

曲阜师范大学

# 本科生毕业论文（设计）



|        |                                |      |            |
|--------|--------------------------------|------|------------|
| 题    目 | 基于 ReduNet 的高维数据低维特征<br>抽取方法研究 |      |            |
| 姓    名 | 李家熙                            | 学号   | 2019411984 |
| 院    系 | 物理工程学院                         |      |            |
| 专    业 | 电子信息工程                         |      |            |
| 指导教师   | 张明强                            | 职  称 | 讲师         |

2023 年 5 月 16 日

曲阜师范大学教务处制

# 目 录

|                                      |    |
|--------------------------------------|----|
| 摘要 .....                             | I  |
| 关键词 .....                            | I  |
| Abstract.....                        | I  |
| Key words.....                       | I  |
| 1 引言 .....                           | 1  |
| 2 高维数据特征抽取的理论与方法 .....               | 2  |
| 2.1 传统高维数据特征抽取方法 .....               | 2  |
| 2.2 ReduNet 的预备理论 .....              | 4  |
| 2.2.1 香农信息编码理论 .....                 | 4  |
| 2.2.2 反馈控制论基本思想 .....                | 4  |
| 2.3 本章小结 .....                       | 4  |
| 3 基于 ReduNet 的高维数据低维特征抽取 .....       | 4  |
| 3.1 最大化编码速率降低原理 .....                | 5  |
| 3.1.1 通过压缩实现分类和聚类 .....              | 5  |
| 3.1.2 线性判别表示 .....                   | 6  |
| 3.1.3 基于香农信息论的解释 .....               | 6  |
| 3.2 基于 ReduNet 的高维数据低维特征最紧凑的表示 ..... | 8  |
| 3.2.1 ReduNet 模型 .....               | 8  |
| 3.2.2 压缩闭环转录 .....                   | 9  |
| 3.3 实验设计与模型测试 .....                  | 11 |
| 3.3.1 实验环境配置与数据集 .....               | 11 |
| 3.3.2 模型的设计 .....                    | 11 |
| 3.3.3 模型的训练与测试 .....                 | 12 |
| 3.4 实验结果分析 .....                     | 13 |
| 3.5 本章小结 .....                       | 15 |
| 4 结论与展望 .....                        | 15 |
| 致谢 .....                             | 16 |
| 参考文献 .....                           | 16 |

# 基于 ReduNet 的高维数据低维特征抽取方法研究

电子信息工程专业学生 李家熙

指导教师 张明强

**摘要：**传统高维数据降维方法通常采用线性映射，存在表达能力不足的缺点，近年来，深度神经网络凭借优秀的表达学习能力，能够实现特征的高效抽取，但仍存在可解释性和泛化能力不足等问题。为此，在对高维数据降维方法充分调研上，重点分析了基于最大化编码速率降低原理的白盒深度网络 ReduNet 模型，分别从信息论和反馈控制论角度讨论了 ReduNet 的基本原理和实现方法，并在 PyTorch+Colab 环境下予以实现。实验结果表明，ReduNet 具有较强的可解释性和更好地解决过拟合问题的能力，该方法在图像分类、识别等任务上具有规模应用的前景。

**关键词：**ReduNet 最大化编码速率降低 数据降维 香农信息编码理论 维纳反馈控制论

## Low-Dimensional Feature Extraction Methods of High-Dimensional Data Based on ReduNet

Student Majoring in Electronic Information Engineering Li Jiayi

Tutor Zhang Mingqiang

**Abstract:** Traditional methods for dimensionality reduction of high-dimensional data usually apply linear mapping, which has the disadvantage of insufficient expressiveness. In recent years, deep neural networks, with their excellent representation learning ability, can achieve efficient extraction of features, but still suffer from problems such as insufficient interpretability and generalization ability. In this regard, based on the full investigation of high-dimensional data dimensionality reduction methods, we focus on analyzing the ReduNet model, a white-box deep network based on the principle of the Maximizing Coding Rate Reduction, and discuss the basic principles and implementation methods of the perspectives of the Information Theory and the Feedback Cybernetics, respectively, and implement them in the PyTorch+Colab environment. The experimental results show that ReduNet has strong interpretability and better ability to solve the overfitting problem, and the method has the prospect of valuable application in image classification and recognition tasks.

**Key words:** ReduNet; Maximal Rate Coding Reduction; Data Dimensionality Reduction; Shannon's Information Coding Theory; Wiener Feedback Cybernetics

## 1 引言

在大数据时代,随着信息技术飞速发展,数据维度和规模以前所未有的速度增长<sup>[1]</sup>,在图像分析、信息处理、语音文本等产生的高维数据<sup>[2]</sup>,需要从中准确、快速地抽取有效特征来分析相关任务。Tian (USC) 等人对高维数据是这样定义的: High-Dimensional Data refers to data with a large number of variables, often large the number of observations<sup>[3]</sup>,即数据的属性数量远大于样本数量。随着维度的增加,高维空间参数估计量会要求更多的样本数据。如果直接对高维数据进行降维处理会引起维度灾难的现象,指的是低维空间不会遇到的在高维空间中分析、组织数据时出现的各种问题,最早由 Richard Bellman<sup>[4]</sup>提出,这样的问题导致数据质量下降。因此,通常会先对数据做如图 1-1 所示处理:

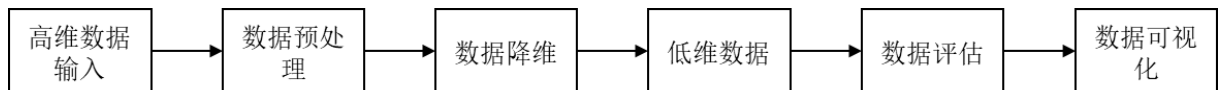


图 1-1 高维数据的处理过程

高维数据往往因为其稀疏性的样本空间,难以找到有效的数据特征(低维特征),若直接处理高维数据通常会产生引入噪声等异常值的问题,导致一些算法的失效。因此,通常会使用降维方法抽取数据特征,以便更好地获取有价值的信息,再加以分析处理这些关键特征,提高了数据分析的效率。

近年来,无论是传统的高维数据低维特征抽取方法还是热门的深度网络,由于其黑盒属性凭经验设计的网络,都出现了泛化能力不足的问题。

1901 年, Karl Pearson<sup>[5]</sup>提出了主成分分析 (Principle Components Analysis, PCA),成功地描述了相关性数据的主要特征。RA Fisher<sup>[6]</sup>于 1936 年提出了线性判别分析 (Linear Discriminant Analysis, LDA), 主要思想是最大化数据点类间距离,同时最小化数据点类内距离,通过分类实现了数据降维。1986 年, Hérault 和 Jutten<sup>[7]</sup>提出 ICA (Independent Component Analysis, 独立成分分析) 的基本思想,将观测信号分解为若干独立成分,解决了盲源分离问题。

这些传统的高维数据降维方法,如 PCA、LDA、ICA 等,通常是基于统计或代数的方法,通过最大化数据方差或类间距离等准则来进行降维,仅能对数据进行线性映射,并且存在表达能力较弱的缺点。同时占用较高的计算成本和存储空间,易出现过拟合等问题。2012 年 AlexNet 诞生以来,深度神经网络 (Deep Neural Network, DNN) 成为了数据降维的热门模型,它们凭借优秀的表达学习能力在图像、语音和自然语言处理等大规模数据实践中,解决过拟合问题取得了一定进展,实现了高维数据特征的有效抽取,但仍存在可解释性和泛化能力不足等问题<sup>[1]</sup>。

2020 年, Ma Yi (IEEE Fellow) 等人提出从第一性原理出发,通过数学推导,提出最大化编码速率降低 (Maximizing Coding Rate Reduction, MCR<sup>2</sup>) 原则<sup>[8]</sup>,通过优化 MCR<sup>2</sup> 目标,推导出基于数据压缩和线性判别表示 (Liner Discriminant Representation, LDR) 框架的 ReduNet 模型<sup>[9]</sup>。并基于简约、自洽两个原则<sup>[10]</sup>,从香农信息论和反馈控制论角度探索高维数据内部特征的有效表示问题。ReduNet 是一种基于 MCR<sup>2</sup> 原理的白盒深度神经网络,不需要用反向传播算法来学习参数且可直接推导出多通道卷积的结构,能够清晰地解释每一层网络的功能和参数含义,从数据压缩的角度构造 DNN 说明现代深度层次架构、线性算子和非线性激活,甚至网络中所有的参数,都可利用数据压缩中“编码率降低” (Rate Reduction) 的原理解释<sup>[11]</sup>,从而达到增强对网络的可解释性。使得 ReduNet

模型能够在图像分类、人脸识别、语音识别等任务上具有规模应用价值。

第一部分是引言，总结高维数据低维特征抽取的应用背景和研究意义，以及高维数据降维方法的研究现状。第二部分是高维数据低维特征抽取的基本方法和原理，包括高维数据降维经典方法的概念以及在实际应用的特点；并简要介绍信息率失真理论和反馈控制论基本思想。第三部分研究基于 ReduNet 的高维数据低维特征抽取，先重点介绍 MCR<sup>2</sup>原理，再阐释 ReduNet 模型算法，并从香农信息编码理论和反馈控制论两个角度，探究如何准确地量化和测量高维数据中包含的信息，寻求最紧凑的信息表示方法。然后设计实验编码予以实现，分析实验结果。最后得出文章的研究结论，说明 ReduNet 在诸多方向的应用前景。

## 2 高维数据特征抽取的理论与方法

本章主要介绍部分传统高维数据特征的抽取方法的原理和基本的机器学习理论，并概括传统降维方法的优劣势。阐述香农信息论率失真函数理论<sup>[12]</sup>和反馈控制论的主要思想。

### 2.1 传统高维数据特征抽取方法

经典的高维数据特征抽取的方法包括主成分分析（PCA）、独立成分分析（ICA）、线性判别分析（LDA）等等，为了将其与本文研究的MCR<sup>2</sup>方法、ReduNet 模型做比较，现简要说明以上三种方法的基本概念与应用特点。

