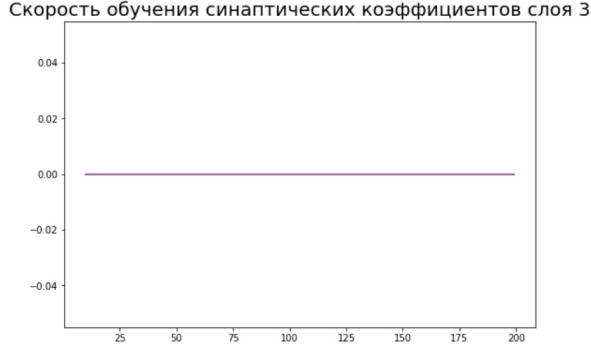
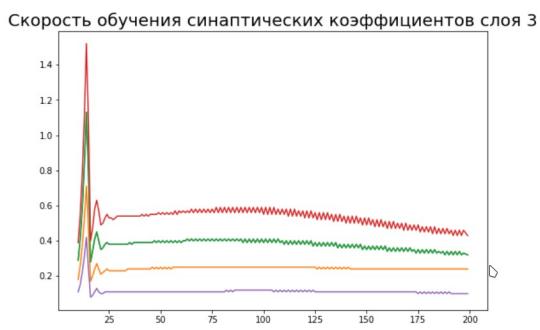
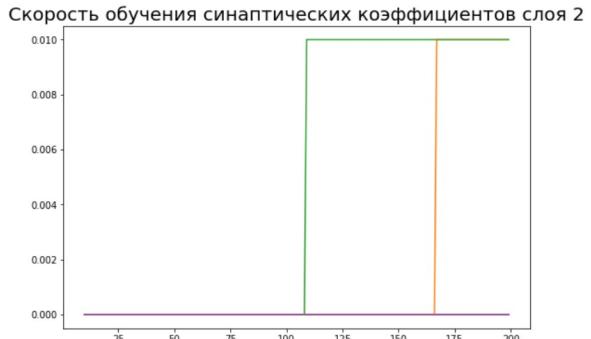
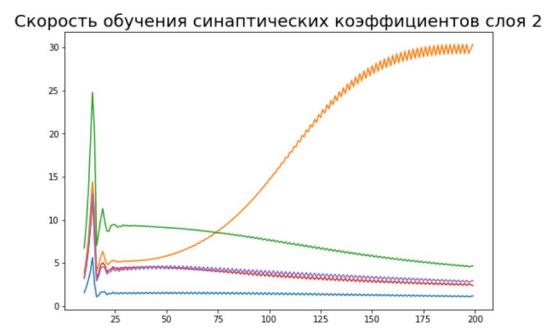
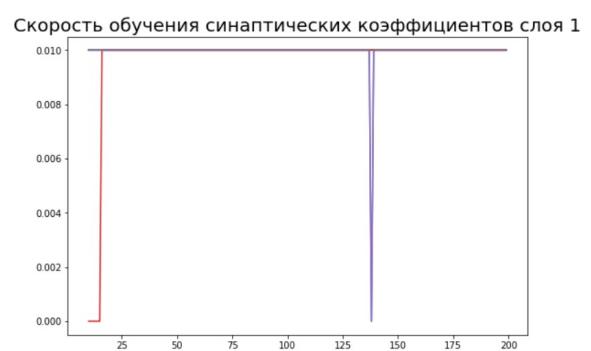
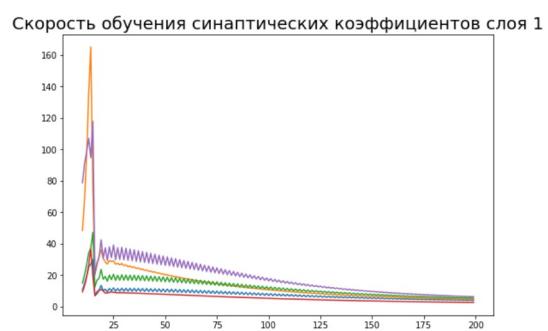


б) Исследование динамики скорости обучения

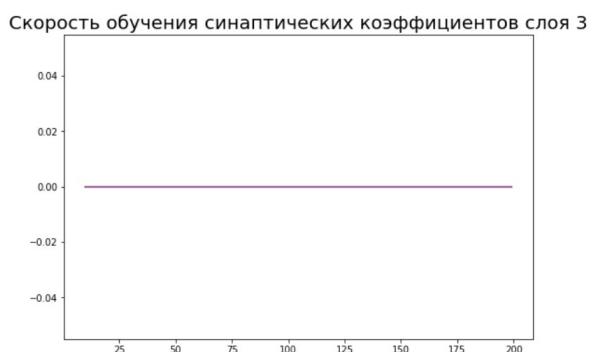
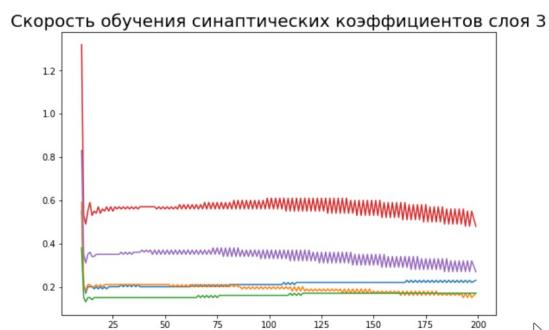
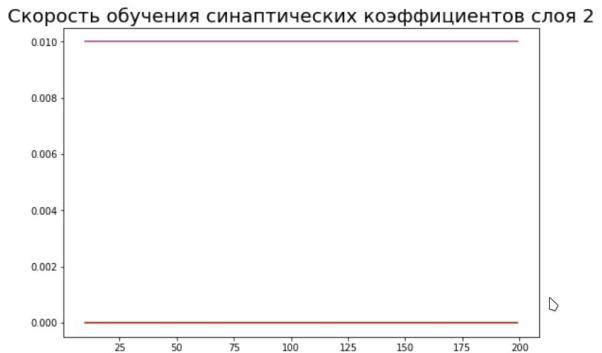
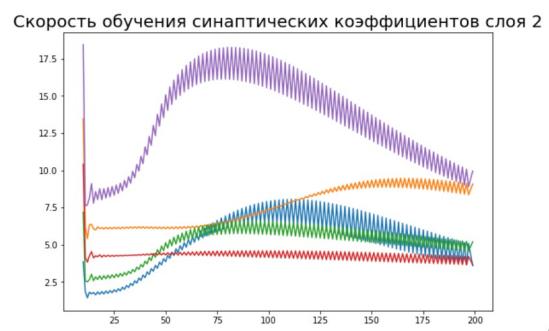
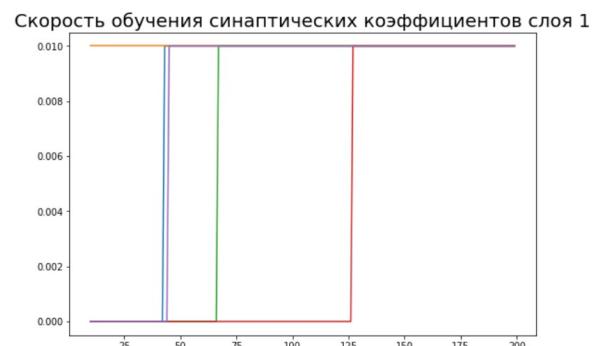
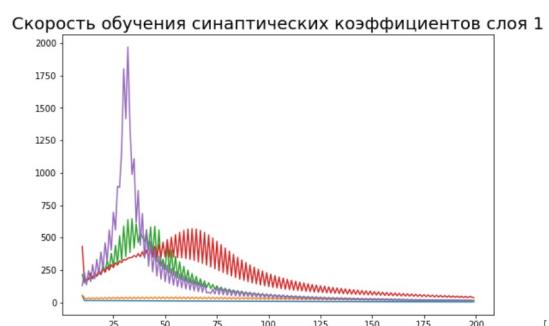
При различных значениях параметра сглаживания построить графики зависимости скорости обучения отдельных синаптических коэффициентов сети (выбрать произвольно из разных слоёв сети) от времени обучения, а также графики зависимости минимальной, максимальной и средней (по всем настраиваемым параметрам сети) скорости обучения от времени обучения для методов RMSProp и AdaDelta.

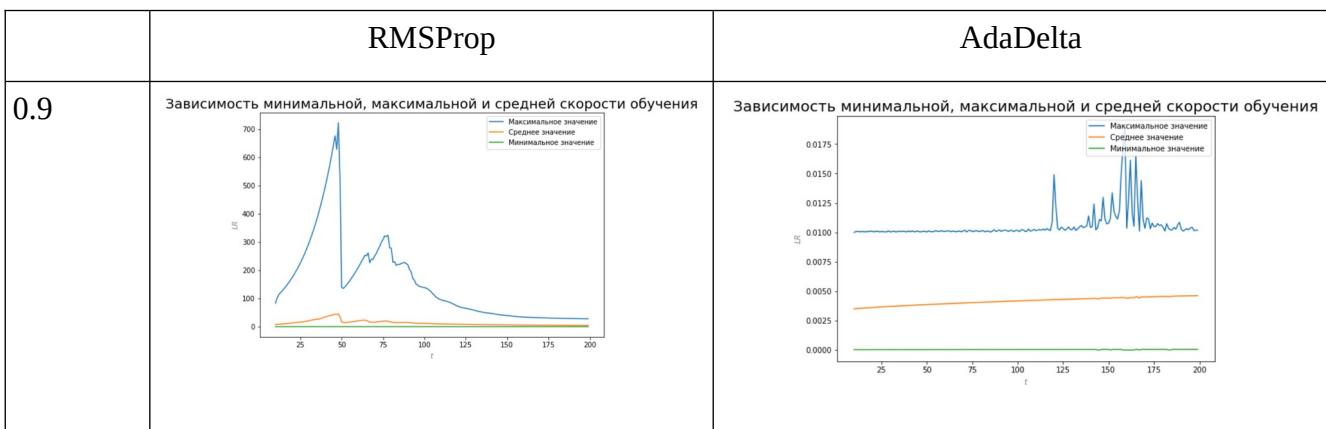
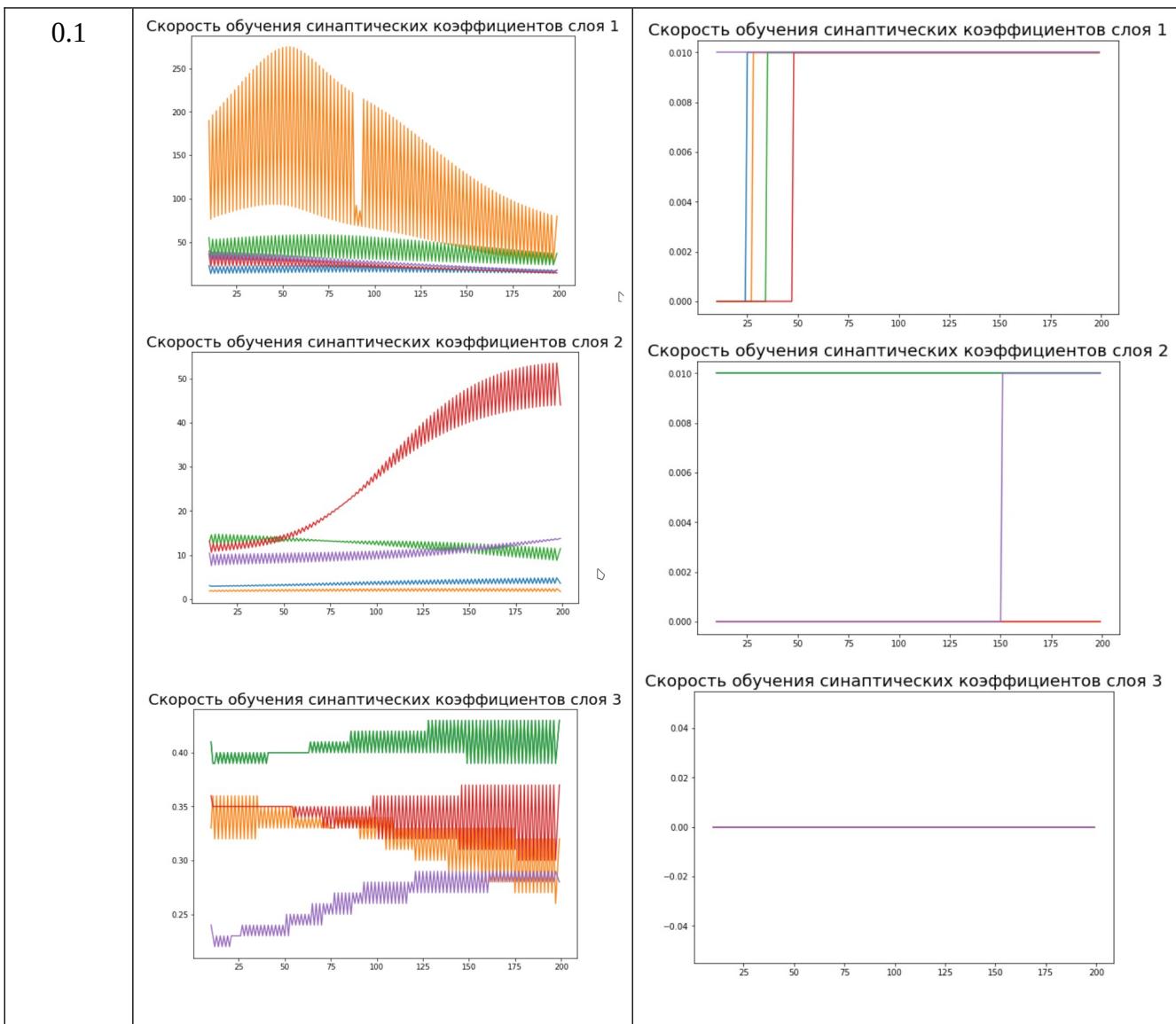


