

神经网络在数据挖掘中的应用

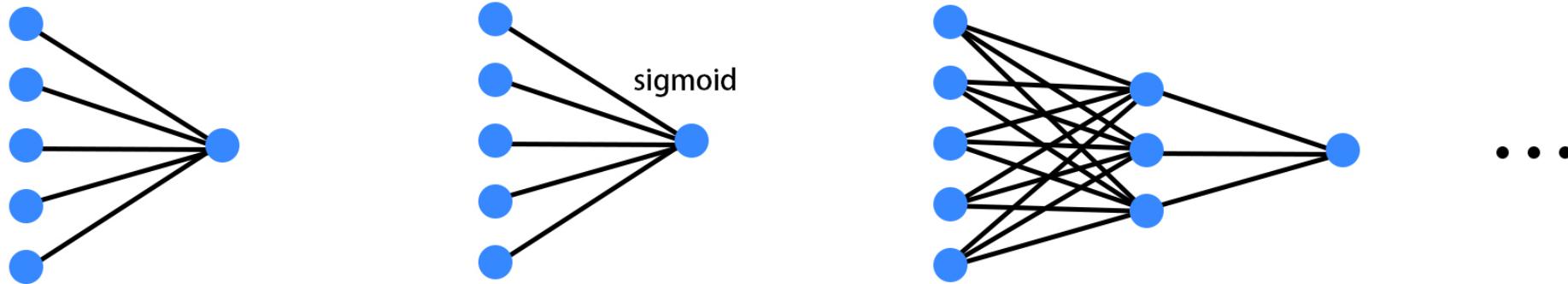
毕业论文开题答辩

数学与应用数学163班 段忠杰

研究内容-理论部分

各类神经网络的原理

1. 简单的单层神经网络（多元线性回归、对数几率回归）
2. 含隐层的神经网络（感知机）
3. 结构复杂的神经网络（递归神经网络、卷积神经网络、生成对抗网络等）



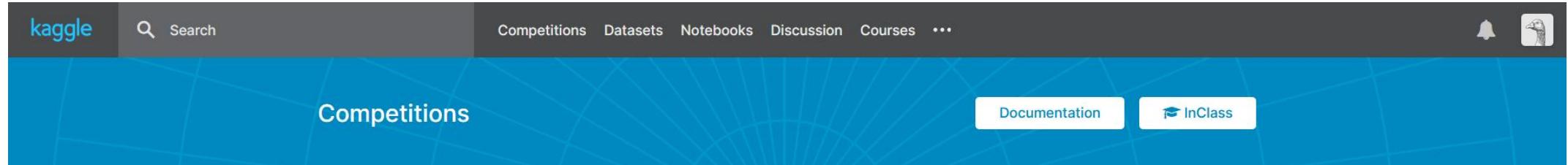
神经网络的相关算法

1. 梯度下降法
2. BP算法

研究内容-应用部分

数据挖掘

1. kaggle



2. 天池



研究内容-应用部分

Example

 Getting Started Prediction Competition

Digit Recognizer

Learn computer vision fundamentals with the famous MNIST data

 Kaggle 2,246 teams Ongoing

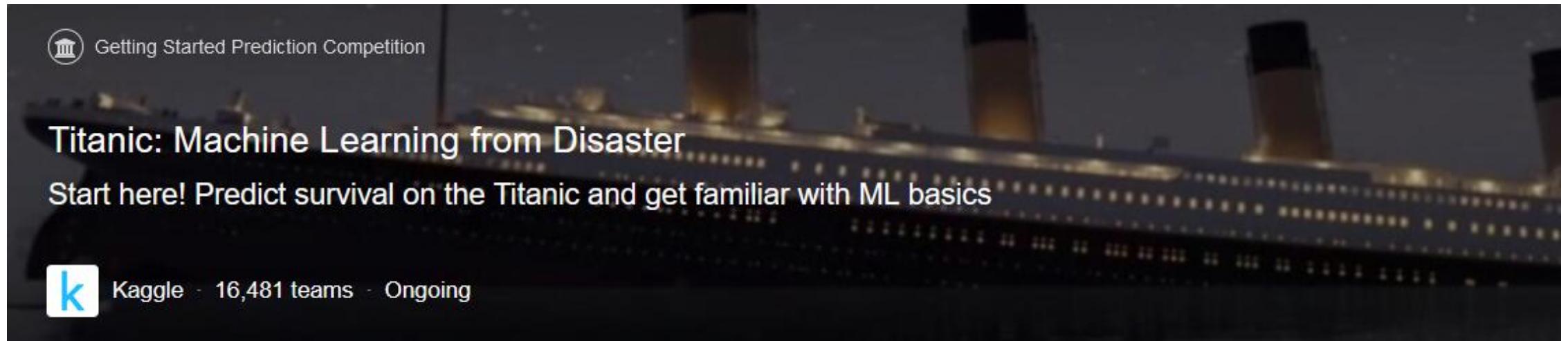
43774574472385075512090127923300424053119592925967978652191809929366043204879287452323608290
50210537141904328438821356087215922908963890574851966851832029350348733154711406434053732404307
90120537141904328438821356087215922908963890574851966851832029350348733154711406434053732404307
3372099951443808315895740350519020782129468324711678004954634326251266876890619780435057817529
81940386858144924978455794805314115266561279440306173191173420955823611836587386305400026165970430
09731238453381314217612544113500234314927520277596390226310156516474235698527833256217273251741
361291242471046746769119531902845879839819619224862113324603579425898604709261068246438308989855
6576204771390302514070280045300463254843920181962724424170284386916310810102908476799415362963
5205204771390302514070280045300463254843920181962724424170284386916310810102908476799415362963
5514813272849080198848101451509185741675984610932944064969478012701113711373358221937021875397085
71980561380867241434799869760351163797833055408252905172665419951226287243801134214712981048196
024120221414160125451135400616401693721050897547891139513462639145112154998684136185718
2919447231414160125451135400616401693721050897547891139513462639145112154998684136185718
7321602346049800830749843349842872459260845787240445189322432742814501111978088068264995292074529
