

北京邮电大学

本科毕业设计（论文）



题目： 社猜猜看这个毕设题目是什么

姓 名 猜 猜

学 院 信息与通信工程学院

专 业 通信工程

班 级 2014211199

学 号 2014210999

班内序号 99

指导教师 猜 猜

2018 年 5 月

目 录

第一章 基础模块示例	1
1.1 特殊文本类型	1
1.1.1 脚注	1
1.1.2 定义、定理与引理等	1
1.1.3 中英文文献、学位论文引用	2
1.2 图表及其引用	2
1.3 公式与算法表示	5
1.3.1 例子：基于主成分分析	5
1.3.1.1 主成分分析算法	5
1.3.1.2 主成分分析可信度评估方法	6
1.4 代码表示	6
1.4.1 直接书写代码在.tex 中	6
1.4.2 引用代码文件	7
1.5 列表样式	7
1.5.1 使用圆点作为项目符号	7
1.5.2 使用数字作为项目符号	7
1.5.3 句中数字编号列表样式	7
参考文献	8
致 谢	9
附 录	10
附录 1 缩略语表	10
附录 2 数学符号	11

第一章 基础模块示例

1.1 特殊文本类型

1.1.1 脚注

社交媒体是一种供用户创建在线社群来分享信息、观点、个人信息和其它内容(如视频)的电子化交流平台,社交网络服务(social network service, SNS)和微博客(microblogging)都属于社交媒体的范畴^[1],国外较为知名的有 Facebook¹、Instagram²、Twitter³、LinkedIn⁴等,国内较为知名的有新浪微博⁵。

在社交媒体的强覆盖下,新闻信息的传播渠道也悄然发生了变化。^[2]

1.1.2 定义、定理与引理等

定义 1.1 这是一条我也不知道在说什么的定义,反正我就是写在这里做个样子罢了,也没人会仔细读。^[3]

定理 1.1 这是一条我也不知道在说什么的定理,反正我就是写在这里做个样子罢了,也没人会仔细读。

公理 1.1 这是一条我也不知道在说什么的公理,反正我就是写在这里做个样子罢了,也没人会仔细读。

引理 1.1 这是一条我也不知道在说什么的引理,反正我就是写在这里做个样子罢了,也没人会仔细读。

命题 1.1 这是一条我也不知道在说什么的命题,反正我就是写在这里做个样子罢了,也没人会仔细读。

推论 1.1 这是一条我也不知道在说什么的推论,反正我就是写在这里做个样子罢了,也没人会仔细读。

本项目来源于科研项目“基于 \LaTeX 的本科毕业设计”,项目编号 1124

¹<http://www.facebook.com/>

²<https://www.instagram.com/>

³<http://www.twitter.com/>

⁴<http://www.linkedin.com/>

⁵<http://www.weibo.com/>

1.1.3 中英文文献、学位论文引用

根据美国皮尤研究中心的 2017 年 9 月发布的调查结果^[4]，67% 的美国民众会从社交媒体上获取新闻信息，其中高使用频率用户占 20%。在国内，中国互联网信息中心《2016 年中国互联网新闻市场研究报告》^[5] 也显示，社交媒体已逐渐成为新闻获取、评论、转发、跳转的重要渠道，在 2016 年下半年，曾经通过社交媒体获取过新闻资讯的用户比例高达 90.7%，在微信、微博等社交媒体参与新闻评论的比例分别为 62.8% 和 50.2%。社交媒体正在成为网络上热门事件生成并发酵的源头，在形成传播影响力后带动传统媒体跟进报道，最终形成更大规模的舆论浪潮。^[6]

在国内，新浪微博由于其发布方便、传播迅速、受众广泛且总量大的特点，成为了虚假信息传播的重灾区：《中国新媒体发展报告（2013）》^[7] 显示，2012 年的 100 件微博热点舆情案例中，有超过 1/3 出现谣言；《中国新媒体发展报告（2015）》^[8] 对 2014 年传播较广、比较典型的 92 条假新闻进行了多维度分析，发现有 59% 的虚假新闻首发于新浪微博。

