



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ _____ «Информатика и системы управления»

КАФЕДРА _____ «Теоретическая информатика и компьютерные технологии»

Домашняя работа № 1
по курсу «Теория искусственных нейронных сетей»
«Реализация однослойного персептрона»

Студент группы ИУ9-71Б Афанасьев И.

Преподаватель Каганов Ю. Т.

Москва 2024

1 Цель работы

1. Изучение многослойного персептрона, исследование его работы на основе использования градиентного спуска.

2 Постановка задачи

- Реализовать на языке высокого уровня многослойный персептрон и проверить его работоспособность на примере данных, выбранных из MNIST dataset.
- Исследовать работу персептрона на основе использования различных целевых функций (среднеквадратичная ошибка, перекрестная энтропия, дивергенция Кульбака-Лейблера).
- Провести исследование эффективности работы многослойного персептрона при изменении гиперпараметров (количества нейронов и количества слоев).
- Подготовить отчет с распечаткой текста программы, графиками результатов исследования и анализом результатов.

3 Реализация

4 Результаты экспериментов

На протяжении всех экспериментов значения некоторых гиперпараметров были фиксированными:

- количество эпох: 200;
- размер пакета данных (mini-batch): 100;
- коэффициент обучения (learning rate): 0.025.

Изменялись значения следующих гиперпараметров:

- количество скрытых слоёв: 1 или 3;
- количество нейронов в скрытых слоях: 20 или 40.

Таким образом, для каждой функции ошибки — среднеквадратичной, перекрёстной энтропии и дивергенции Кульбака-Лейблера — рассматривались четыре конфигурации с разным числом скрытых слоёв и нейронов в них. В скрытых слоях всегда используется функция активации Leaky ReLU, на выходном слое — Softmax.

4.1 Среднеквадратичная ошибка (MSE)

4.1.1 1 скрытый слой, 20 нейронов

На рисунке¹ 1 изображено изменение функции ошибки за период обучения персептрона. Здесь и далее график синей функции соответствует данным для обучения, график оранжевой функции — тестовым данным. В результате, ошибка на данных для обучения составляет 0.166286, ошибка на тестовых данных — 0.193274.

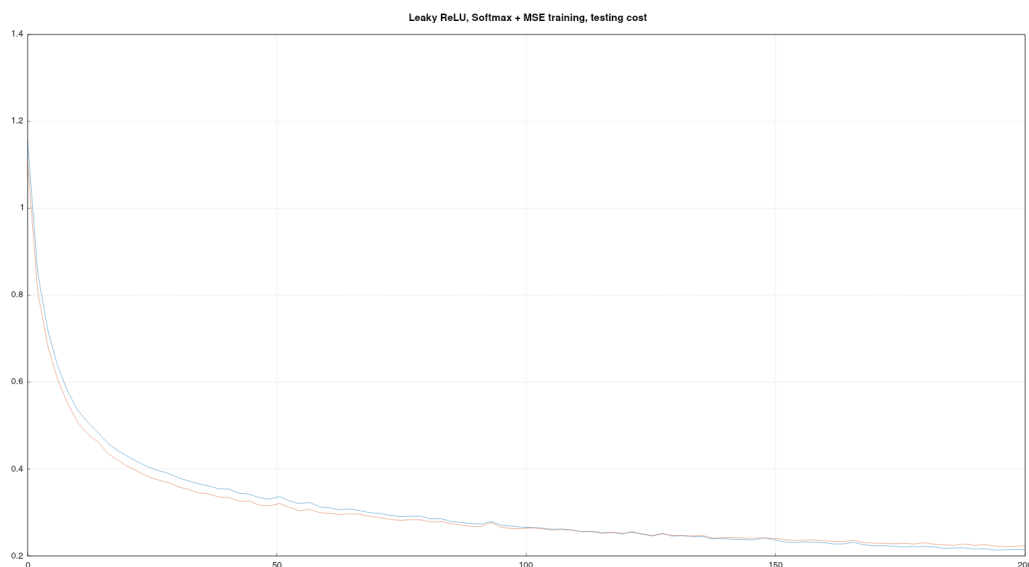


Рис. 1

¹На некоторых рисунках шкала оси *Oy* начинается с 0.2, а не 0.0.

На рисунке 2 изображено изменение точности за период обучения персептрона. В результате, точность на данных для обучения составляет $\frac{47518}{50000} \approx 95,04\%$, на тестовых данных — $\frac{9421}{10000} = 94,21\%$.

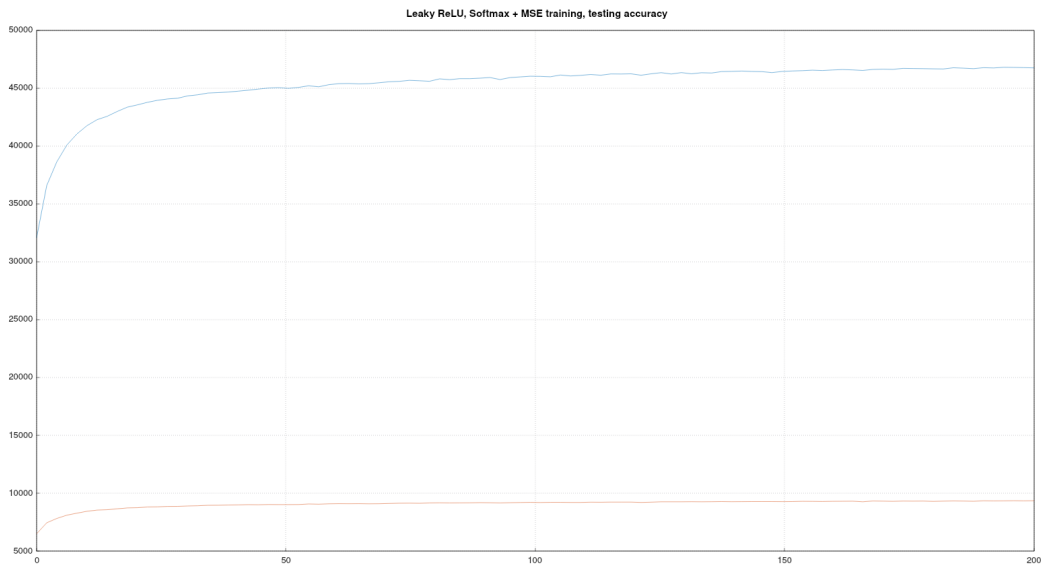


Рис. 2

4.1.2 1 скрытый слой, 40 нейронов

На рисунке 3 изображено изменение функции ошибки за период обучения персептрона. В результате, ошибка на данных для обучения составляет 0.123245, ошибка на тестовых данных — 0.161724.

На рисунке 4 изображено изменение точности за период обучения персептрона. В результате, точность на данных для обучения составляет $\frac{48121}{50000} \approx 96,24\%$, на тестовых данных — $\frac{9527}{10000} = 95,27\%$.

4.1.3 3 скрытых слоя, 20 нейронов

На рисунке 5 изображено изменение функции ошибки за период обучения персептрона. В результате, ошибка на данных для обучения составляет 0.163502, ошибка на тестовых данных — 0.206251.