1103261288LP18742445131380341886600394814917801203791533289150937643795189267675128316371416330
244041731888654697358785211547107216311671912283123574344341266588709401490176282607287413200441
8L718877306003620095126911039984583852938846948740256479370041863813501458393691488798686453
401718877306003620095126911039984583852938846948740256479370041863813501458393691488798686453
1771813681818194511459321325943274983854940658184985390282060610481125023434236021434014062353686
711813681818194511459321325943274983854940658184985390282060610481125023434236021434014062353686
31149022194417131254348447658612143901978537375169630819003692156773220545302739888502116083886

 Getting Started Prediction Competition

Titanic: Machine Learning from Disaster

Start here! Predict survival on the Titanic and get familiar with ML basics

 Kaggle 16,481 teams Ongoing



研究方法

1. 学习机器学习相关的知识，掌握神经网络的原理

	Python Learn the most important language for Data Science		Pandas Short hands-on challenges to perfect your data man		Advanced SQL Take your SQL skills to the next level
	Intro to Machine Learning Learn the core ideas in machine learning and build your f		Feature Engineering The most effective way to improve your models		Geospatial Analysis Create interactive maps, and discover patterns in geospatial data
	Intermediate Machine Learning Learn to handle missing values, non-numeric values, data accurate and useful.		Deep Learning Use TensorFlow to take Machine Learning to the nex		Microchallenges Ultra-short challenges to build and test your skill
	Data Visualization Make great data visualizations. A great way to see the po		Intro to SQL Learn SQL for working with databases, using Google		Machine Learning Explainability Extract human understandable insights from any Machine Learning model