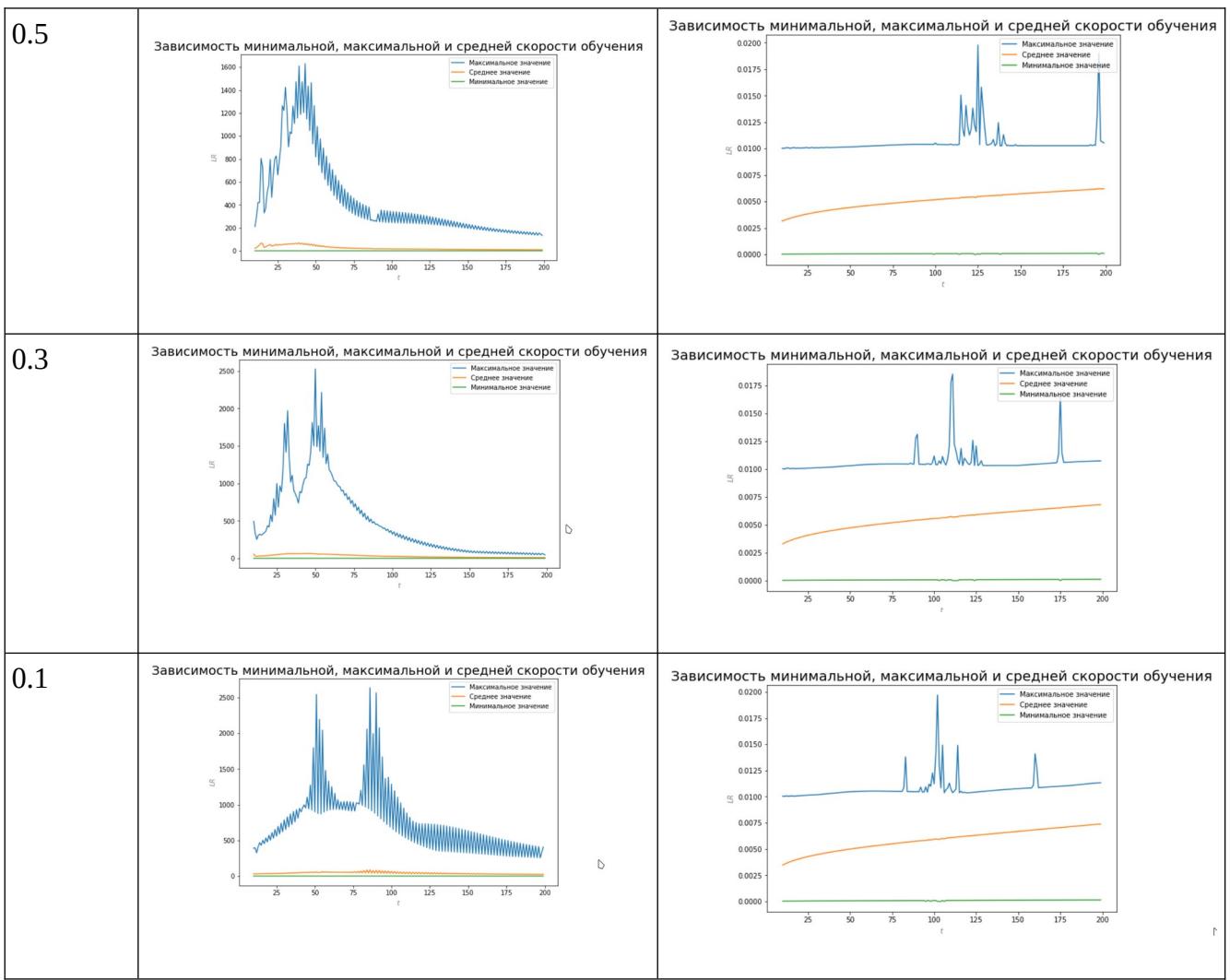
0.5



0.3







в) Заполнить таблицу по результатам обучения

№ п/п	Метод	Параметр сглаживания, ρ	Число эпох обучения	Ошибка на обучающей выборке, $E_{обуч}$	Ошибка на тестовой выборке, $E_{тест}$
1	RMSProp	0.1	200	0.044	0.044
2	AdaDelta		200	0.533	0.528
3	RMSProp	0.3	200	0.041	0.041
4	AdaDelta		200	0.545	0.540
5	RMSProp	0.5	200	0.038	0.038
6	AdaDelta		200	0.558	0.553
7	RMSProp	0.9	200	0.033	0.033
8	AdaDelta		200	0.575	0.570

г) Выводы

Метод RMSProp сходится достаточно быстро и, в отличии от метода AdaDelta, дает неплохие результаты.

При увеличении параметра сглаживания у RMSProp уменьшается ошибка, у AdaDelta — увеличивается.

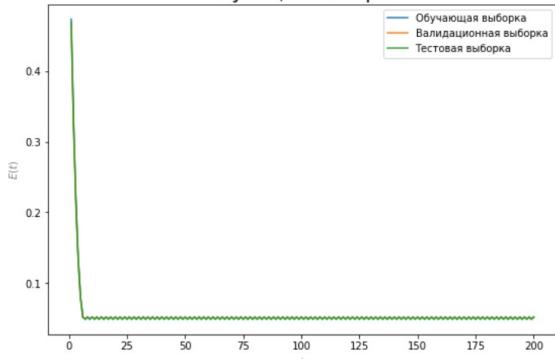
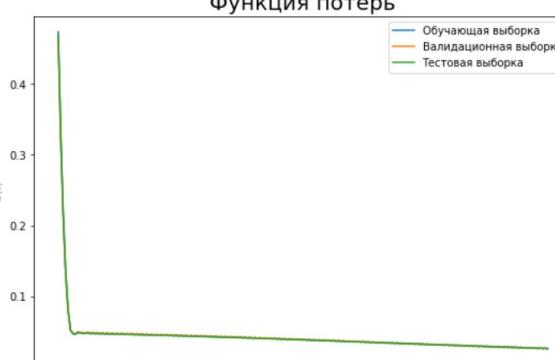
4.7. Исследование метода Adam

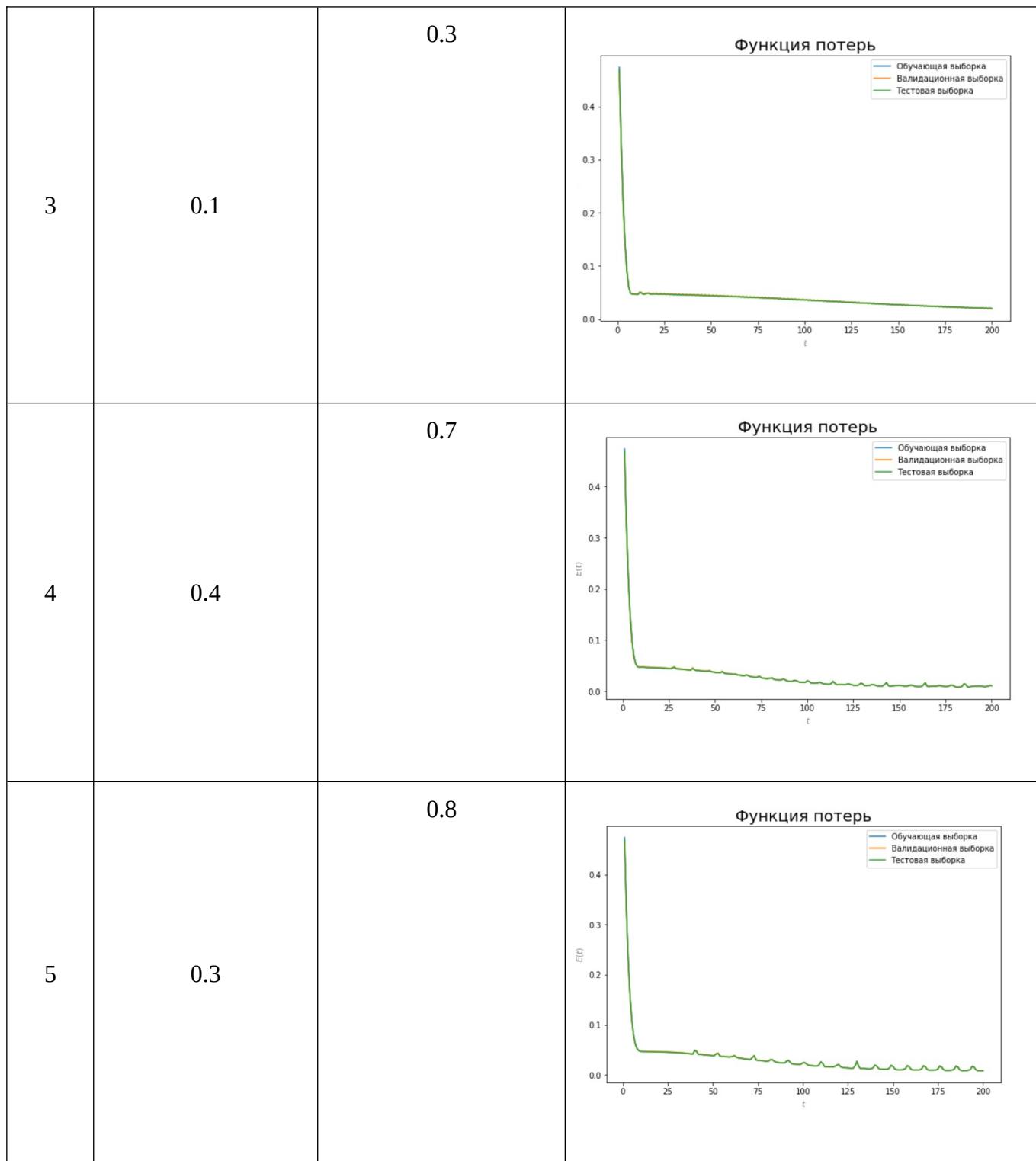
a) Сравнение кривых обучения

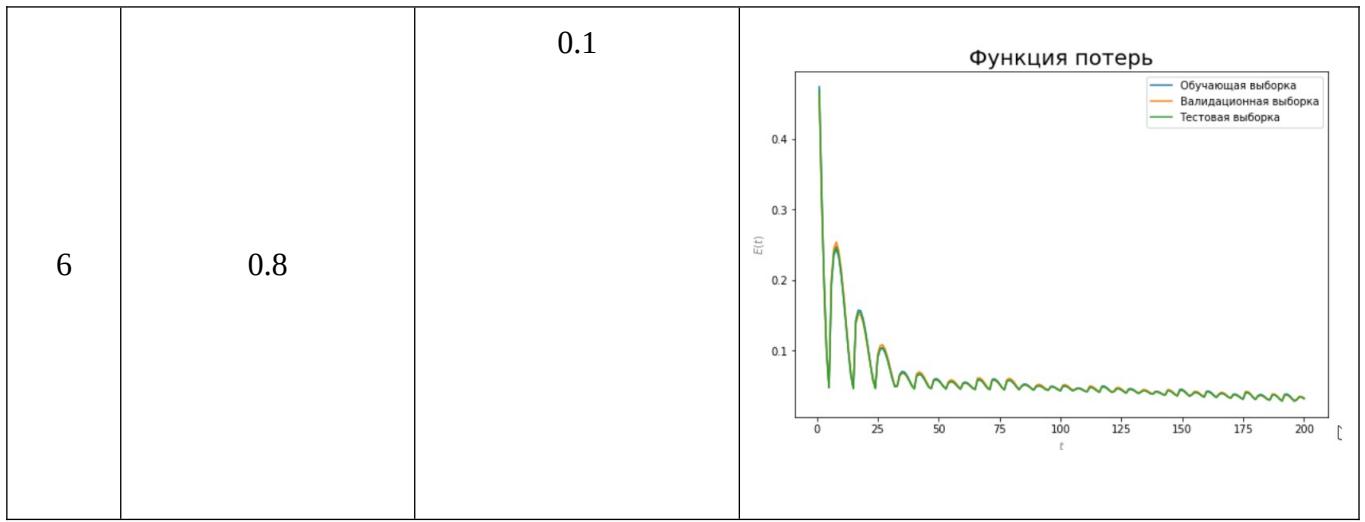
Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) при различных значениях параметров сглаживания β_1, β_2 (значения указать в таблице ниже).

Указание 1: базовую скорость обучения выбрать наилучшей по результатам исследований п. 4.2.

Указание 2: в одном из экспериментов выбрать β_1 и β_2 равными наилучшему значению параметра ρ по результатам исследований п. 4.6.