На рисунке 6 изображено изменение точности за период обучения персептрона. В результате, точность на данных для обучения составляет $\frac{47504}{50000} \approx 95,01\%$, на тестовых данных — $\frac{9416}{10000} = 94,16\%$.

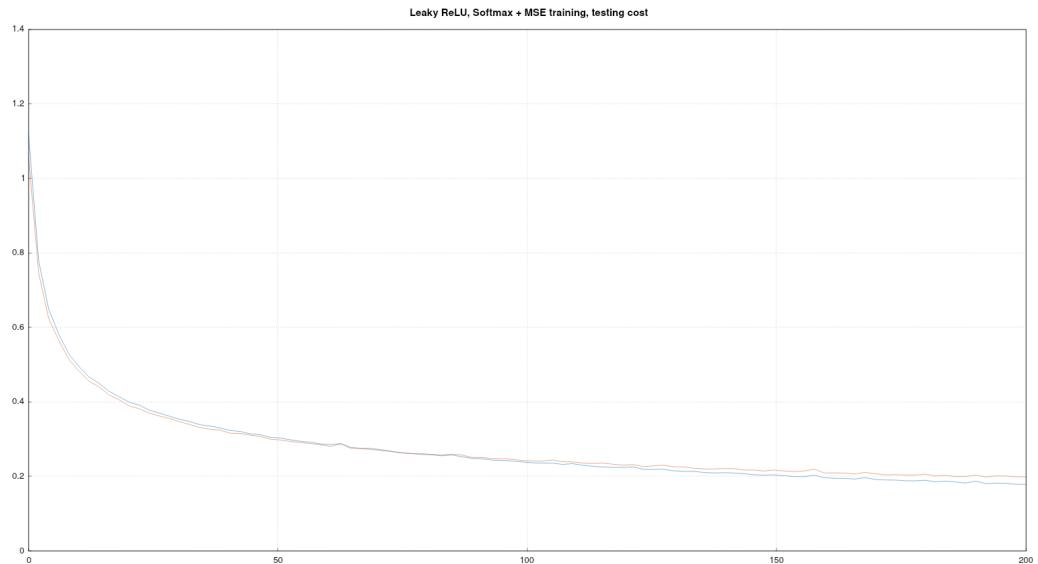


Рис. 3

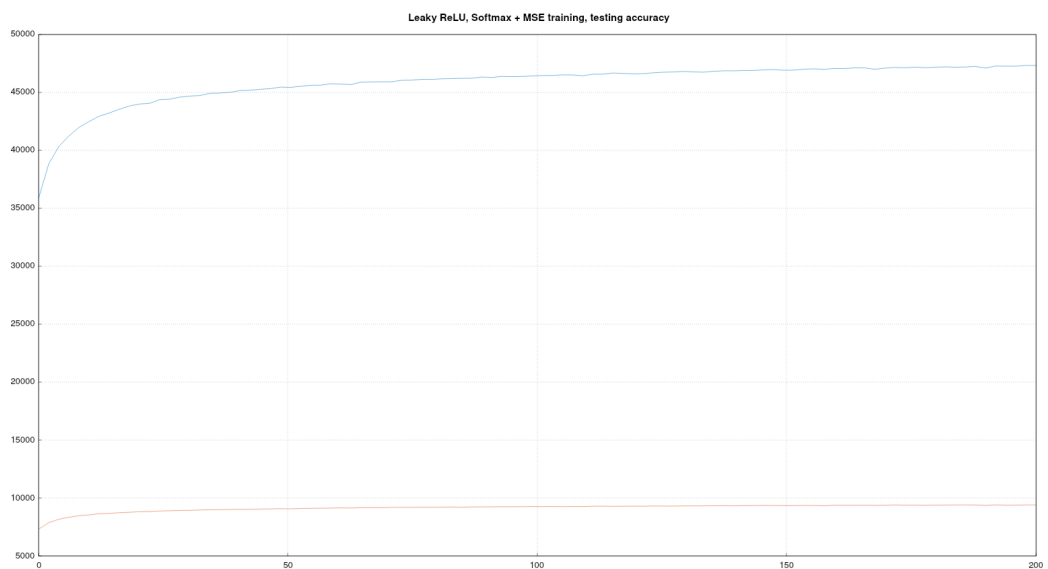


Рис. 4

4.1.4 3 скрытых слоя, 40 нейронов

На рисунке 7 изображено изменение функции ошибки за период обучения персептрона. В результате, ошибка на данных для обучения составляет 0.105351, ошибка на тестовых данных — 0.190718.