此等信息的传播严重损害了有关公众人物的名誉权，降低了社交媒体服务商的商业美誉度，扰乱了网络空间秩序，冲击着网民的认知，极易对民众造成误导，带来诸多麻烦和经济损失，甚至会导致社会秩序的混乱。针对社交媒体谣言采取行动成为了有关部门、服务提供商和广大民众的共同选择。^[3]

1.2 图表及其引用

此处引用了简单的表 1-1。

请注意， \LaTeX 的图表排版规则决定了图表**不一定会乖乖呆在你插入的地方**，这是为了避免 Word 中由于图片尺寸不匹配在页面下部出现的空白，所以请不要使用“下图”“下表”作为指向文字，应使用“图 1-1 所示”这样的表述。

表 1-1 基于浏览者行为的特征

特征	描述	形式与理论范围
点赞量	微博的点赞数量	数值，N
评论量	微博的评论数量	数值，N
转发量	微博的转发数量	数值，N

此处引用了复杂的表 1-2。

此处展示了更专业的表 1-3，一个好的表格没有竖线。

此处引用了一张图。图 1-1 表示的是一个由含有 4 个神经元的输入层、含有 3 个神经元的隐藏层和含有 4 个神经元的输出层组成的自编码器，+1 代表偏置项。

表 1-2 基于浏览者行为的复杂特征

类别	特征	不知道叫什么的表头	
		描述	形式与理论范围
正常互动	点赞量	微博的点赞数量	数值，N
	评论量	微博的评论数量	数值，N
	转发量	微博的转发数量	数值，N
非正常互动	羡慕量	微博的羡慕数量	数值，N

表 1-3 红警 2 名词解释

术语类别	缩略语	解释
游戏	兵营	兵营（Barracks），《命令与征服 红色警戒 2：尤里的复仇》游戏中的一种生产建筑，用以生产步兵单位
	建造场	建造场（Construction Yard），《命令与征服 红色警戒 2：尤里的复仇》游戏中的一种基础建筑，用以支持其他建筑的建造
	矿厂	矿石精炼厂（Ore Refinery），《命令与征服 红色警戒 2：尤里的复仇》游戏中的一种资源建筑，用以将矿车采集的矿石转化为游戏资金
	空指	空指部（Airforce Command Headquarters），《命令与征服 红色警戒 2：尤里的复仇》游戏中的一种资源建筑，用以提供雷达功能和 T2 科技及生产部分空军单位
	相机	游戏术语，特指游戏内的观察区域和视角
	重工	战车工厂（War Factory），《命令与征服 红色警戒 2：尤里的复仇》游戏中的一种生产建筑，用以生产载具单位
	战争迷雾	游戏术语，《命令与征服 红色警戒 2：尤里的复仇》中指黑色的未探索区域

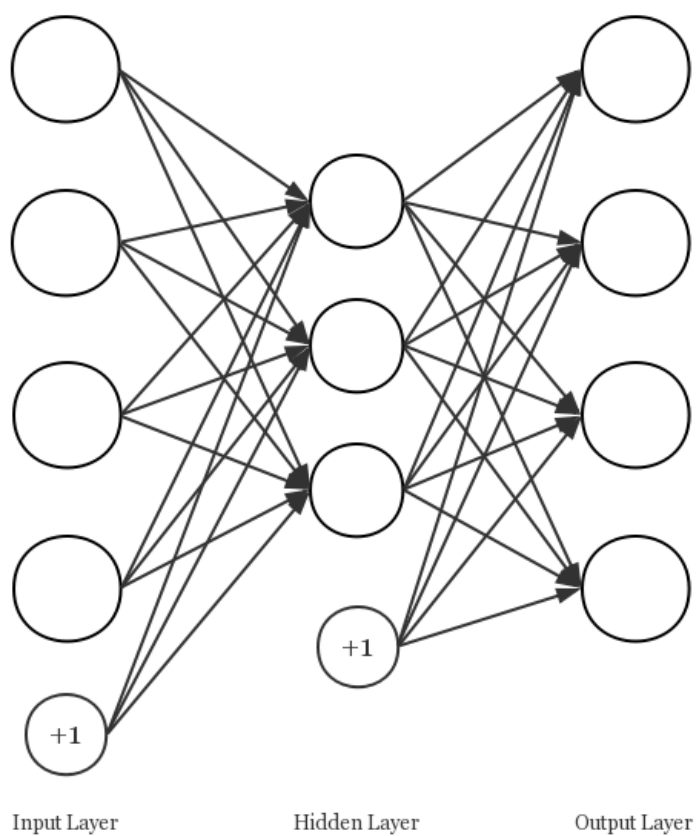
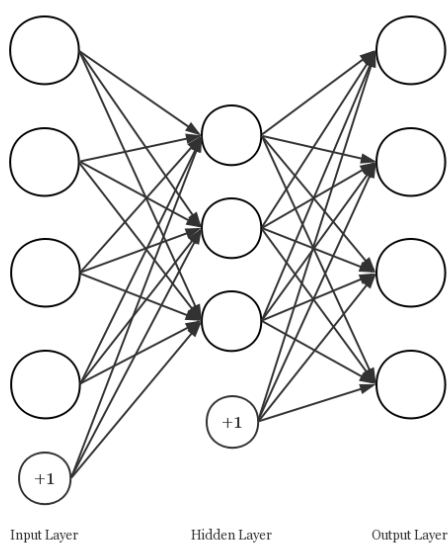