№ п/п	β_1	β_2	График
1	0	0	<p style="text-align: center;">Функция потерь</p> 
2	0.1	0.1	<p style="text-align: center;">Функция потерь</p> 

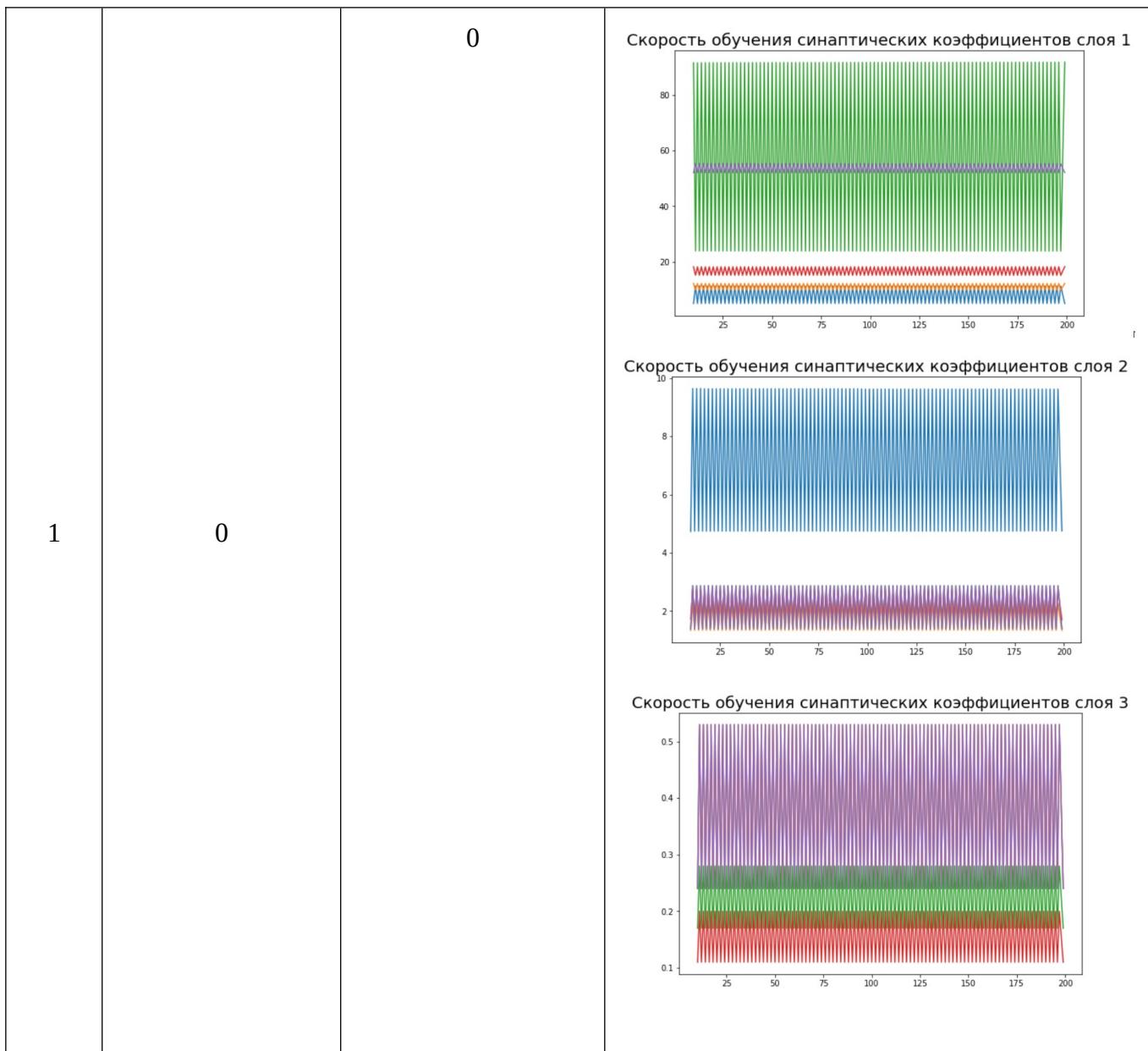


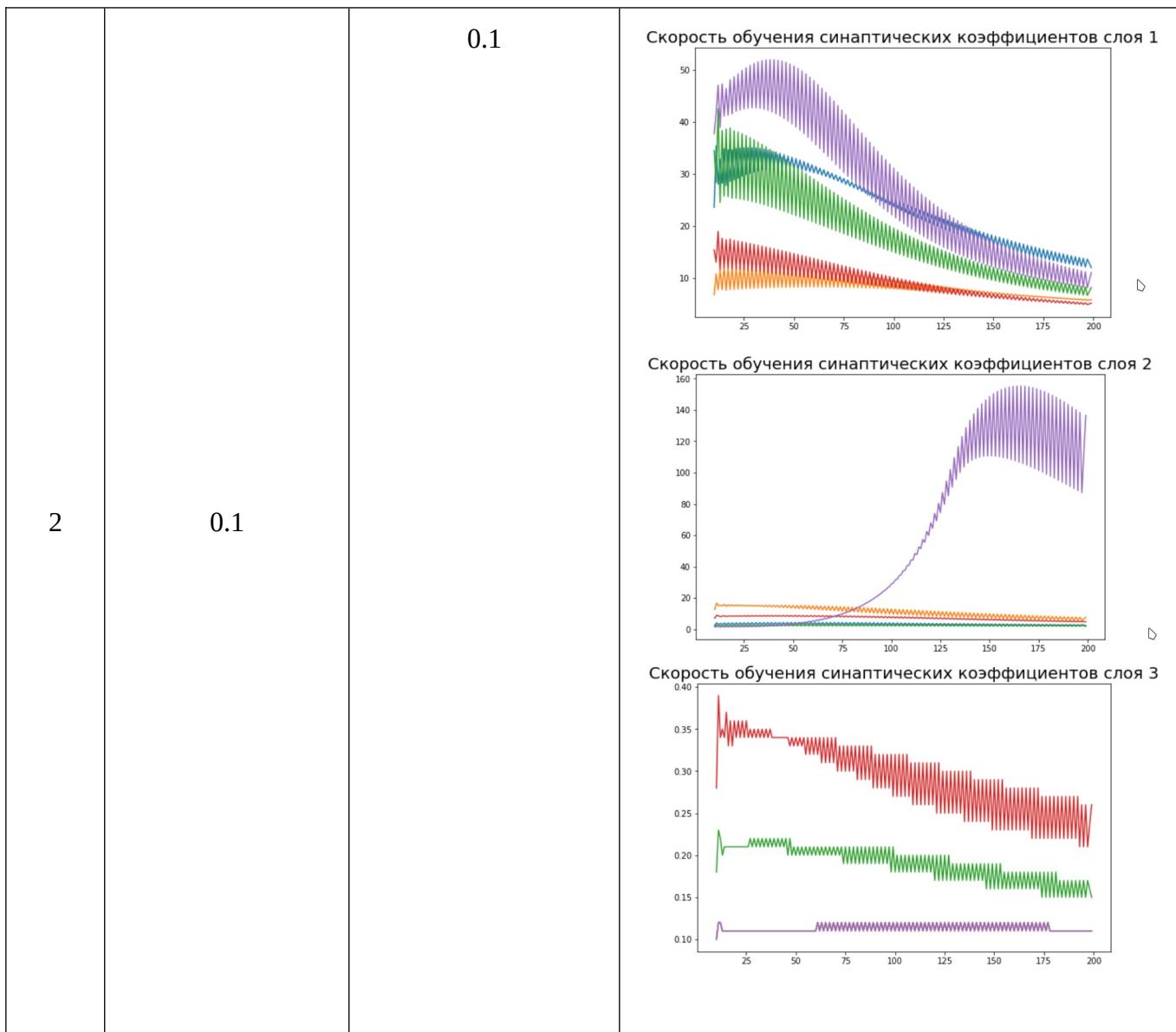


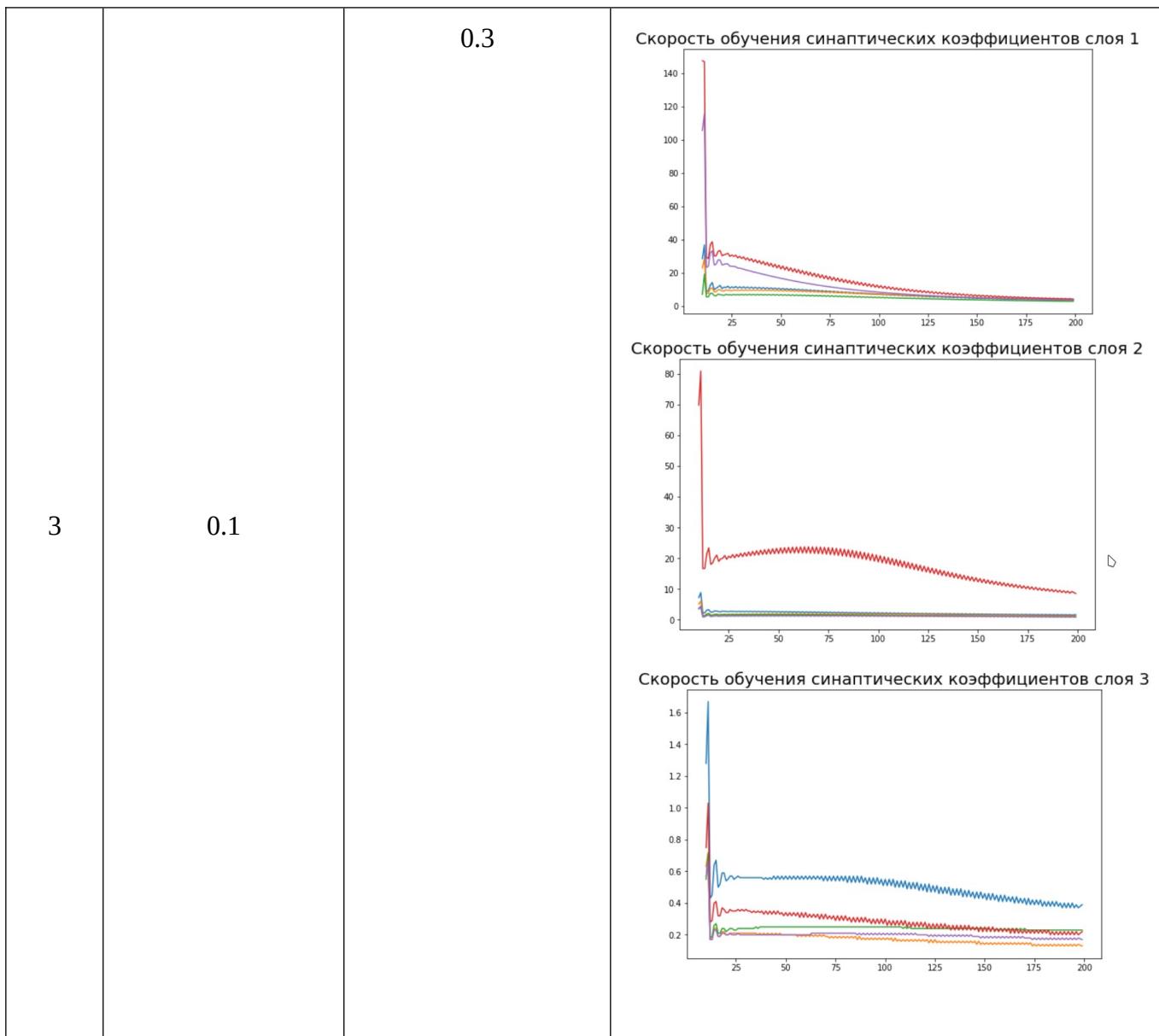
б) Исследование динамики скорости обучения

При различных значениях параметров сглаживания построить графики зависимости скорости обучения отдельных синаптических коэффициентов сети (выбрать произвольно из разных слоёв сети) от времени обучения, а также графики зависимости минимальной, максимальной и средней (по всем настраиваемым параметрам сети) скорости обучения от времени обучения.

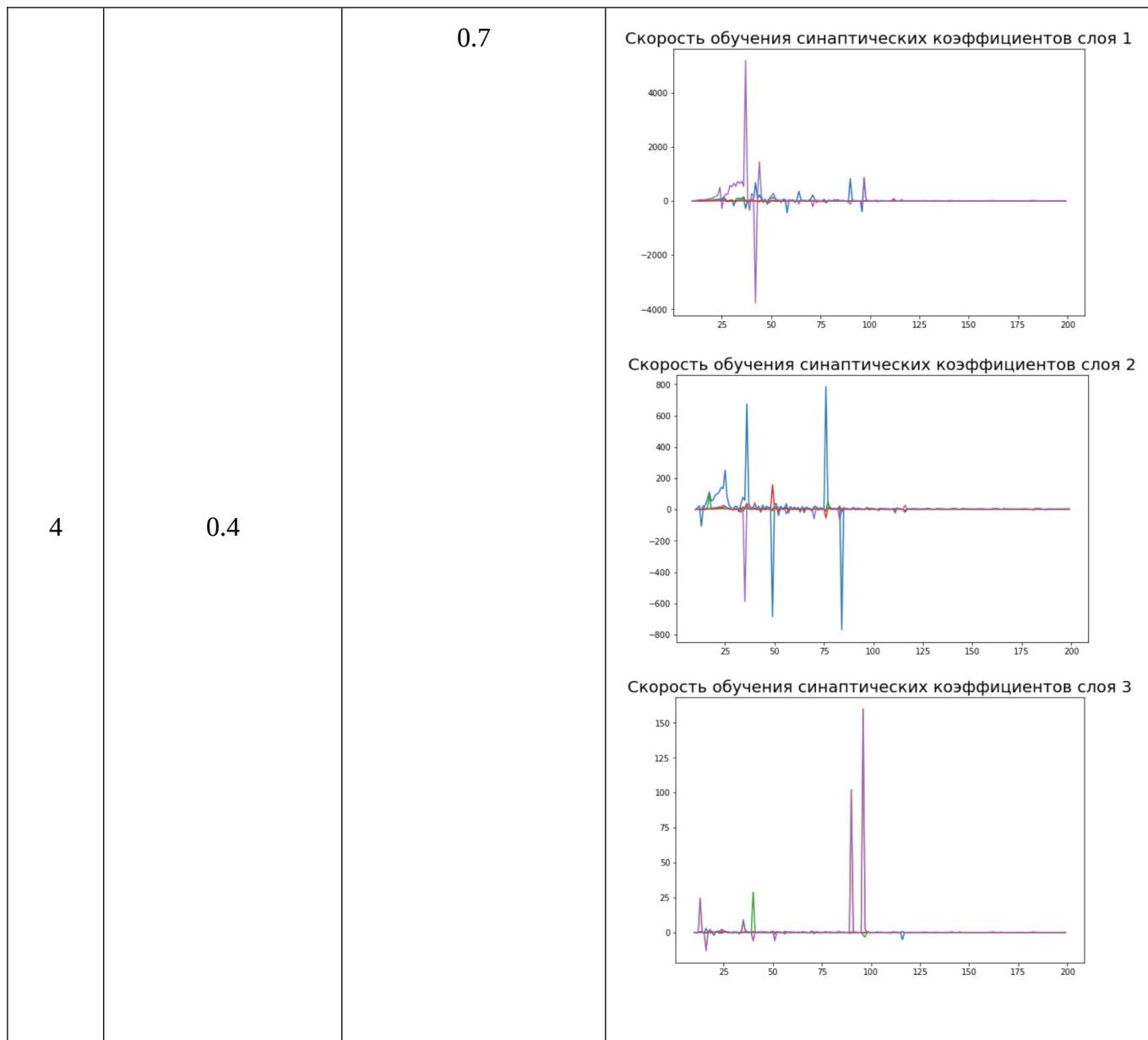
№ п/п	β_1	β_2	График
-------	-----------	-----------	--------