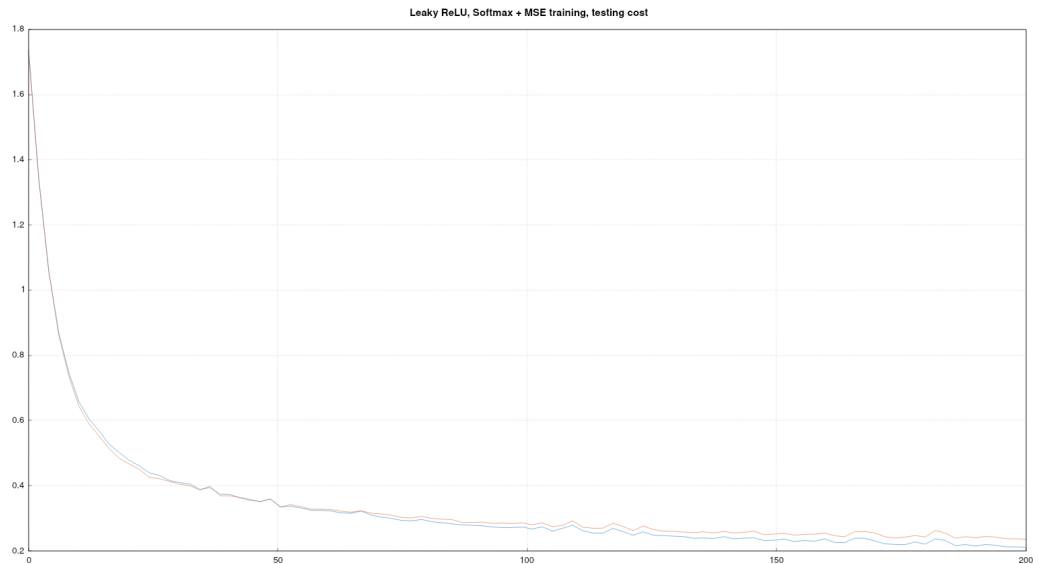


Рис. 5

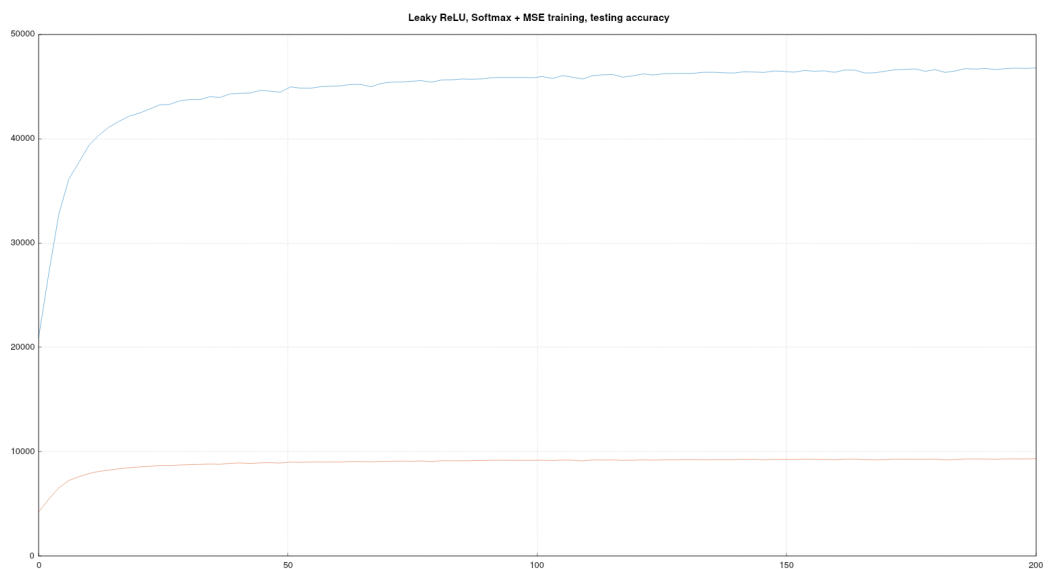


Рис. 6

На рисунке 8 изображено изменение точности за период обучения персептрона. В результате, точность на данных для обучения составляет $\frac{48422}{50000} \approx 96,84\%$, на тестовых данных — $\frac{9481}{10000} = 94,81\%$.

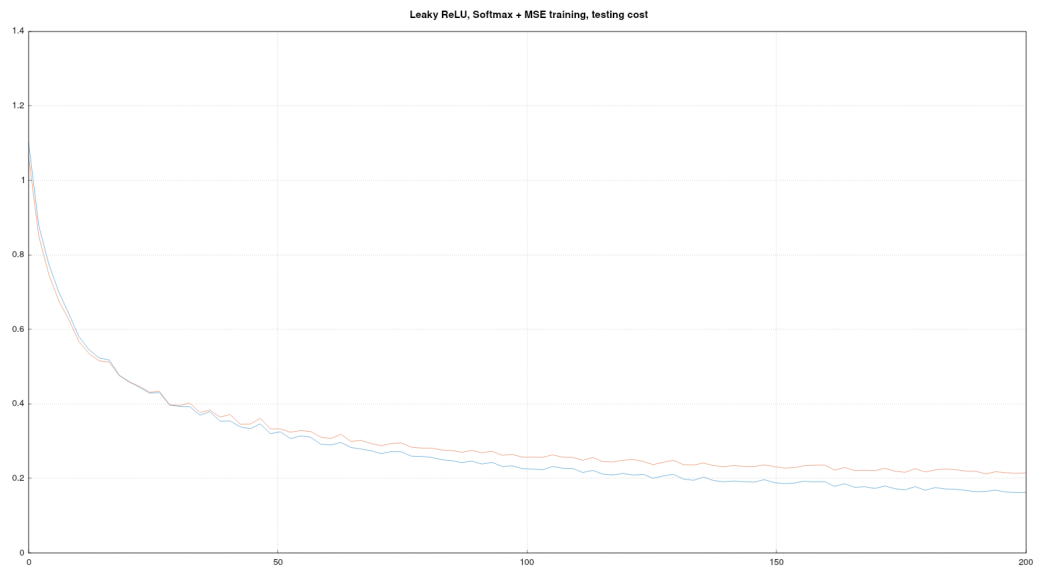


Рис. 7

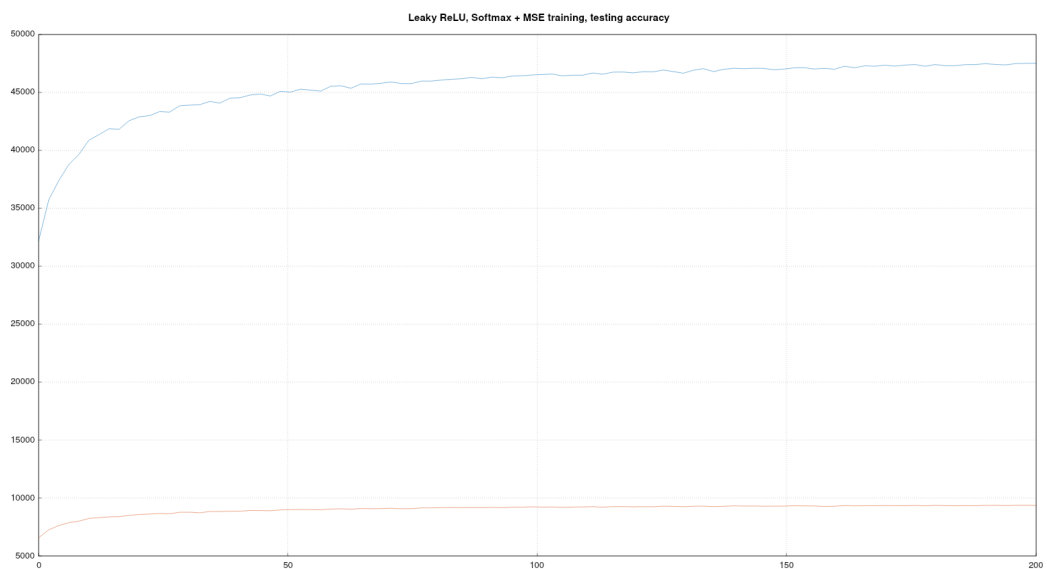


Рис. 8

4.2 Перекрёстная энтропия

4.2.1 1 скрытый слой, 20 нейронов

На рисунке 9 изображено изменение функции ошибки за период обучения персептрона. В результате, ошибка на данных для обучения составляет 0.192402, ошибка на тестовых данных — 0.217352.