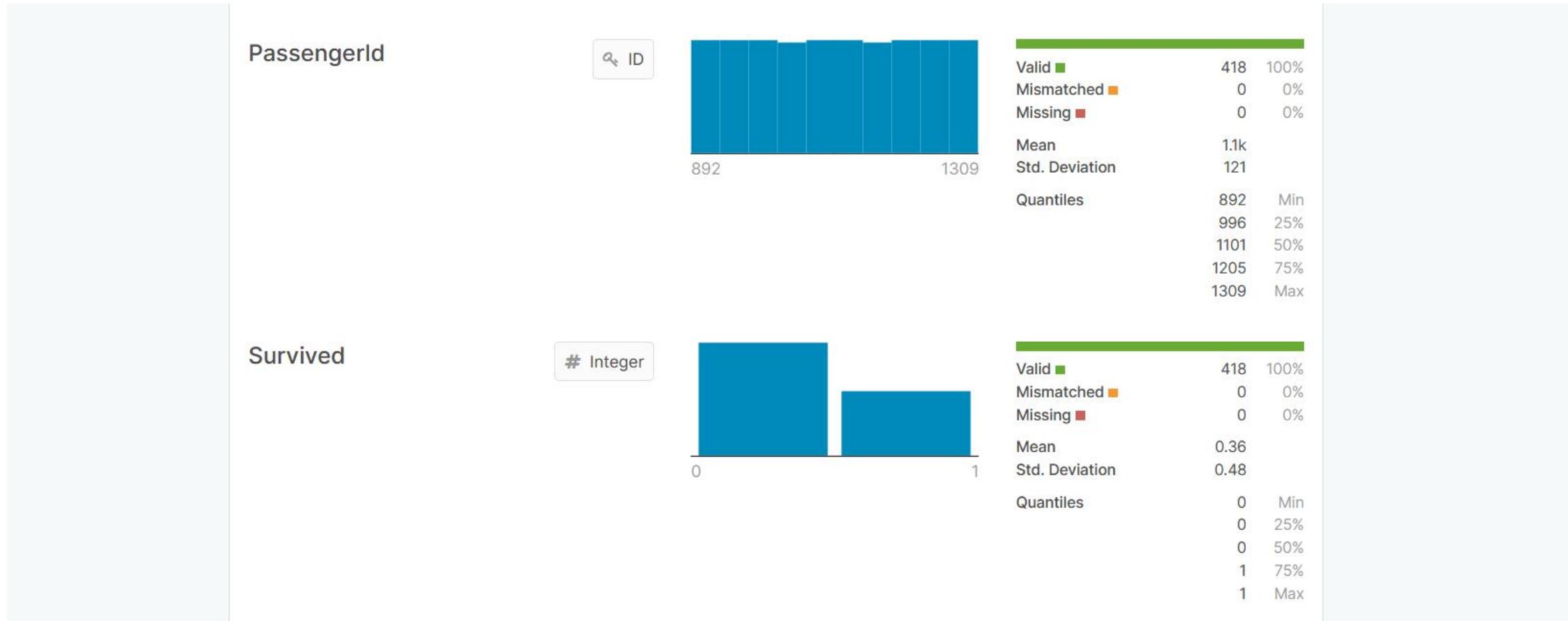
研究方法

2. 利用计算机编程，从原理出发实现简单的神经网络

```
61 class NeuralNetwork {
62 private:
63     Tensor dw;
64 public:
65     std::vector<NeuralNetworkLayer> ly;
66     double eta = 0.1;
67     NeuralNetwork() {}
68     NeuralNetwork(std::vector<int> siz, std::vector<int> flag) {
69         assert(siz.size() == flag.size());
70         ly.resize(siz.size());
71         ly[0].set_input_layer(siz[0]);
72         for (int i = 1; i < siz.size(); i++) {
73             ly[i].set_connect_layer(ly[i - 1], siz[i], flag[i]);
74             ly[i].weight_initialize();
75         }
76         int maxd = 0;
77         for (auto &i : siz)maxd = std::max(maxd, i);
78         dw.resize((int)siz.size(), maxd, maxd);
79     }
80     void show()const {
81         for (int i = 0; i < ly.size(); i++) {
82             if (i == 0) {
83                 std::cout << " 0: input layer" << std::endl;
84             } else {
85                 std::cout << " " << i << ": layer" << std::endl;
86                 ly[i].show();
87             }
88         }
89     }
90     Vector predict(const Vector &x) {
91         assert(x.size() == ly[0].size());
92         ly[0].value = ly[0].activeValue = x;
93         for (int i = 1; i < ly.size(); i++) {
94             ly[i].calculate(ly[i - 1].activeValue);
95         }
96         return ly.back().activeValue;
97     }
98     void calculate_diff(int I, int J, int K) {
99         double temp;
100        for (int j = 0; j < ly[I].size(); j++) {
101            temp = active_function_diff(ly[I].activeFlag, ly[I].value[j]);
102            ly[I].diffValue[j] = (j == J) ? (ly[I - 1].value[K] * temp) : 0;
103        }
104        for (int i = I + 1; i < ly.size(); i++) {
105            for (int j = 0; j < ly[i].size(); j++) {
106                for (int k = 0; k < ly[i].size(); k++) {
107                    if (j == K) {
108                        ly[i].diffValue[j] += ly[i - 1].value[k] * temp;
109                    }
110                }
111            }
112        }
113    }
114 }
```