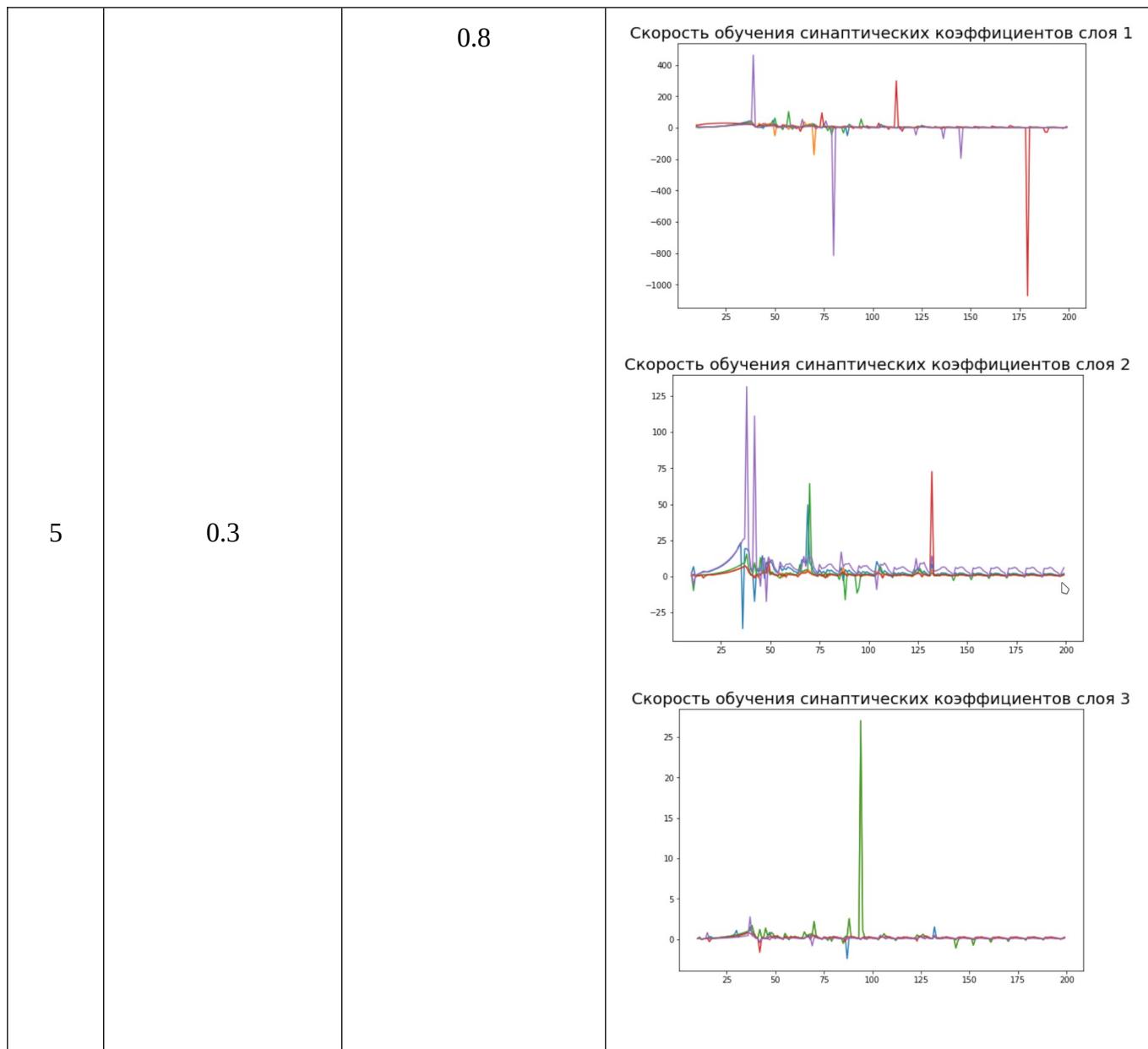


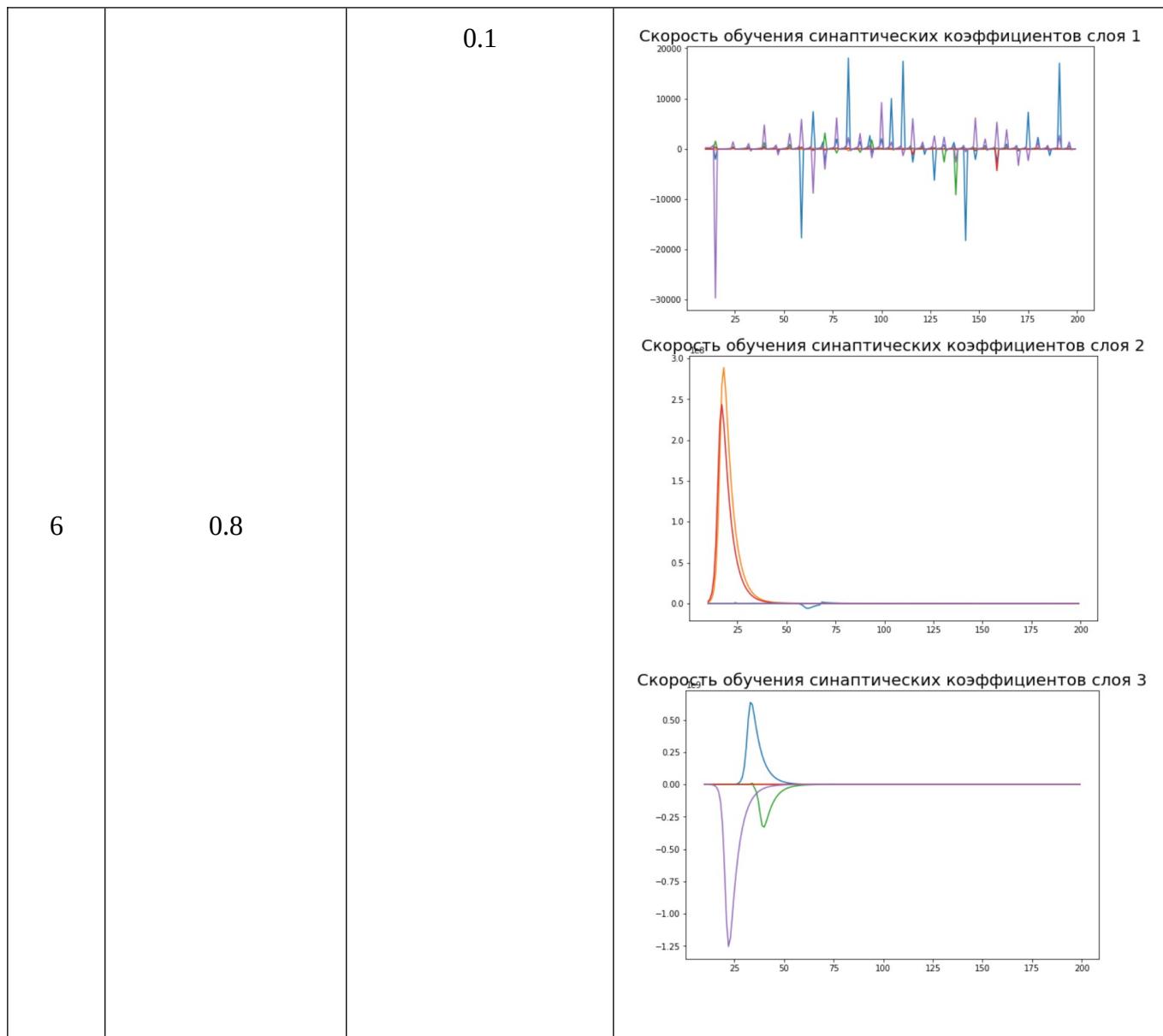




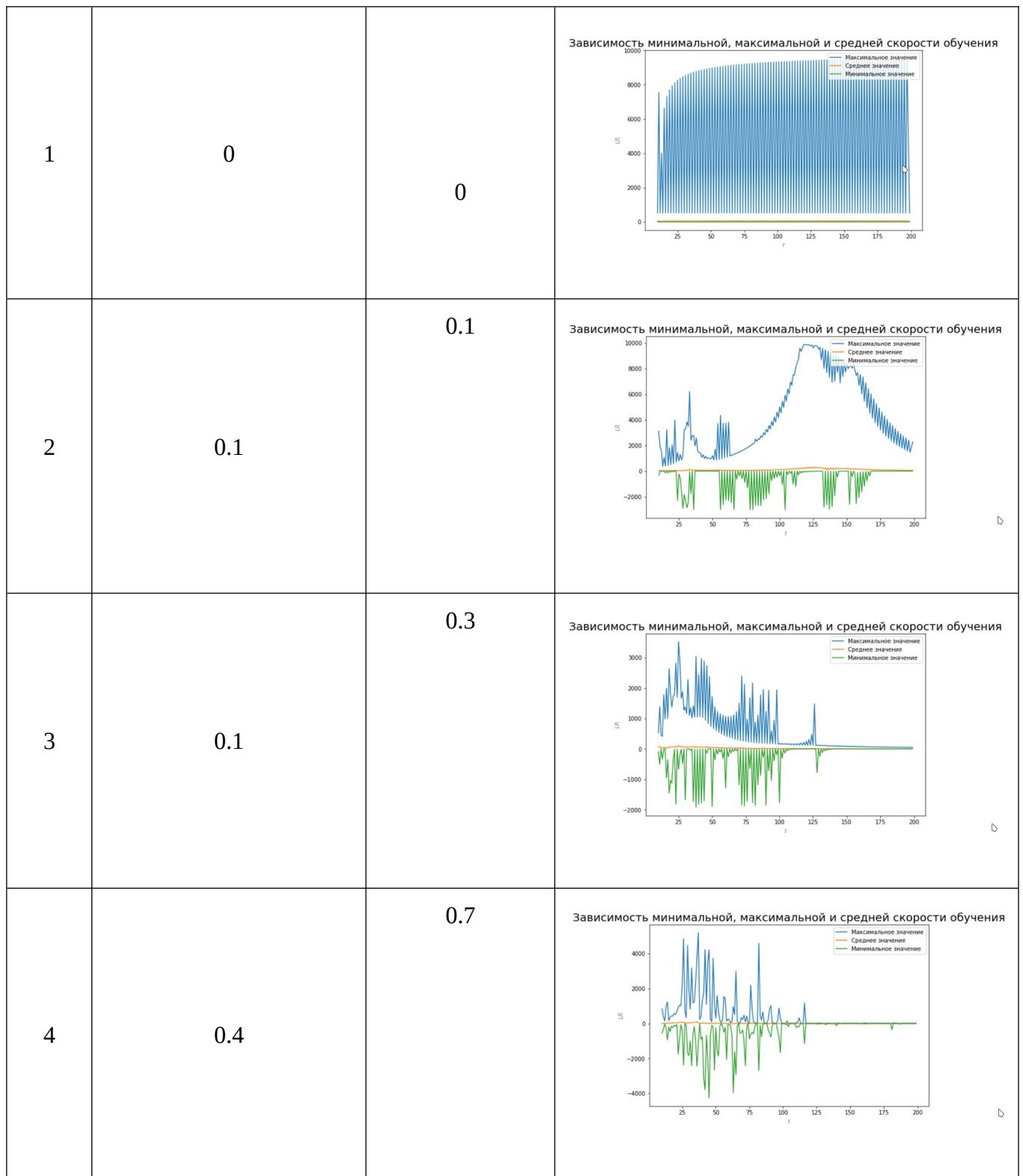
▷

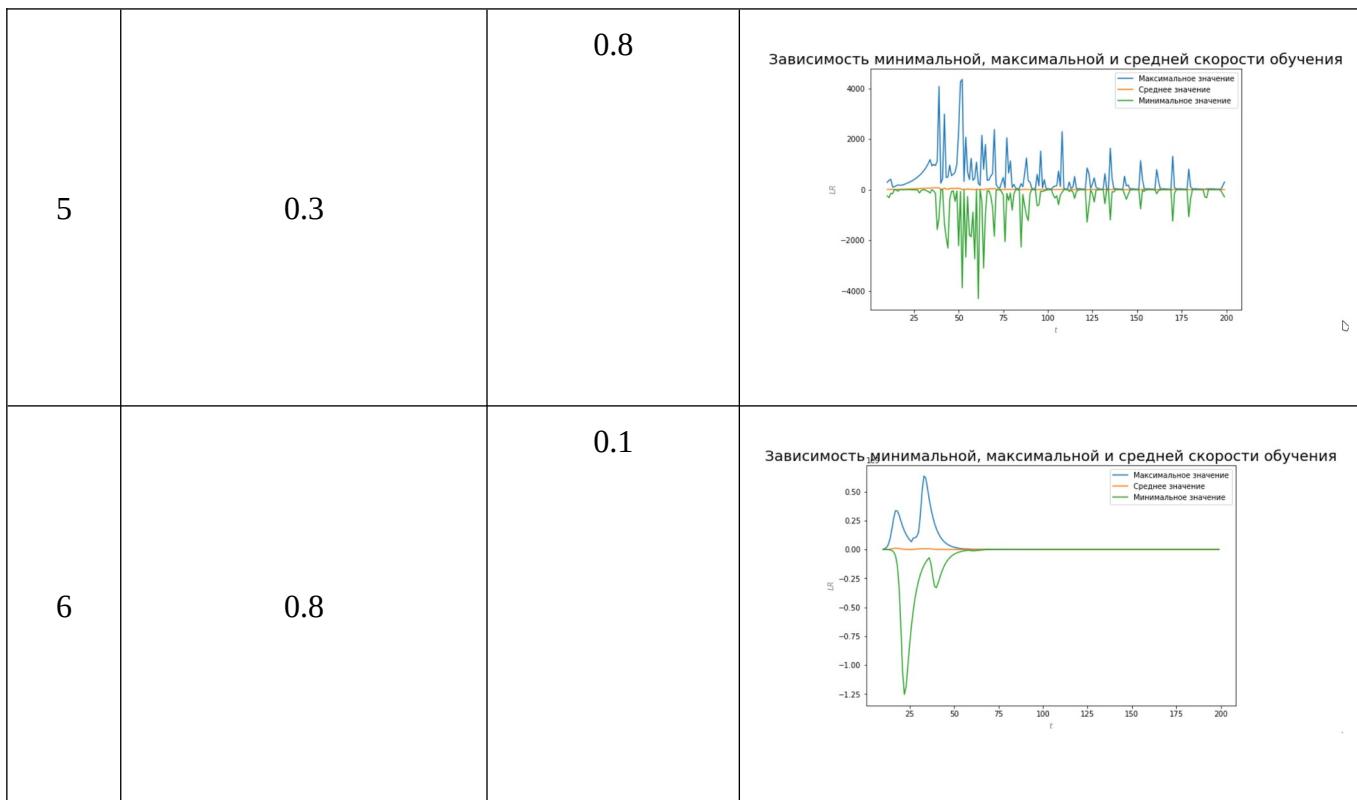






№ п/п	β_1	β_2	График
-------	-----------	-----------	--------





в) Заполнить таблицу по результатам обучения

№ п/п	β_1	β_2	Число эпох обучения	Ошибка на обучающей выборке, $E_{обуч}$	Ошибка на тестовой выборке, $E_{тест}$
1	0	0	200	0.048	0.048
2	0.1	0.1	200	0.039	0.039
3	0.1	0.3	200	0.036	0.036
4	0.4	0.7	200	0.020	0.019
5	0.3	0.8	200	0.025	0.024
6	0.8	0.1	200	0.049	0.049

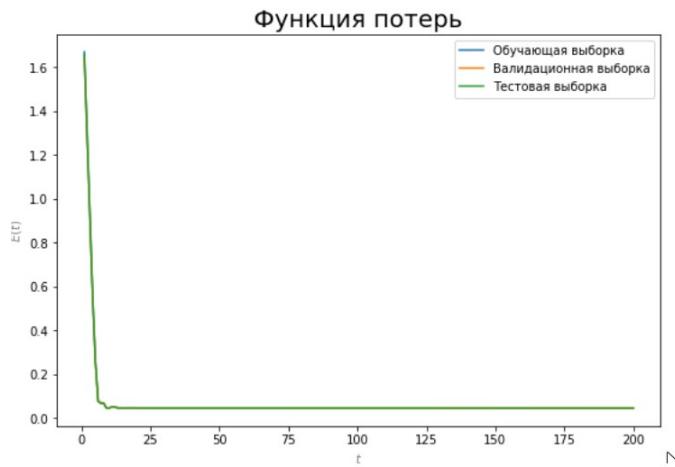
г) Выводы

Adam достаточно качественно обучает модель. При этом на скорость обучения коэффициентов влияют коэффициенты бета. Лучше всего показывает себя при $\beta_1=0.3$ и $\beta_2=0.8$.

4.8. Исследование метода RProp

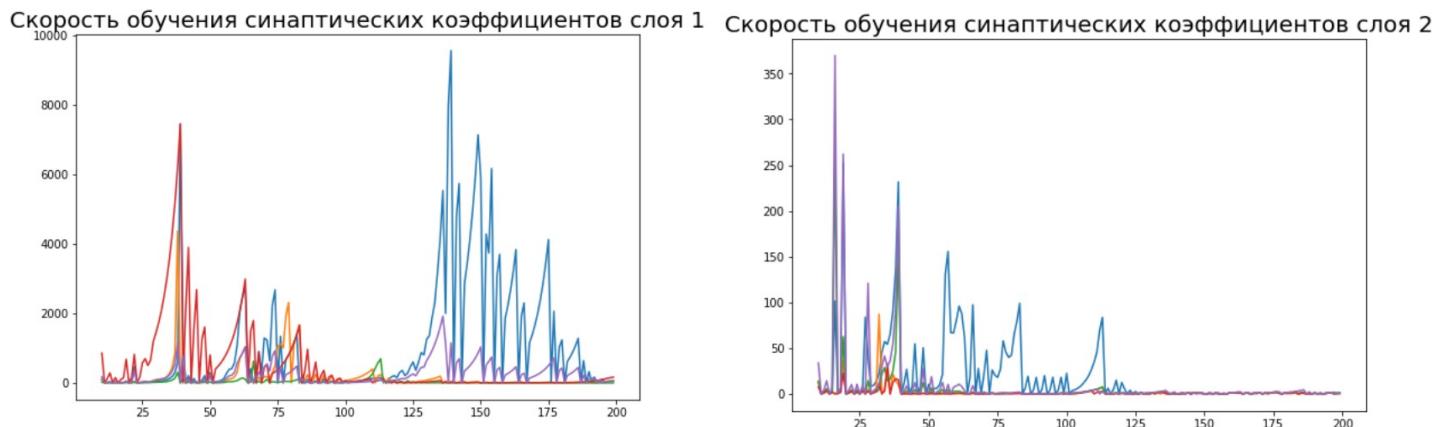
a) Кривые обучения

Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения).

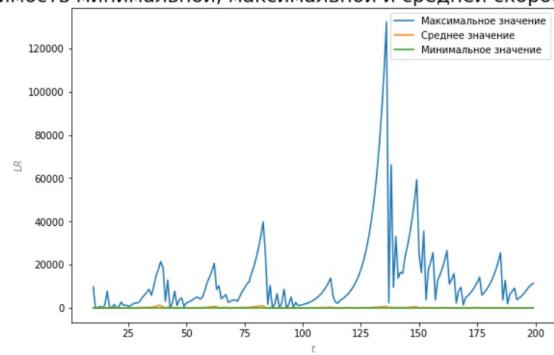
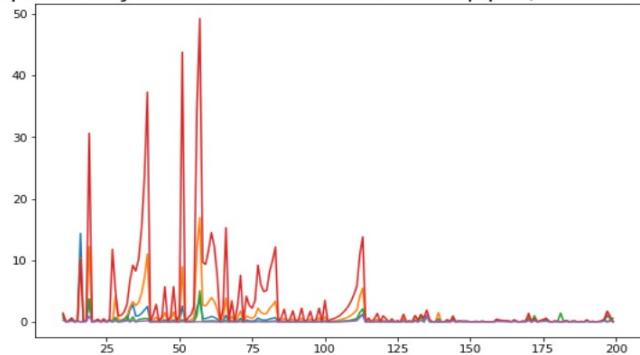


б) Исследование динамики приращений весов

Построить графики зависимости приращений отдельных синаптических коэффициентов сети (выбрать произвольно из разных слоёв сети) от времени обучения, а также графики зависимости минимального, максимального и среднего (по всем настраиваемым параметрам сети) приращения от времени обучения.



Скорость обучения синаптических коэффициентов слоя 3 Зависимость минимальной, максимальной и средней скорости обучения



в) Заполнить таблицу по результатам обучения

№ п/п	Метод	Число эпох обучения	Ошибка на обучающей выборке, $E_{обуч}$	Ошибка на тестовой выборке, $E_{тест}$
1	RProp	200	0.046	0.046

г) Выводы