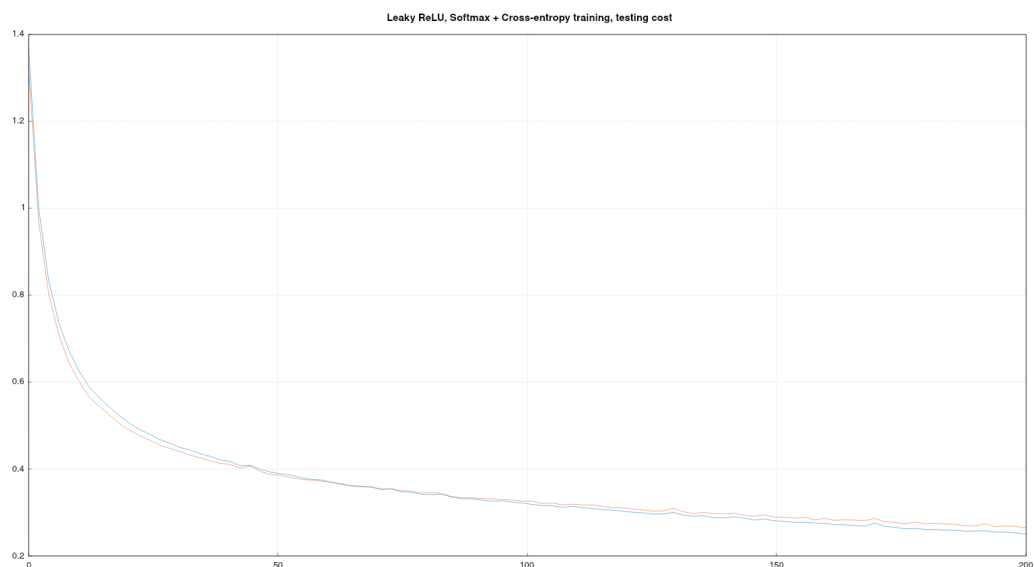


Рис. 9

На рисунке 10 изображено изменение точности за период обучения персептрона. В результате, точность на данных для обучения составляет $\frac{47194}{50000} \approx 94,34\%$, на тестовых данных — $\frac{9345}{10000} = 93,45\%$.

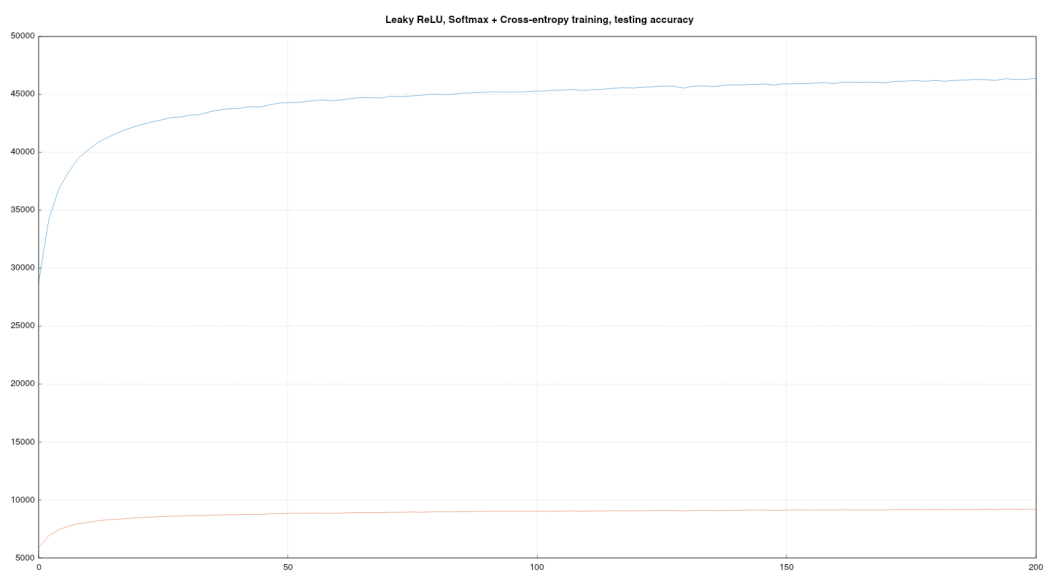


Рис. 10

4.2.2 1 скрытый слой, 40 нейронов

На рисунке 11 изображено изменение функции ошибки за период обучения персептрона. В результате, ошибка на данных для обучения составляет 0.131022, ошибка на тестовых данных — 0.173189.

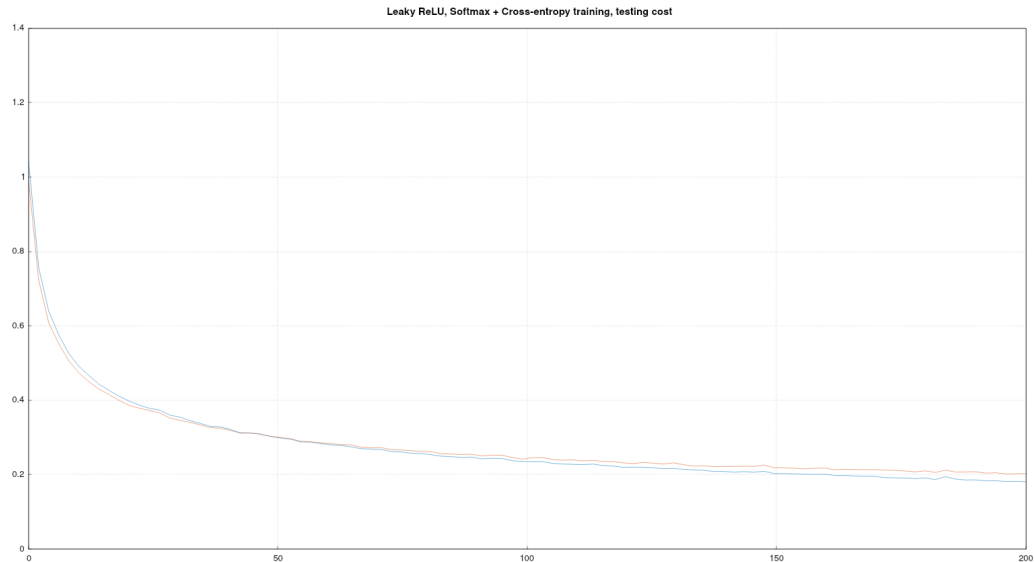


Рис. 11

На рисунке 12 изображено изменение точности за период обучения персептрона. В результате, точность на данных для обучения составляет $\frac{48097}{50000} \approx 96,19\%$, на тестовых данных — $\frac{9483}{10000} = 94,83\%$.

4.2.3 3 скрытых слоя, 20 нейронов

На рисунке 13 изображено изменение функции ошибки за период обучения персептрона. В результате, ошибка на данных для обучения составляет 0.157613, ошибка на тестовых данных — 0.216353.

На рисунке 14 изображено изменение точности за период обучения персептрона. В результате, точность на данных для обучения составляет $\frac{47681}{50000} \approx 95,36\%$, на тестовых данных — $\frac{9386}{10000} = 93,86\%$.