图 1-1 自编码器结构

(a)



(b)

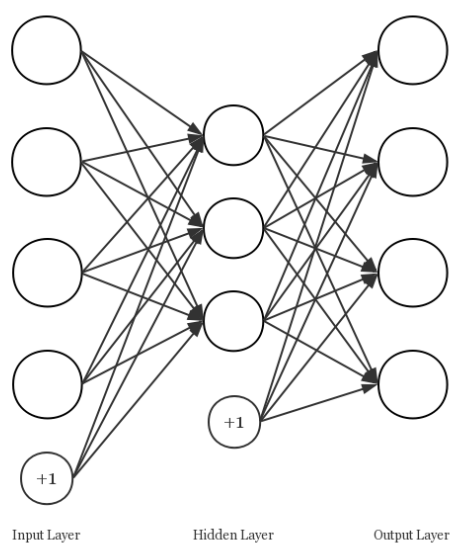


图 1-2 这是两个自编码器结构，我就是排一下子图的效果：(a)左边的自编码器，(b)右边的自编码器

1.3 公式与算法表示

1.3.1 例子：基于主成分分析

1.3.1.1 主成分分析算法

下面对主成分分析进行介绍。

主成分分析是一种简单的机器学习算法，其功能可以从两方面解释：一方面可以认为它提供了一种压缩数据的方式，另一方面也可以认为它是一种学习数据表示的无监督学习算法。^[9] 通过 PCA，我们可以得到一个恰当的超平面及一个投影矩阵，通过投影矩阵，样本点将被投影在这一超平面上，且满足最大可分性（投影后样本点的方差最大化），直观上讲，也就是能尽可能分开。

对中心化后的样本点集 $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_m\}$ （有 $\sum_{i=1}^m \mathbf{x}_i = 0$ ），考虑将其最大可分地投影到新坐标系 $\mathbf{W} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_i, \dots, \mathbf{w}_d\}$ ，其中 \mathbf{w}_i 是标准正交基向量，满足 $\|\mathbf{w}_i\|_2 = 1$ ， $\mathbf{w}_i^T \mathbf{w}_j = 0$ ($i \neq j$)。假设我们需要 d' ($d' < d$) 个主成分，那么样本点 \mathbf{x}_i 在低维坐标系中的投影是 $\mathbf{z}_i = (z_{i1}; z_{i2}; \dots; z_{id'})$ ，其中 $z_{ij} = \mathbf{w}_j^T \mathbf{x}_i$ ，是 \mathbf{x}_i 在低维坐标系下第 j 维的坐标。对整个样本集，投影后样本点的方差是

$$\begin{aligned}
 & \frac{1}{m} \sum_{i=1}^m \mathbf{z}_i^T \mathbf{z}_i \\
 &= \frac{1}{m} \sum_{i=1}^m (\mathbf{x}_i^T \mathbf{W})^T (\mathbf{x}_i^T \mathbf{W}) \\
 &= \frac{1}{m} \sum_{i=1}^m \mathbf{W}^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{W} \\
 &= \frac{1}{m} \mathbf{W}^T \mathbf{X} \mathbf{X}^T \mathbf{W}
 \end{aligned} \tag{1-1}$$