RProp обучает коэффициенты со значительными флюктуациями, однако также сходится к локальному экстремуму.

4.9. Исследование методов Левенберга-Маркардта и BFGS

а) Кривые обучения

Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) для: 1) метода Левенберга-Маркардта; 2) метода BFGS.

б) Заполнить таблицу по результатам обучения

№ п/п	Метод	Число эпох обучения	Ошибка на обучающей выборке, $E_{обуч}$	Ошибка на тестовой выборке, $E_{тест}$
1	LM			
2	BFGS			

в) Выводы

4.10. Исследование метода стохастического градиента

a) Сравнение кривых обучения

Построить графики зависимости ошибки сети на обучающей, валидационной и тестовой выборках от времени обучения (кривые обучения) при различных размерах mini-batch'ей (указаны в таблице ниже).

Указание: параметр скорости обучения простого градиентного метода выбрать наилучшим по результатам исследований п. 4.2.

№ п/п	Метод	Размер mini-batch'a	График
1	GD	1	
2	GD	20	
3	GD	100	

4	GD	равен объёму выборки	<p>Функция потерь</p> <ul style="list-style-type: none"> Обучающая выборка Валидационная выборка Тестовая выборка
---	----	----------------------	--

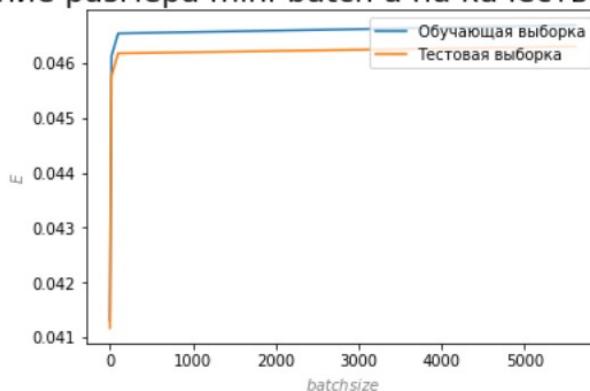
б) Заполнить таблицу по результатам обучения

№ п/п	Метод	Размер mini-batch'a	Число эпох обучения	Ошибка на обучающей выборке, $E_{обуч}$	Ошибка на тестовой выборке, $E_{тест}$
1	GD	1	3	0.46	0.46
2	GD	20	7	0.47	0.47
3	GD	100	10	0.47	0.46
4	GD	равен объёму выборки	100	0.47	0.46

в) Исследование влияния размера mini-batch'a на качество обучения

Построить графики зависимости ошибок обученной сети на обучающей и тестовой выборках от размера mini-batch'a.

Влияние размера mini-batch'a на качество обучения



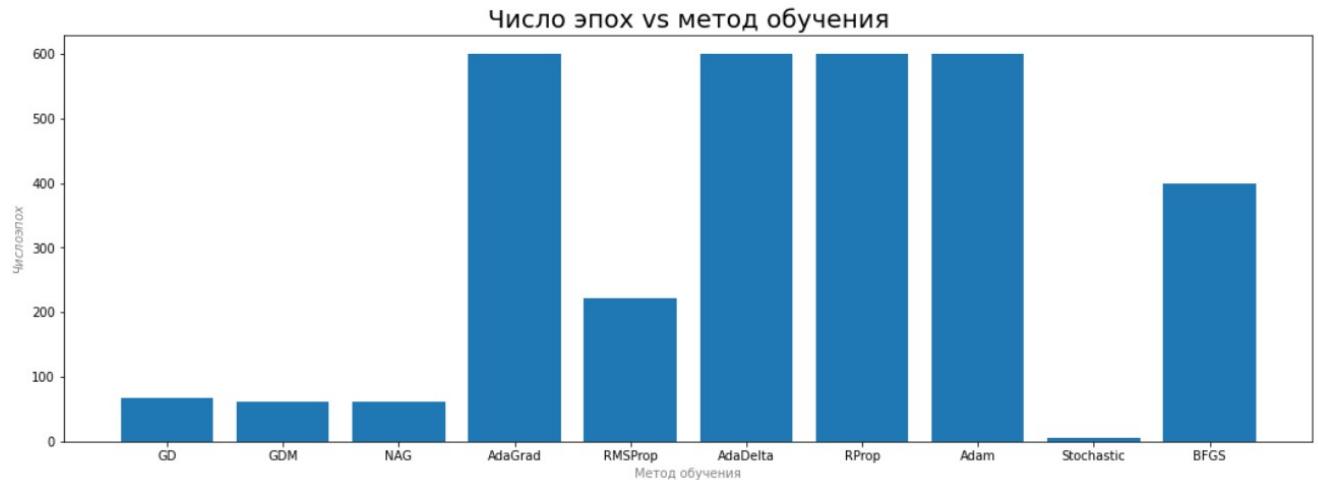
г) Выводы

Чем меньше размер пакета, тем лучше результат.