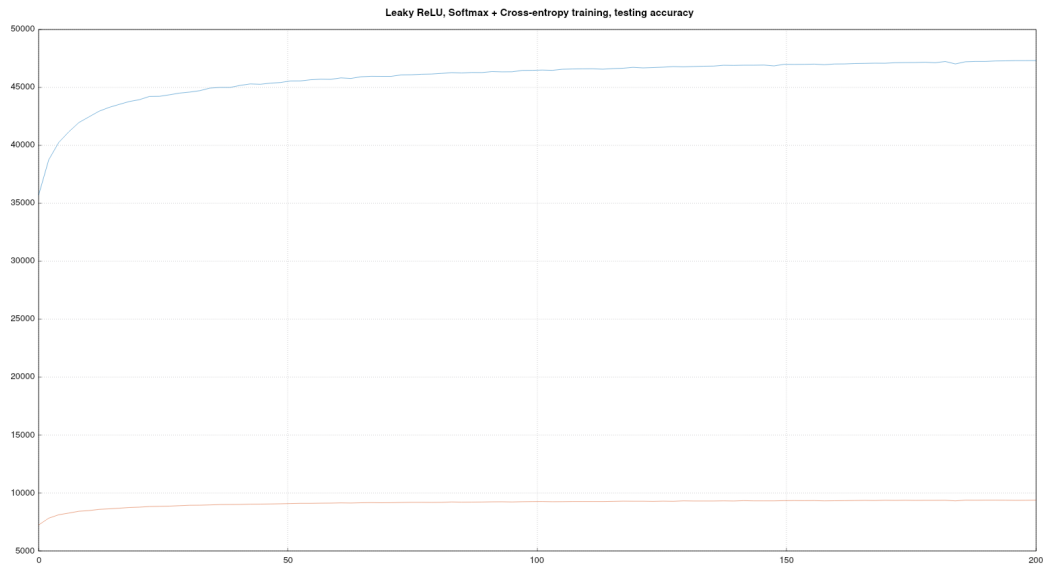


Рис. 12

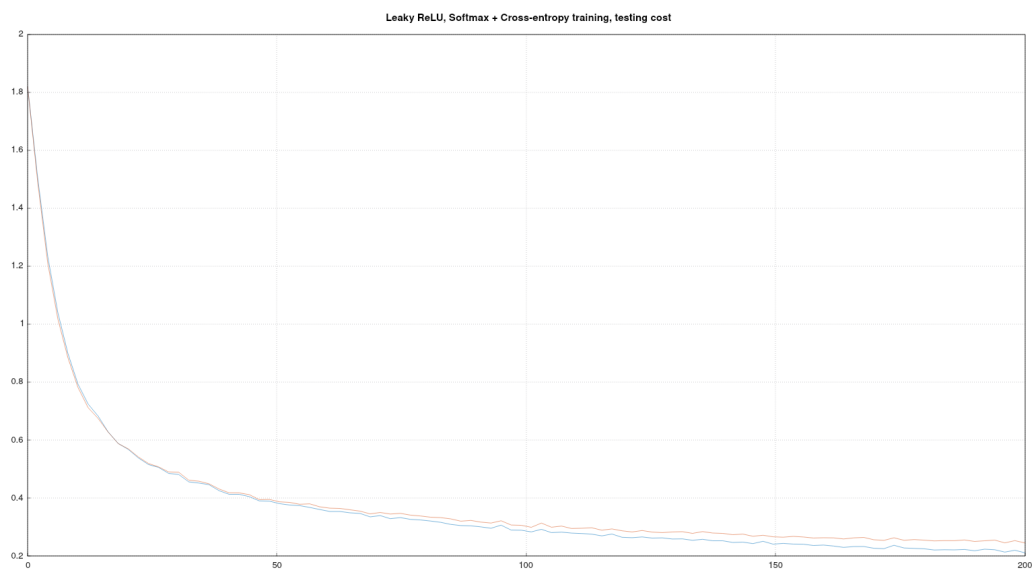


Рис. 13

4.2.4 3 скрытых слоя, 40 нейронов

На рисунке 15 изображено изменение функции ошибки за период обучения персептрона. В результате, ошибка на данных для обучения составляет 0.118457, ошибка на тестовых данных — 0.17582.

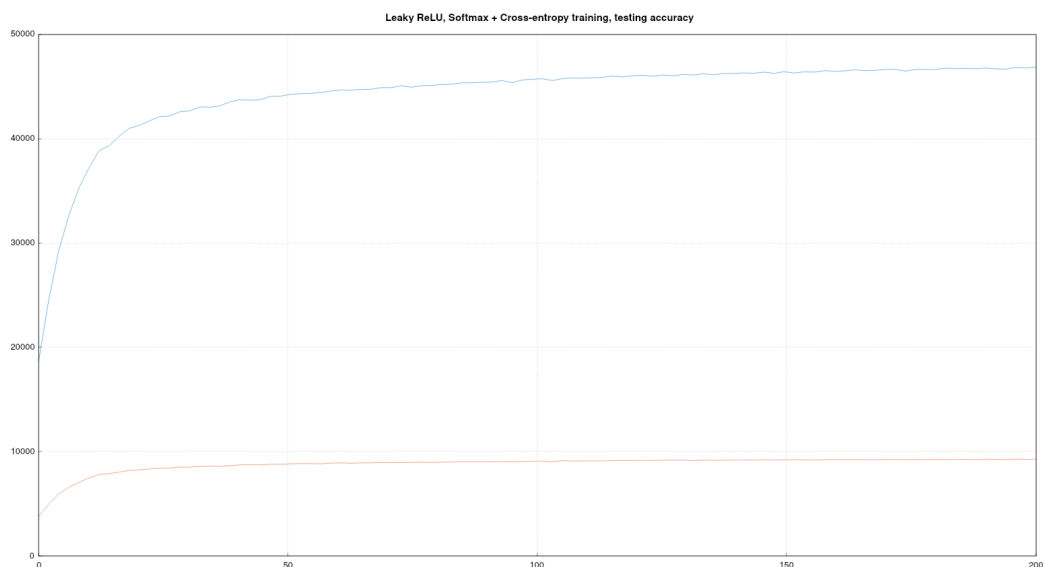


Рис. 14

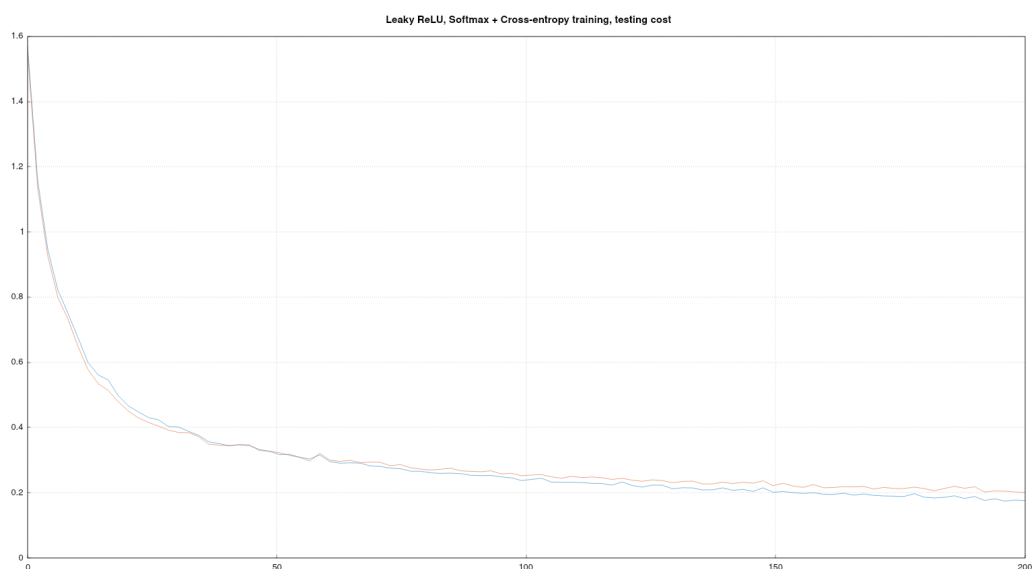


Рис. 15

На рисунке 16 изображено изменение точности за период обучения персептрона. В результате, точность на данных для обучения составляет $\frac{48232}{50000} \approx 96,46\%$, на тестовых данных — $\frac{9502}{10000} = 95,02\%$.