由于我们知道新坐标系 \mathbf{W} 的列向量是标准正交基向量，且样本点集 \mathbf{X} 已经过中心化，则 PCA 的优化目标可以写为

$$\begin{aligned}
 & \max_{\mathbf{W}} \quad \text{tr}(\mathbf{W}^T \mathbf{X} \mathbf{X}^T \mathbf{W}) \\
 & \text{s. t.} \quad \mathbf{W}^T \mathbf{W} = \mathbf{I}
 \end{aligned} \tag{1-2}$$

由于 $\mathbf{X} \mathbf{X}^T$ 是协方差矩阵，那么只需对它做特征值分解，即

$$\mathbf{X}^T \mathbf{X} = \mathbf{W} \mathbf{\Lambda} \mathbf{W}^T \tag{1-3}$$

其中 $\mathbf{\Lambda} = \text{diag}(\boldsymbol{\lambda})$ ， $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \dots, \lambda_m\}$ 。

具体地，考虑到它是半正定矩阵的二次型，存在最大值，可对式 (1-2) 使用拉格

朗日乘法

$$\mathbf{X}\mathbf{X}^T\mathbf{w}_i = \lambda_i\mathbf{w}_i \quad \text{式 (1-4)}$$

之后将求得特征值降序排列, 取前 d' 个特征值对应的特征向量组成所需的投影矩阵 $\mathbf{W}' = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{d'})$, 即可得到 PCA 的解。PCA 算法的描述如算法1所示。

算法 1 主成分分析 (PCA)

输入: 样本集 $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_m\}$, 低维空间维数 d'

输出: 投影矩阵 $\mathbf{W}' = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{d'})$

- 1: 对所有样本中心化 $\mathbf{x}_i \leftarrow \mathbf{x}_i - \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i$
 - 2: 计算样本的协方差 $\mathbf{X}\mathbf{X}^T$
 - 3: 对协方差矩阵 $\mathbf{X}\mathbf{X}^T$ 做特征值分解
 - 4: 取最大的 d' 个特征值所对应的特征向量 $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{d'}$
-

1.3.1.2 主成分分析可信度评估方法

记待判定微博 \mathbf{w}_0 的经典特征向量为 \mathbf{f}_0^c , 它的发布者在 \mathbf{w}_0 前发布的 k 条微博为 $\mathbf{W} = \mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k$, 这 k 条微博对应的经典特征向量集为 $\mathbf{F}_W^c = \{\mathbf{f}_1^c, \mathbf{f}_2^c, \dots, \mathbf{f}_k^c\}$ 。令 $label = 1$ 代表谣言, $label = 0$ 代表非谣言。算法的具体流程如算法2所示。

算法 2 基于 PCA 的信息可信度评估

输入: $\mathbf{f}_0^c, \mathbf{F}_W^c$, 保留主成分数 n

输出: 标签 $label \in \{0, 1\}$

- 1: 对所有特征向量应用 PCA, 保留前 n 个主成分 $\mathbf{o}_i^c \leftarrow PCA(\mathbf{f}_i^c, n)$ ($i = 0, 1, \dots, k$)
 - 2: 计算 \mathbf{F}_W^c 中各向量的平均距离 μ 和标准差 σ
 - 3: 计算阈值 $thr = \mu/\sigma$
 - 4: **if** $\min_{1 \leq j \leq k} \|\mathbf{o}_0^c - \mathbf{o}_j^c\|_2 > thr$ **then**
 - 5: $label \leftarrow 1$
 - 6: **else**
 - 7: $label \leftarrow 0$
 - 8: **end if**
-

1.4 代码表示

1.4.1 直接书写代码在.tex 中

下面的代码1-1是用 Python 编写的加法函数。

代码 1-1 加法

```
1 def plusFunc(a, b):  
2     return a + b
```

1.4.2 引用代码文件

下面的代码1-2是用 Python 文件中引入的倒序打印 x 到 1 的函数，请查看 code 文件夹。

代码 1-2 倒序打印数字

```
1 def numbers(x):  
2     if x > 0:  
3         print(x)  
4         numbers(x-1)
```