4.11. Сравнение методов обучения

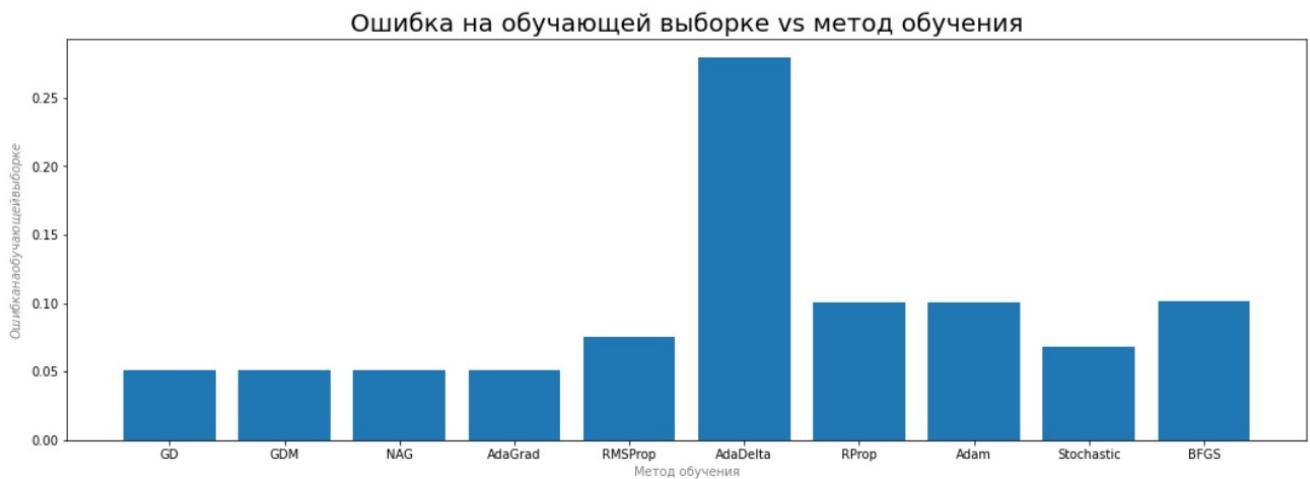
a) Сравнение числа эпох обучения

Построить столбиковую диаграмму числа эпох обучения (по горизонтальной оси – метод обучения (значения параметров методов взять наилучшими по результатам соответствующего исследования), по вертикальной оси – число эпох обучения).



б) Сравнение качества обученных нейросетевых моделей

Построить столбиковую диаграмму ошибок обученной сети на обучающей и тестовой выборках (по горизонтальной оси – метод обучения (значения параметров методов взять наилучшими по результатам соответствующего исследования), по вертикальной оси – ошибки обученной сети).



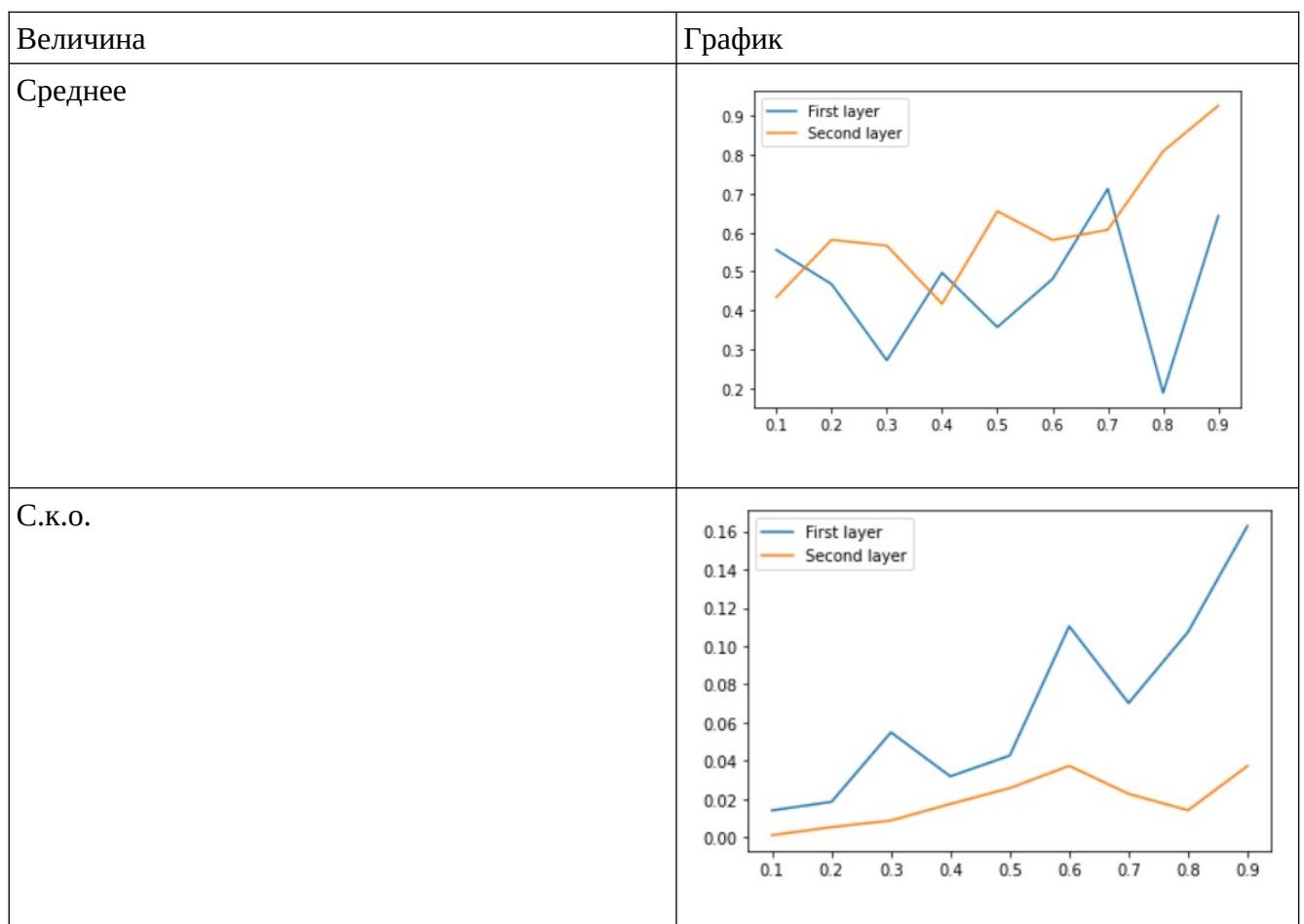
в) Выводы

Лучше всего показал себя Adam (с определенными коэффициентами бета). Хуже всего — AdaDelta.

4.12. Исследование влияния способа инициализации весов на качество обучения

а) Исследование прямого информационного потока в сети

Для произвольно взятых нейронов из каждого слоя построить графики зависимости среднего значения и с.к.о. (по примерам обучающей выборки) выходов при инициализации весов сети случайными числами, распределёнными по нормальному закону $N(0; \sigma)$, от значения с.к.о. σ . Отметить на графике значения σ , рассчитанные по методу Хавьера для различных слоёв сети.



б) Исследование обратного информационного потока в сети

Для тех же нейронов построить графики зависимости среднего значения и с.к.о. (по примерам обучающей выборки) двойственных потенциалов при инициализации весов сети случайными числами, распределёнными поциальному закону $N(0; \sigma)$, от значения с.к.о. σ . Отметить на графике значения σ , рассчитанные по методу Хавьера для различных слоёв сети.

в) Исследование распределений выходов и двойственных потенциалов нейронов

Построить гистограммы распределения выходов и двойственных потенциалов (по примерам обучающей выборки) тех же нейронов, получаемые: 1) при слишком малых значениях σ ; 2) при значениях σ , рассчитанных по методу Хавьера; 3) при слишком больших значениях σ .

г) Кривые обучения

Построить графики зависимости ошибки сети на обучающей и тестовой выборках от времени обучения (кривые обучения) при инициализации весов сети случайными числами, распределёнными по нормальному закону $N(0; \sigma)$: 1) при слишком малых значениях σ ; 2) при значениях σ , рассчитанных по методу Хавьера; 3) при слишком больших значениях σ .

Указание: для обучения использовать простой градиентный метод, параметр скорости обучения выбрать наилучшим по результатам исследований п. 4.2.

д) Заполнить таблицу по результатам обучения

№ п/п	Распределение начальных весов	С.к.о. начальных весов	Средняя ошибка на обучающей выборке \pm с.к.о., $\bar{E}_{обуч} \pm \sigma [E_{обуч}]$	Средняя ошибка на тестовой выборке \pm с.к.о., $\bar{E}_{тест} \pm \sigma [E_{тест}]$
1	Равномерное	0.1	(0.0459, 0.0460)	(0.04573242, 0.0459785)
2	Нормальное		(0.0457809, 0.04587111)	(0.04534607, 0.04543432)
3	Усеченное нормальное		(0.04576394, 0.0458283)	(0.0453348, 0.04539169)
4	Равномерное	0.5	(0.04573366, 0.04573367)	(0.045393811, 0.045393833)
5	Нормальное		(0.042529102, 0.0557317)	(0.04182919, 0.05589583)
6	Усеченное нормальное		(0.0434834, 0.050289936)	(0.042891786, 0.05022656)
7	Равномерное	1	(0.0457337, 0.045733706)	(0.045393854, 0.045393854)
8	Нормальное		(0.05428554, 0.11423449)	(0.053134314, 0.114997165)
9	Усеченное		(0.04646321,	(0.0458374,

	нормальное		0.07474811)	0.07482262)
--	------------	--	-------------	-------------

Указание 1: для обучения использовать простой градиентный метод, параметр скорости обучения выбрать наилучшим по результатам исследований п. 4.2.

Указание 2: для расчета среднего значения и с.к.о. ошибки обученной сети проводить многократное (не менее 10 раз) обучение из различных начальных точек.

e) Выводы

Равномерная инициализация показывает более плохие результаты в сравнении с нормальным и усеченным нормальным. Нормальная и усеченная нормальная инициализация почти не отличаются по эффективности.

4.13. Методы кросс-валидации

a) Заполнить таблицу по результатам кросс-валидации различными методами

Метод кросс-валидации	Число запусков обучения	Средняя ошибка на обучающей выборке \pm с.к.о., $\bar{E}_{обуч} \pm \sigma [E_{обуч}]$	Средняя ошибка на тестовой выборке \pm с.к.о., $\bar{E}_{тест} \pm \sigma [E_{тест}]$
Монте-Карло	10	(-0.010, 0.130)	(-0.009, 0.132)
Holdout 60/30/10	1		
10-fold	1	(0.034, 0.055)	(0.042, 0.093)
LOOCV	1		

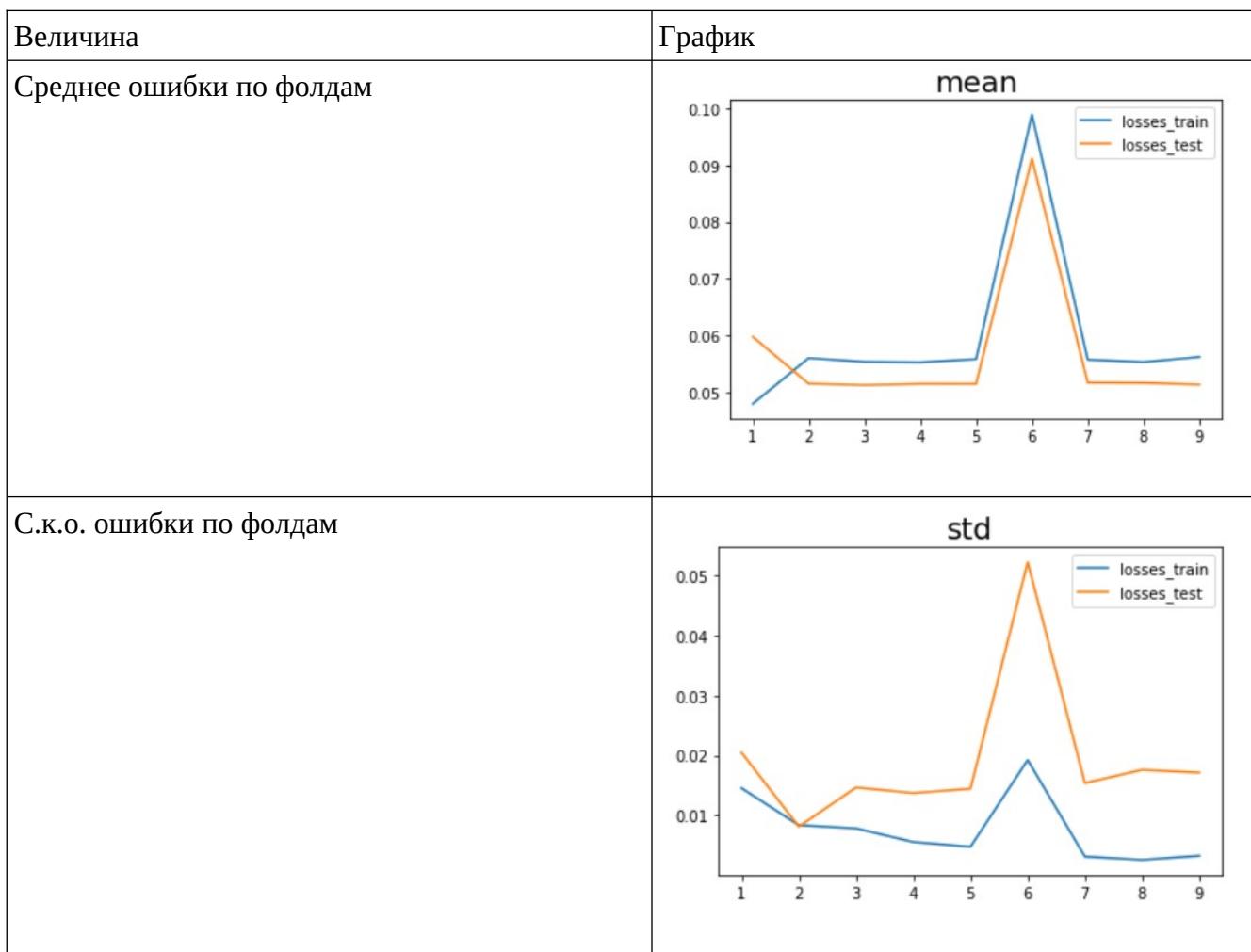
Указание 1: для обучения использовать простой градиентный метод, параметр скорости обучения выбрать наилучшим по результатам исследований п. 4.2.