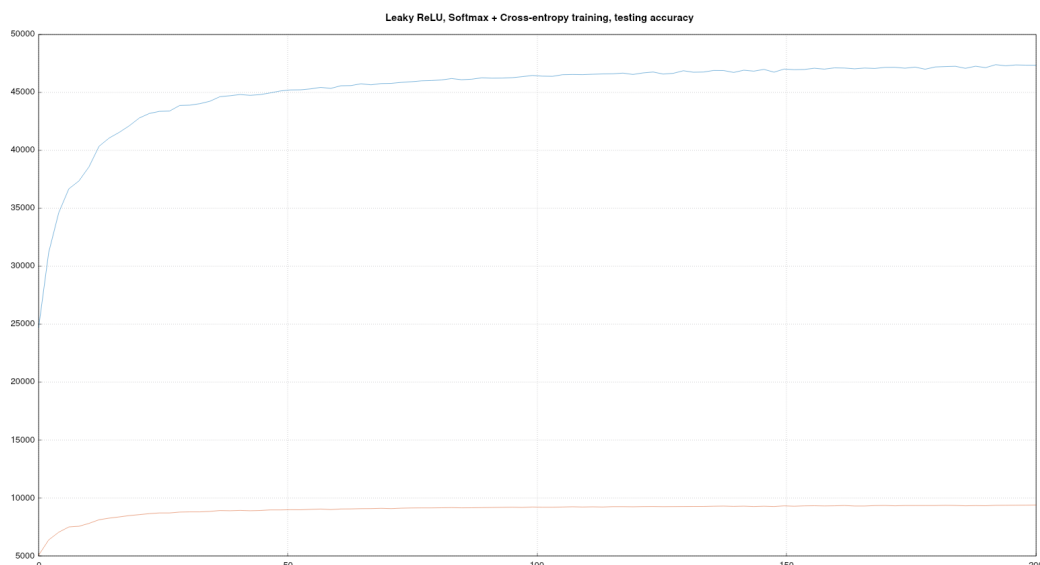


Рис. 16

4.3 Дивергенция Кульбака-Лейблера

4.3.1 1 скрытый слой, 20 нейронов

На рисунке 17 изображено изменение функции ошибки за период обучения персептрона. В результате, ошибка на данных для обучения составляет 0.17601, ошибка на тестовых данных — 0.201559.

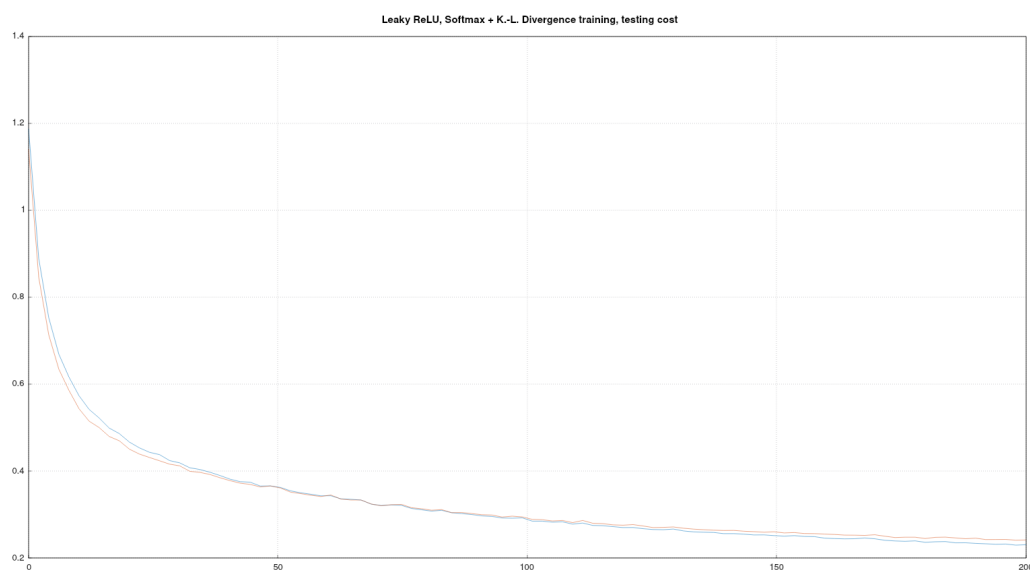


Рис. 17

На рисунке 18 изображено изменение точности за период обучения персептрона. В результате, точность на данных для обучения составляет $\frac{47346}{50000} \approx 94,69\%$, на тестовых данных — $\frac{9415}{10000} = 94,15\%$.

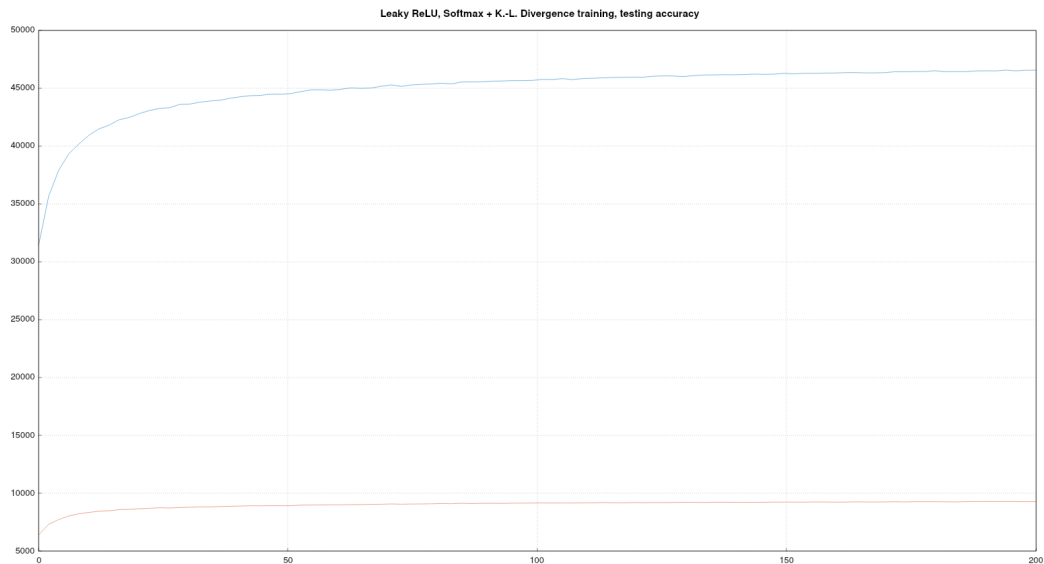


Рис. 18

4.3.2 1 скрытый слой, 40 нейронов

На рисунке 19 изображено изменение функции ошибки за период обучения персептрона. В результате, ошибка на данных для обучения составляет 0.118821, ошибка на тестовых данных — 0.168346.