1.5 列表样式

1.5.1 使用圆点作为项目符号

- 第一章为基础模块示例，是的，本章的名字就是基础模块示例，正如你看到这个样子。
- 第二章为不存在，是的，其实它不存在。

1.5.2 使用数字作为项目符号

1. 第一章为基础模块示例，是的，本章的名字就是基础模块示例，正如你看到这个样子。
2. 第二章为不存在，是的，其实它不存在。

1.5.3 句中数字编号列表样式

1. 第一章为基础模块示例，是的，本章的名字就是基础模块示例，正如你看到这个样子; 2. 第二章为不存在，是的，其实它不存在。

参考文献

- [1] Merriam-Webster. Social Media [EB/OL]. 2018 [2018-04-15]. <http://www.merriam-webster.com/dictionary/socialmedia>.
- [2] Vosoughi Soroush, Roy Deb, Aral Sinan. The spread of true and false news online [J]. Science. 359 (6380). 2018, 3: 1146–1151.
- [3] 周兴. 基于深度学习的谣言检测及模式挖掘 [学位论文]. 中国科学院大学, 2017.
- [4] Pew Research Center. News Use Across Social Media Platforms 2017 [EB/OL]. 2017 [2018-04-15]. <http://www.journalism.org/2017/09/07/news-use-across-social-media-platforms-2017>.
- [5] 中国互联网络信息中心. 2016 年中国互联网新闻市场研究报告 [EB/OL]. 2017 [2018-04-15]. <http://www.cnnic.cn/hlwfzyj/hlwxyzbg/mtbg/201701/P020170112309068736023.pdf>.
- [6] Yang Fan, Liu Yang, Yu Xiaohui et al. Automatic detection of rumor on Sina Weibo [C]. In Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics. New York, NY, USA. 2012 : 1–7.
- [7] 唐绪军, 吴信训, 黄楚新 等. 中国新媒体发展报告 No.4(2013) [M]. 社会科学文献出版社, 2013.
- [8] 唐绪军, 吴信训, 黄楚新 等. 中国新媒体发展报告 No.6(2015) [M]. 社会科学文献出版社, 2015.
- [9] Goodfellow Ian, Bengio Yoshua, Courville Aaron. Deep Learning [M]. The MIT Press, 2016.

致 谢

此处请写致谢的内容。
它可以有多段。

附录

附录 1 缩略语表

表 附-1 基于浏览者行为的特征

特征	描述	形式与理论范围
点赞量	微博的点赞数量	数值，N
评论量	微博的评论数量	数值，N
转发量	微博的转发数量	数值，N

表 附-2 基于浏览者行为的复杂特征

类别	特征	不知道叫什么的表头	
		描述	形式与理论范围
正常互动	点赞量	微博的点赞数量	数值，N
	评论量	微博的评论数量	数值，N
	转发量	微博的转发数量	数值，N
非正常互动	羡慕量	微博的羡慕数量	数值，N

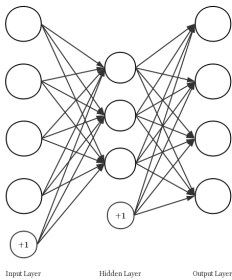


图 附-1 自编码器结构

代码 附-1 减法

```
1 def minusFunc(a, b):
2     return a - b
```

$$\max_{\boldsymbol{W}} tr(\boldsymbol{W}^T \boldsymbol{X} \boldsymbol{X}^T \boldsymbol{W})$$

式（附-1）

附录 2 数学符号

数和数组

a	标量（整数或实数）
\mathbf{a}	向量
$\dim()$	向量的维数
\mathbf{A}	矩阵
\mathbf{A}^T	矩阵 \mathbf{A} 的转置
\mathbf{I}	单位矩阵（维度依据上下文而定）
$\text{diag}(\mathbf{a})$	对角方阵，其中对角元素由向量 \mathbf{a} 确定