Указание 2: инициализацию сети провести один раз наилучшим методом по результатам исследований п. 4.12, запуски процедуры обучения каждый раз проводить из одной и той же начальной точки.

Указание 3: Внутри каждого фолда кросс-валидации разбивать обучающие данные на обучающую и валидационную выборки в отношении 70/30.

б) Исследование k -fold кросс-валидации

Построить графики зависимости среднего значения и с.к.о. (по фолдам) ошибки обученной сети на обучающей и тестовой выборках от числа фолдов k .



в) Выводы

Количество фолдов 3-4 оптимально для кросс-валидации методов k-fold.

4.14. Исследование различных архитектур нейронных сетей

a) Исследование зависимости качества обучения от числа нейронов в скрытых слоях

Провести обучение нейронных сетей с различным числом нейронов в скрытых слоях. По результатам обучения заполнить таблицу.

№ п/п	Число нейронов в скрытых слоях	Средняя ошибка на обучающей выборке $\bar{E}_{обуч} \pm \sigma[E_{обуч}]$ с.к.о.,	Средняя ошибка на тестовой выборке $\bar{E}_{тест} \pm \sigma[E_{тест}]$ с.к.о.,
1	5	(0.03, 0.055)	(0.0298, 0.0506)
2	10	(-6310447, 12620896)	(-299834, 599669)
3	13	(-2519, 5039)	(-22, 45)
4	17	(-33129597956,	(-47504457579,

		66259196413)	95008915530)
5	20	(0.0197, 0.0481)	(0.0241, 0.0401)
6	25	(-203231490, 406465086)	(-1914237304, 3828475715)

Указание 1: при выполнении пп. а)–б) для обучения использовать метод сопряженных градиентов, Левенберга-Маркардта или BFGS.

Указание 2: при выполнении пп. а)–б) среднее значение и с.к.о. ошибок на обучающей и тестовой выборках считать по результатам 10-fold кросс-валидации, внутри каждого фолда разбивать обучающие данные на обучающую и валидационную выборки в отношении 70/30.

б) Исследование зависимости качества обучения от активационных характеристик нейронов
Для наилучшей архитектуры, найденной в п. а), провести обучение сети при различных активационных характеристиках нейронов скрытых слоёв. По результатам обучения заполнить таблицу.

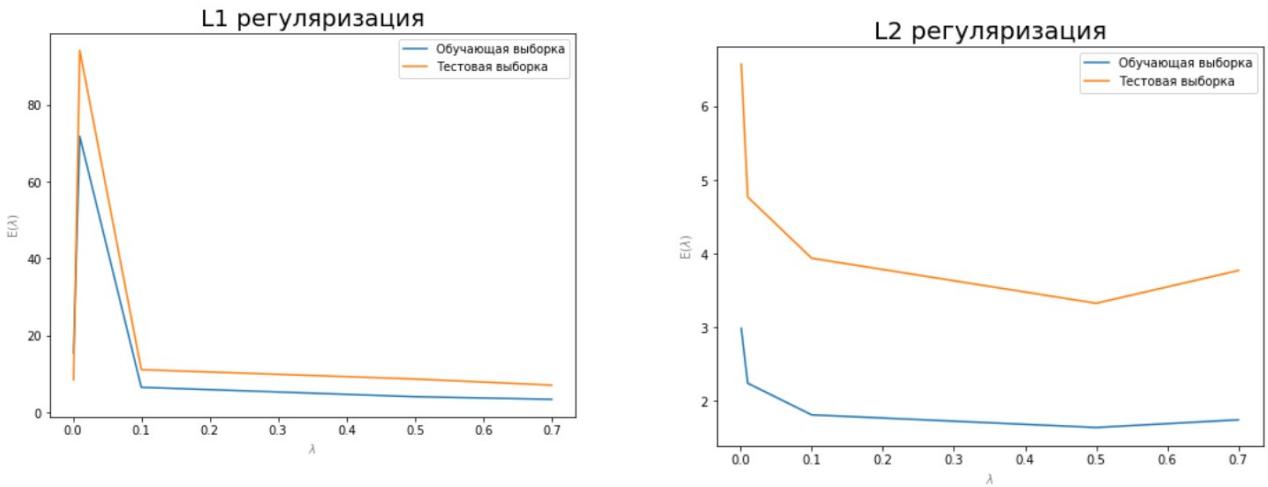
AX нейронов скрытых слоёв	Средняя ошибка на обучающей выборке \pm с.к.о., $\bar{E}_{\text{обуч}} \pm \sigma[E_{\text{обуч}}]$	Средняя ошибка на тестовой выборке \pm с.к.о., $\bar{E}_{\text{тест}} \pm \sigma[E_{\text{тест}}]$
logistic	(0.044, 0.049)	(0.039, 0.061)
tanh	(0.019, 0.033)	(0.017, 0.045)
linear	(13.5, 50.4)	(11.1, 98.4)
softsign	(0.007, 0.021)	(0.012, 0.026)
softplus	(1.78, 4.58)	(1.62, 8.83)

4.15. L_1 и L_2 регуляризация весов

а) Исследование зависимости качества обученной сети от параметра регуляризации

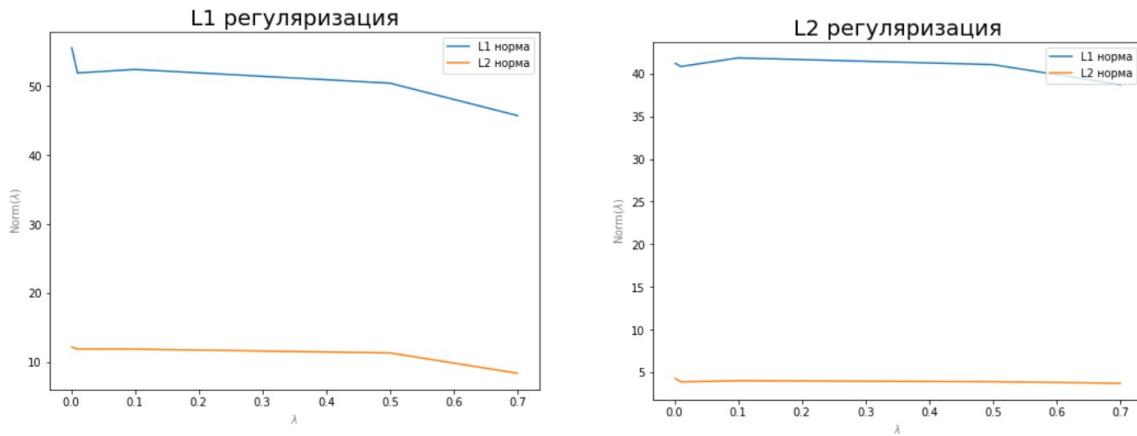
Построить график зависимости ошибки обученной сети на обучающей и тестовой выборках от значения параметра λ : 1) при L_1 -регуляризации весов; 2) при L_2 -регуляризации весов.

Указание: исследования в пп. а)–в) проводить для лучшей архитектуры, найденной в п. 4.14, для обучения использовать один из градиентных методов с адаптивным шагом, параметры метода выбрать наилучшими по результатам исследований пп. 4.5–4.7.



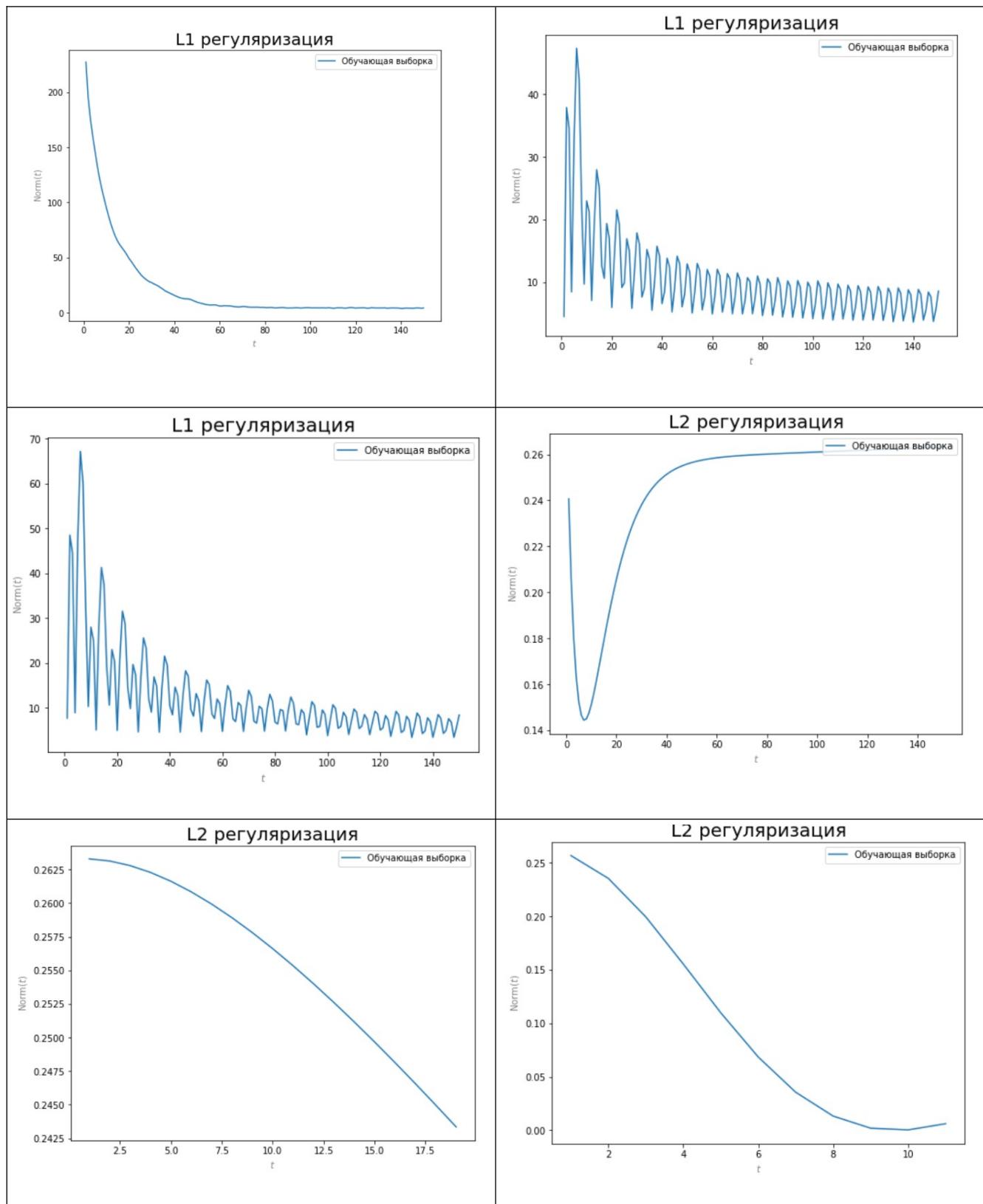
б) Исследование зависимости нормы вектора синаптических коэффициентов от параметра регуляризации

Построить графики зависимости L_1 и L_2 нормы вектора синаптических коэффициентов обученной сети от параметра λ : 1) при L_1 -регуляризации; 2) при L_2 -регуляризации.



в) Сравнение кривых обучения при различных значениях параметра регуляризации

При фиксированных значениях параметра λ ($\lambda \approx 0$, $\lambda > 0$, $\lambda \gg 0$) построить графики зависимости L_1 и L_2 нормы вектора синаптических коэффициентов сети от времени обучения.



г) Выводы

Чем больше коэффициент λ , тем меньше ошибка при обучении модели. При L2 есть вероятность, что ошибка увеличится.

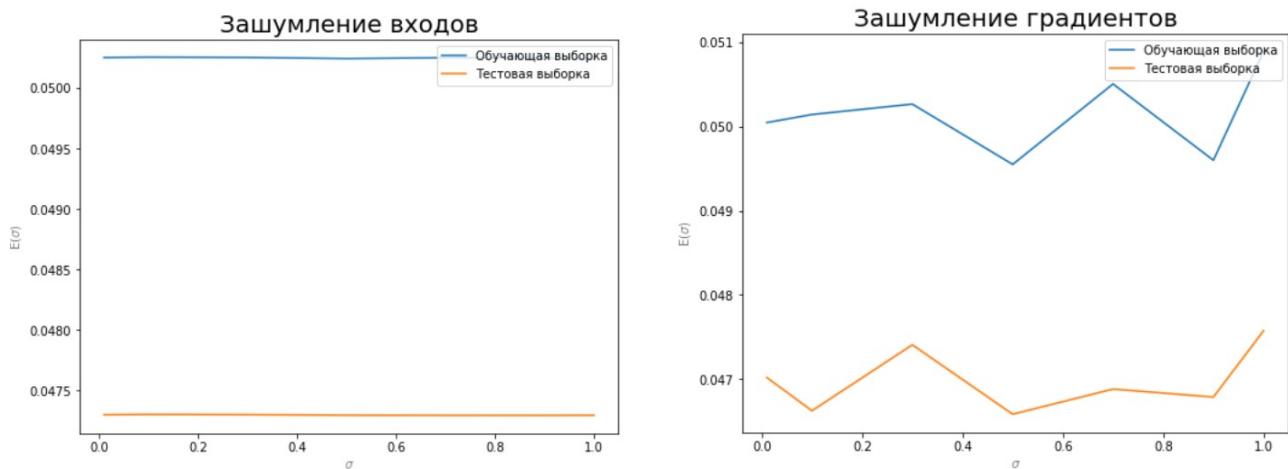
4.16. Инъекция шума

а) Исследование зависимости качества обученной сети от параметра зашумления

Построить графики зависимости ошибки обученной сети на обучающей и тестовой выборках от с.к.о. σ шума при: 1) зашумлении входов; 2) зашумлении градиентов.