На рисунке 20 изображено изменение точности за период обучения персептрона. В результате, точность на данных для обучения составляет $\frac{48233}{50000} \approx 96,47\%$, на тестовых данных — $\frac{9506}{10000} = 95,06\%$.

4.3.3 3 скрытых слоя, 20 нейронов

На рисунке 21 изображено изменение функции ошибки за период обучения персептрона. В результате, ошибка на данных для обучения составляет 0.163792, ошибка на тестовых данных — 0.213412.

На рисунке 22 изображено изменение точности за период обучения персептрона. В результате, точность на данных для обучения составляет $\frac{47487}{50000} \approx 94,97\%$, на тестовых данных — $\frac{9369}{10000} = 93,69\%$.

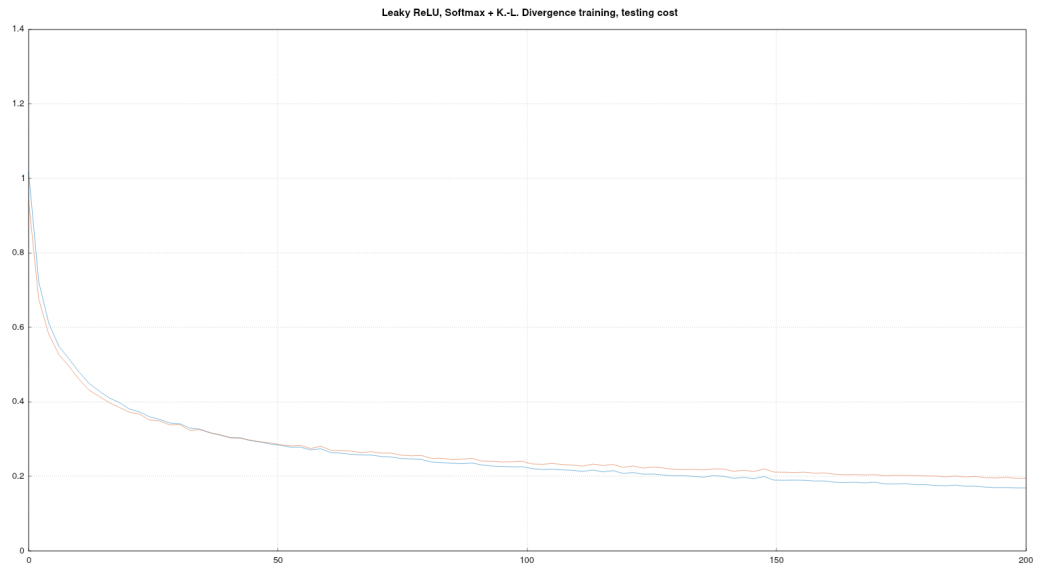


Рис. 19

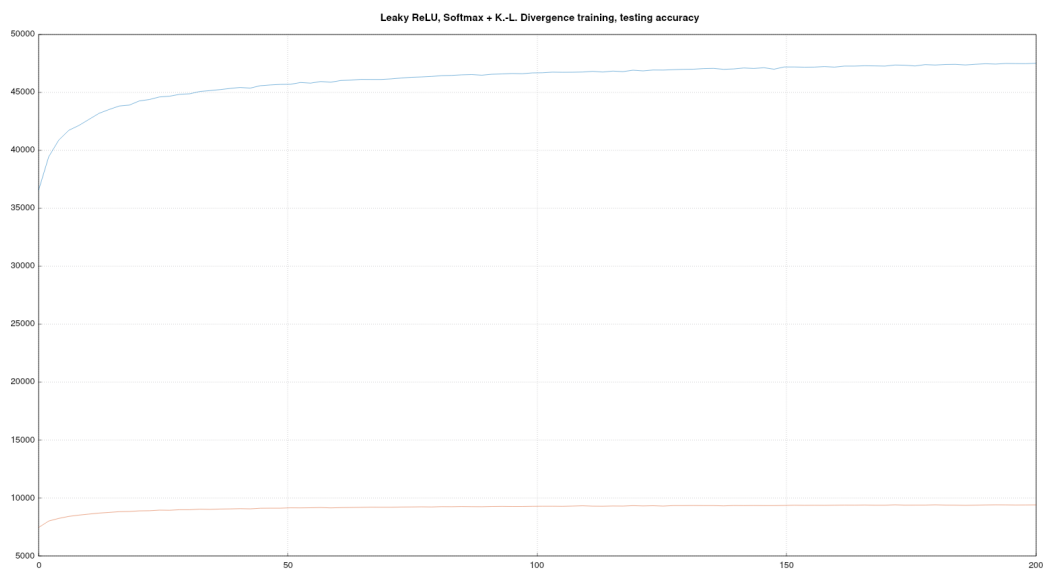


Рис. 20

4.3.4 3 скрытых слоя, 40 нейронов

На рисунке 23 изображено изменение функции ошибки за период обучения персептрона. В результате, ошибка на данных для обучения составляет 0.111865, ошибка на тестовых данных — 0.197848.