研究方法

3. 充分利用 Kaggle 等网站提供的数据，测试神经网络的性能



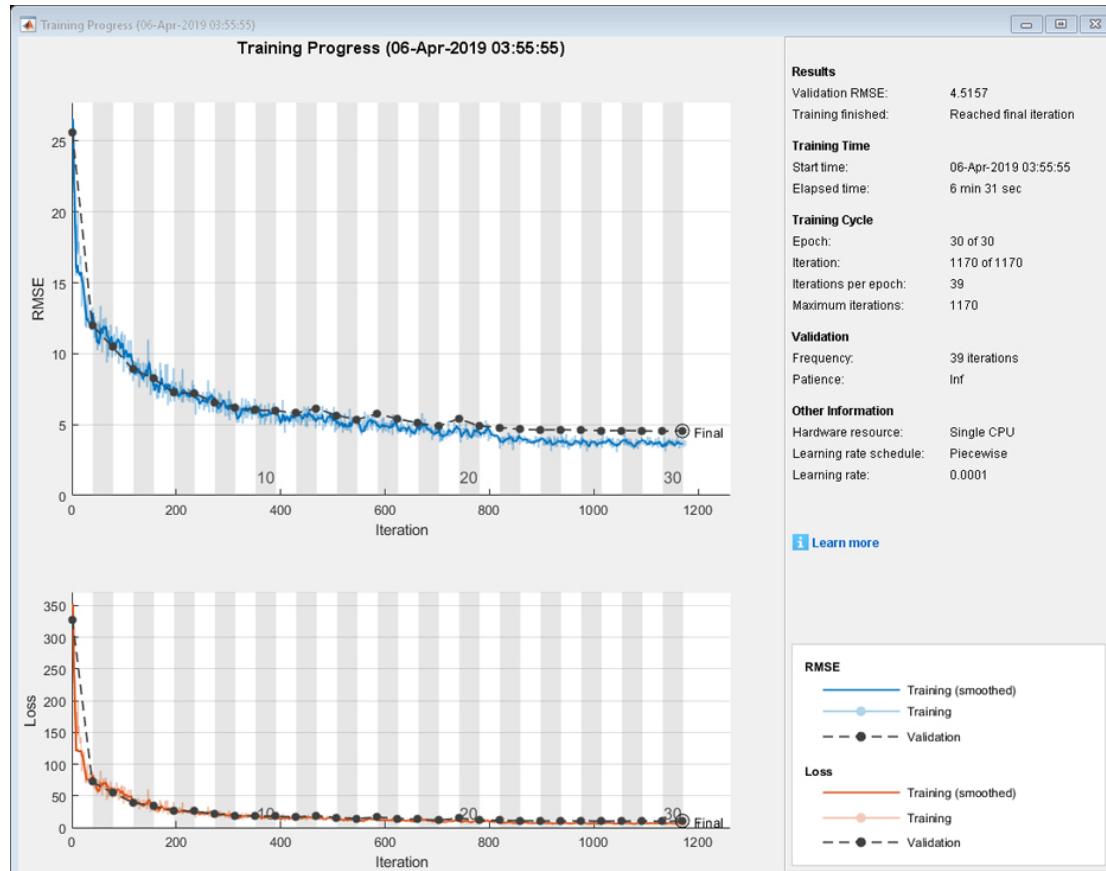
研究方法

4. 调整神经网络的结构与参数，提高其效率

			Overview	Data	Notebooks	Discussion	Leaderboard	Rules	Team		My Submissions	Submit Predictions
1865	leyouqi								0.95871	7	25d	
1866	Mana Masuda								0.95871	1	23d	
1867	Alexandru Zbiera								0.95871	1	14d	
1868	Alnewbie000								0.95842	4	1mo	
1869	Arnold H								0.95828	1	2mo	
1870	Talgat Omarov								0.95814	1	15h	
1871	Luke Reardon								0.95800	10	2mo	
1872	Nidhi Suthar								0.95757	2	10d	
1873	Andreas Gravrok								0.95742	2	2mo	
1874	Pan DENG								0.95671	1	2mo	
1875	dorado								0.95671	1	1mo	

研究方法

5. 使用 Matlab 与 TensorFlow 提供的神经网络框架，复现各种结构复杂的神经网络



The image shows the TensorFlow landing page. The main title is '端到端的开源机器学习平台'. Below it, there are four navigation links: 'TensorFlow' (selected), '针对 JavaScript', '针对移动设备和 IoT 设备', and '针对生产'. A descriptive text block states: '核心开源库，可以帮助您开发和训练机器学习模型。您可以通过直接在浏览器中运行 Colab 笔记本来快速开始使用。' At the bottom, a large orange button says '开始使用 TensorFlow'.

研究方法经网络

6. 尝试自行设计复杂神经网络的结构，来解决数据挖掘中的实际问题
7. 撰写论文，总结研究成果

恳请各位老师批评指正！