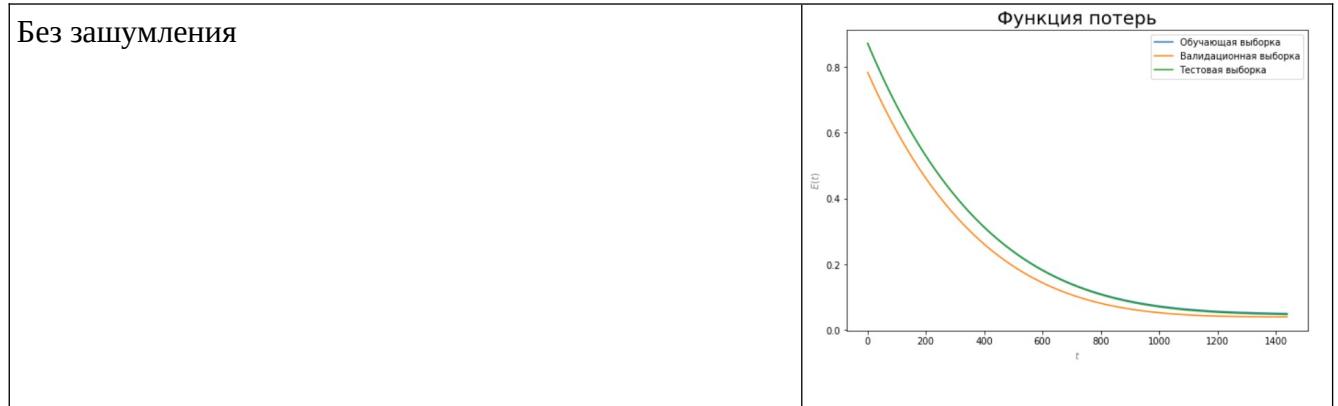
Указание 1: шум генерировать из нормального распределения $N(0; \sigma)$.

Указание 2: исследования в пп. а)–б) проводить для лучшей архитектуры, найденной в п. 4.14, для обучения использовать один из градиентных методов с адаптивным шагом, параметры метода выбрать наилучшими по результатам исследований пп. 4.5–4.7.

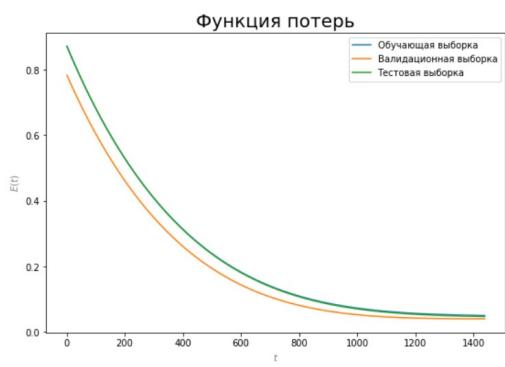


б) Сравнение кривых обучения

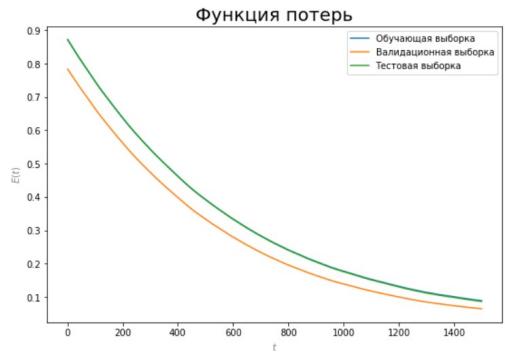
Построить кривые обучения на обучающей, валидационной и тестовой выборках при обучении: 1) без зашумления; 2) с зашумлением входов; 3) с зашумлением градиентов; 4) с зашумлением входов и градиентов.



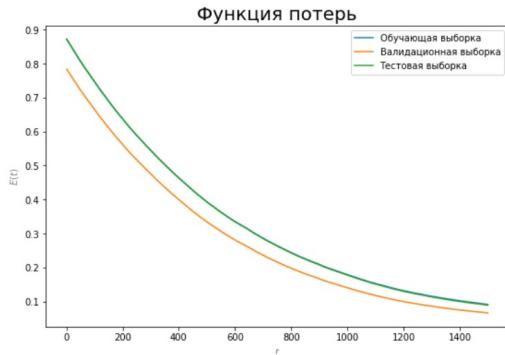
С зашумлением входов



С зашумлением градиентов



С зашумлением входов и градиентов



Указание: шум генерировать из нормального распределения $N(0; \sigma)$, использовать с.к.о. шума σ , при котором наблюдается наименьшая ошибка сети (по результатам исследований в п. а)) либо задать произвольно.

в) Выводы

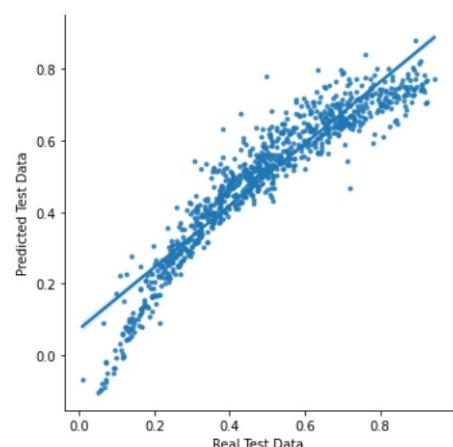
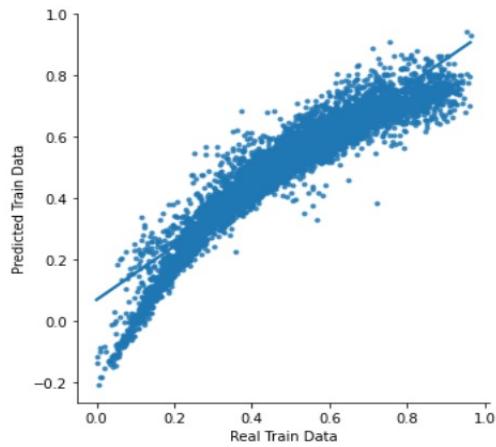
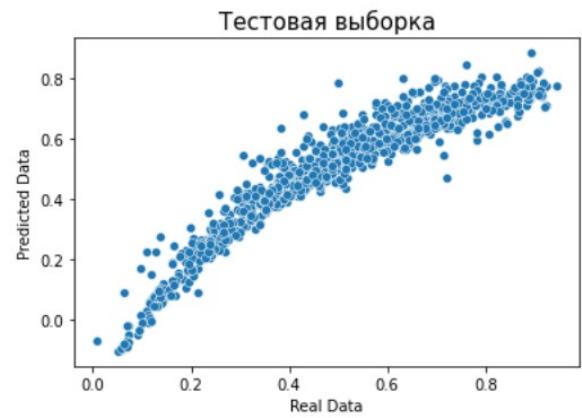
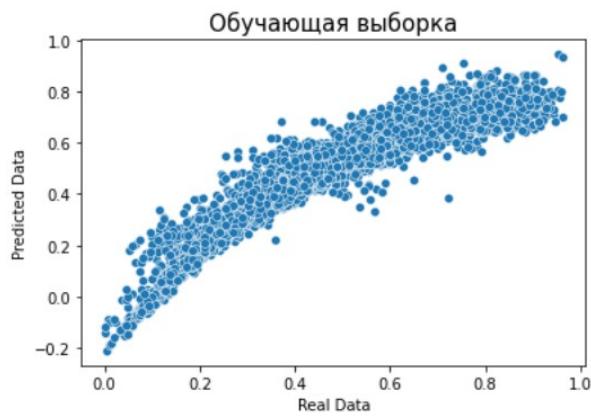
При зашумлении (особенно при одновременном зашумлении градиента и входа) модель обучается медленнее, однако так же сходится в точку локального минимума.

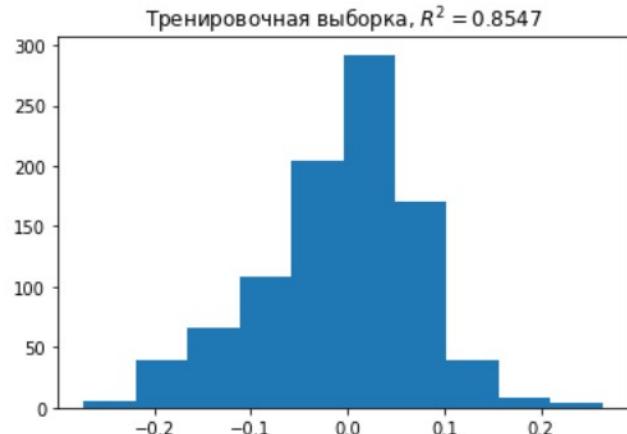
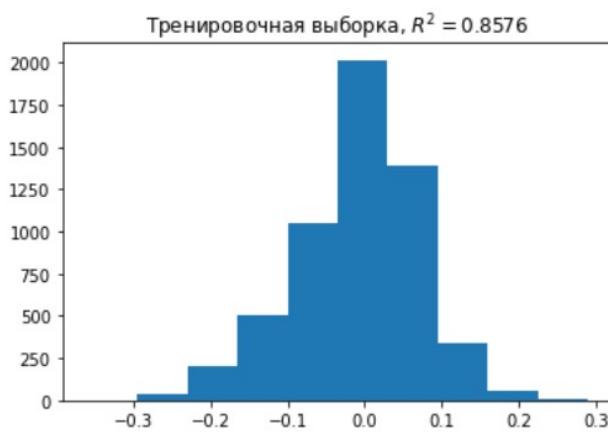
V. Исследование обученной нейросетевой модели

Указание: среди всех построенных в пп. 4.2–4.16 нейросетевых моделей выбрать модель с наименьшими значениями ошибок на обучающей и валидационной выборках.

5.1. Исследование качества обученной модели

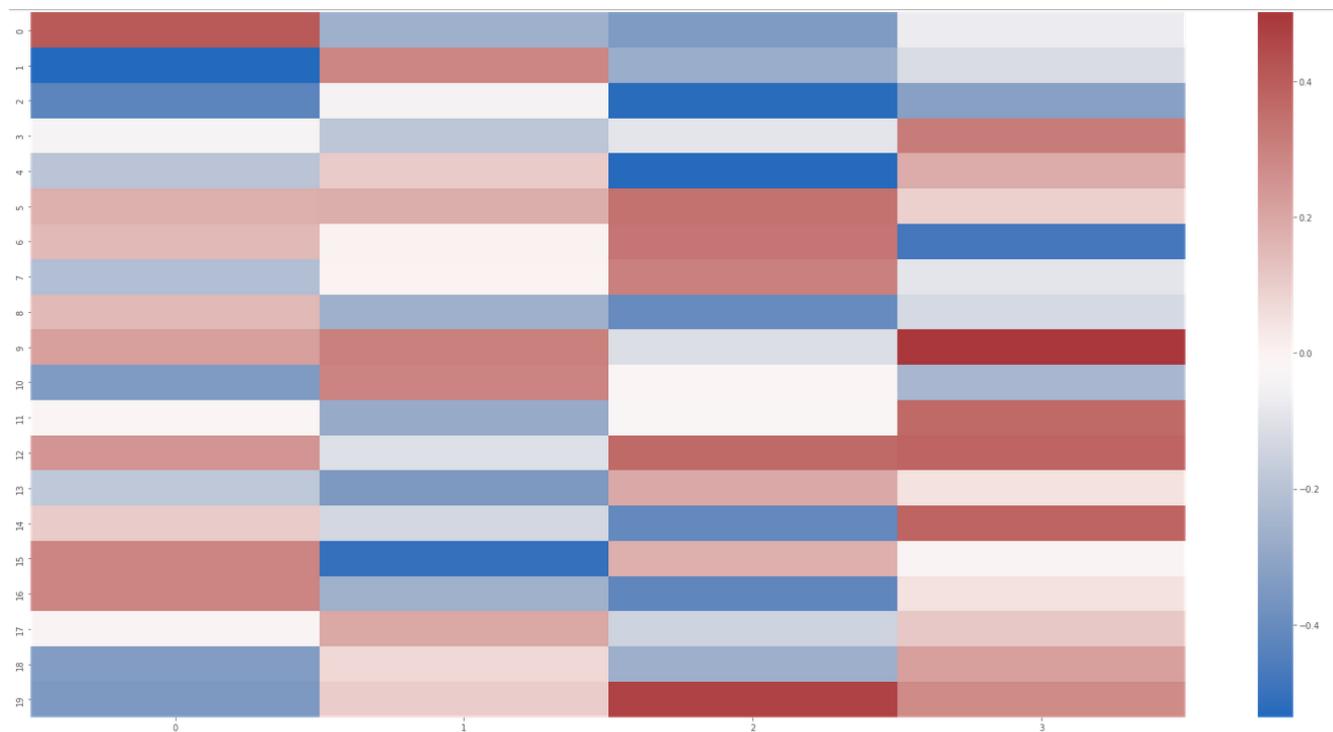
Для задач регрессии: построить диаграммы рассеяния в пространстве «выход модели – желаемый выход» для обучающей и тестовой выборок, построить линейные регрессии выхода модели на желаемый выход, рассчитать коэффициенты детерминации линейных регрессионных моделей для обучающей и тестовой выборок. Построить гистограммы распределения ошибок сети на примерах обучающей и тестовой выборок.





5.2. Оценка важности признаков

Визуализировать матрицу синаптических коэффициентов 1-го слоя обученной сети (использовать heatmap). Сделать предположения о важности используемых признаков для решения поставленной задачи по результатам визуального анализа.



5.3. Выводы

Построенная нейросетевая модель применима в рамках данной задачи с коэффициентом детерминации $R^2 = 0.8576$ на обучающей и $R^2 = 0.8547$ на тестовой выборках. Ошибки сети на примерах обучающей и тестовой выборок распределены нормально, следовательно, построенная модель хорошо описывает истинную зависимость.