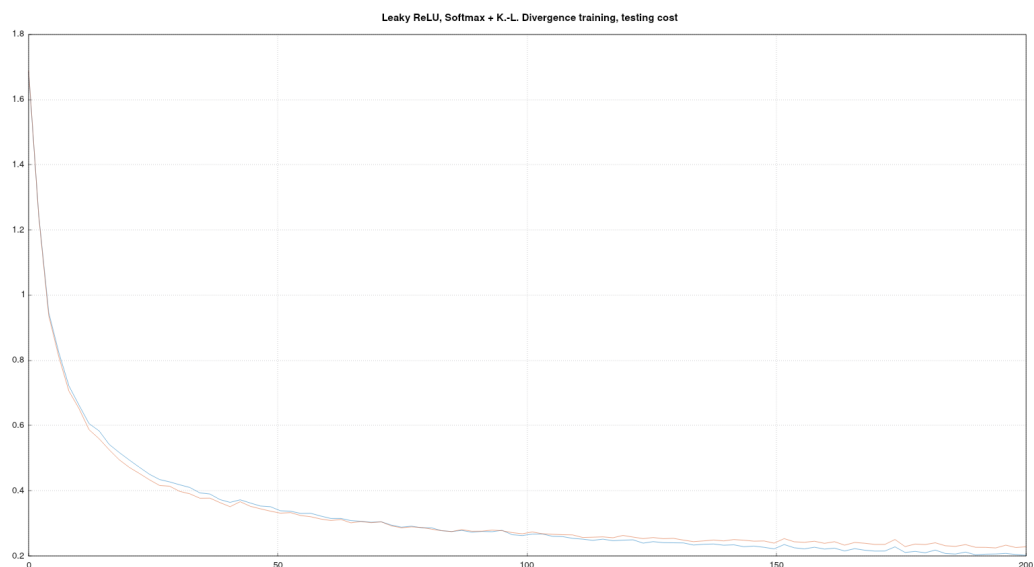


Рис. 21

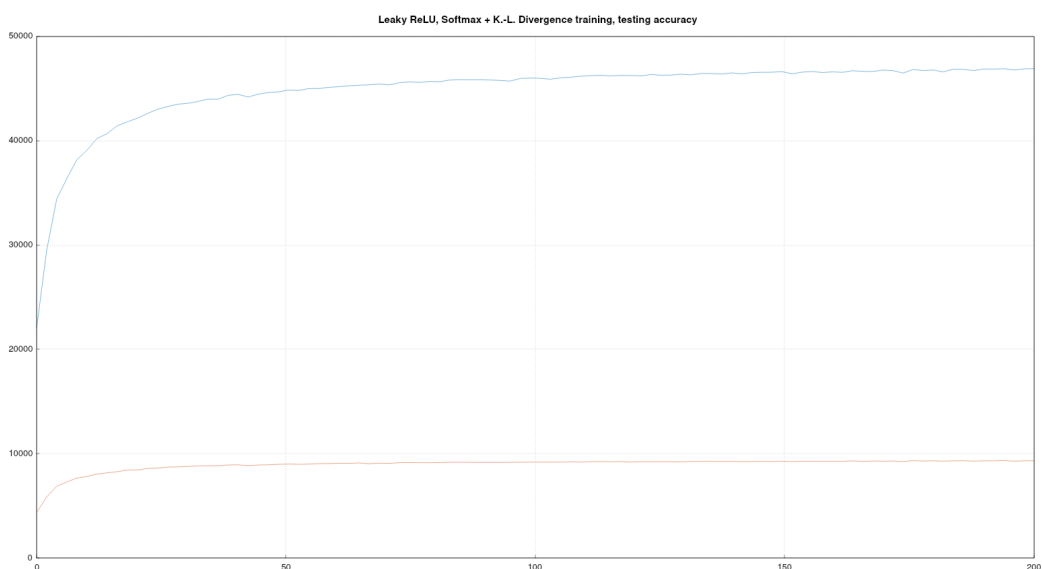


Рис. 22

На рисунке 24 изображено изменение точности за период обучения персептрона. В результате, точность на данных для обучения составляет $\frac{48315}{50000} \approx 96,63\%$, на тестовых данных — $\frac{9457}{10000} = 94,57\%$.

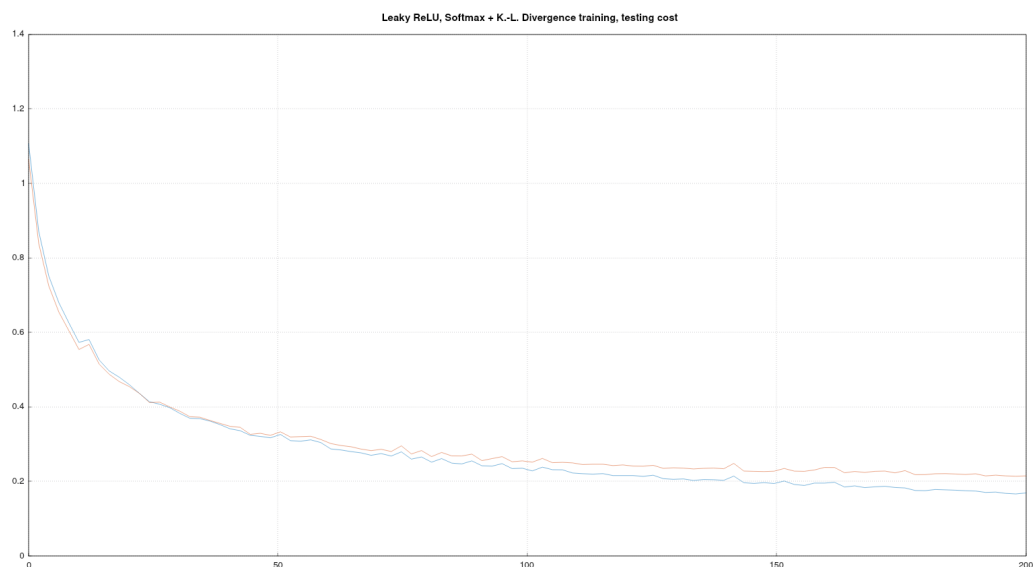


Рис. 23

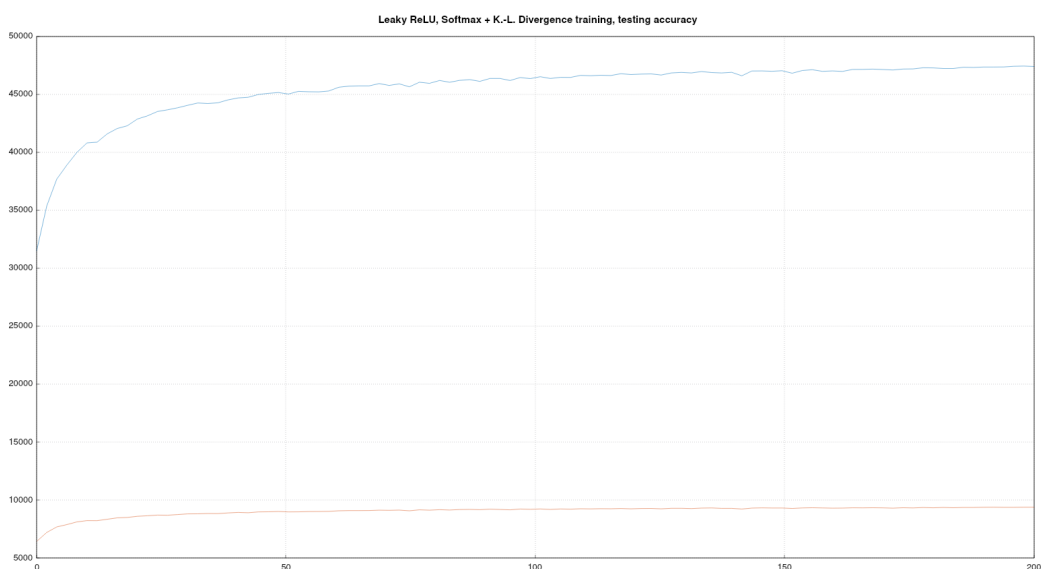


Рис. 24

5 Вывод

Исходя из полученных результатов видно, что наименьшая ошибка на тестовых данных достигается при наличии 1-го скрытого слоя с 40 нейронами: для среднеквадратичной функции, перекрёстной энтропии, дивергенции Кульбака-Лейблера точность на тестовых данных составляет 95,27%, 94,83% и 95,06%

соответственно. Увеличение числа нейронов в скрытых слоях может способствовать уменьшению ошибки и повышению точности персептрона.

При 3-х скрытых слоях с 40 нейронами ошибка на данных для обучения меньше, чем в предыдущем случае, однако ошибка на тестовых данных выше — это может свидетельствовать о переобучении персептрона; более простая архитектура сети (с 1-м скрытым слоем, а не 3-мя) показывает лучший результат.

Между конфигурациями, где скрытые слои содержат по 20 нейронов, особых различий нет; здесь ошибка в целом выше, чем в предыдущих двух случаях.