主成分分析（PCA），主要思想是将高维数据映射到低维空间中，使用正交变换对一系列相关变量值进行线性变换，然后投影为不相关变量值，从而实现降维效果。数据在旋转后的方向轴上方差最大，即投影最为分散，主要的信息被保留。主要数学公式如下：

$$\text{var}(A) = \frac{1}{m} \sum_i^m (a_i - \mu_a)^2, \quad (2-1)$$

公式（2-1）中 $A$ 表示降维后特征， $\mu_a$ 是均值。 $\text{var}(A)$ 一般做零均值化处理： $\text{var}(A) = \frac{1}{m} \sum_i^m (a_i)^2$ ，希望 $\text{Conv}(A, B) = \frac{1}{m} \sum_i^m a_i b_i$ 为0，然后构造 $C = \frac{1}{m} X X^T$ 协方差矩阵。

假如，

$$X = \begin{bmatrix} a_1 & a_2 & \cdots & a_m \\ b_1 & b_2 & \cdots & b_m \end{bmatrix},$$

那么，

$$\frac{1}{m} X X^T = \begin{bmatrix} \frac{1}{m} \sum_{i=1}^m a_i^2 & \frac{1}{m} \sum_{i=1}^m a_i b_i \\ \frac{1}{m} \sum_{i=1}^m a_i b_i & \frac{1}{m} \sum_{i=1}^m b_i^2 \end{bmatrix} = \begin{bmatrix} \text{Cov}(a, a) & \text{Cov}(a, b) \\ \text{Cov}(b, a) & \text{Cov}(b, b) \end{bmatrix},$$

此时，协方差矩阵 $C$ 满足对角矩阵即可。降维后的数据 $Y$ 满足 $Y=XP$ ，达到高维数据特征抽取的效果。

主成分分析（PCA）可以缓解维度灾难，数据降噪，解决过拟合问题，但PCA是线性变换，处理非线性数据效果不佳。并且是一种无监督学习算法，没有考虑相关变量带来的影响。

独立成分分析（ICA）主要思想是在特征空间中寻找与高度非高斯投影对应的方向

的算法，就是为独立成分的线性组合寻找一组转换矩阵，实现成分的分离。其缺点是需要较高的计算复杂度，对于高维数据计算方法复杂，并且要求数据之间独立性要强。

线性判别分析（LDA）与线性判别表示（LDR）思想一致，使得数据类别之间距离最大化，同类内部距离最小化，是一种监督学习方法，相比 PCA 可增强分类的准确性。由图 2-1（a-b）使用 IRIS 数据可以看到 PCA 与 LDA 的差异，LDA 识别了类之间最大差异，而在 PCA 在分析类别之间的差异不够明显。

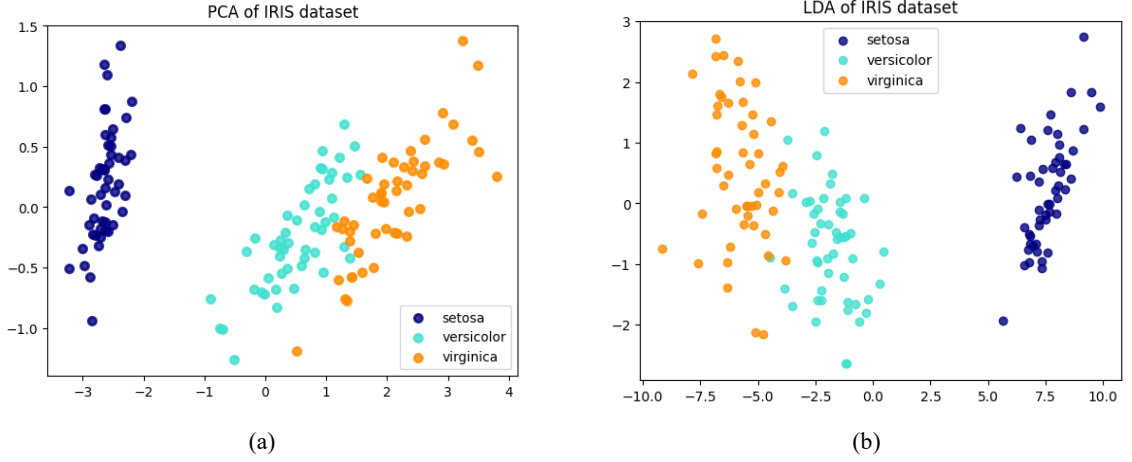


图 2-1 PCA 和 LDA

深度网络的分类器通常可以使用分类交叉熵（Categorical Cross-Entropy）来衡量模型预测的标签和实际的标签之间的差异，衡量模型对于不同类别分类的准确程度，也就是交叉熵的值越小，模型的两个标签误差就越小，即模型的性能越好。

$$L_{CE}(y_i, \hat{y}_i) = -\frac{1}{N} \sum_{i=1}^N (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)), \quad (2-2)$$

$$y_i^*, \hat{y}_i^* = \arg \min_{y_i, \hat{y}_i} L_{CE}(y_i, \hat{y}_i),$$

公式（2-2）以二分类模型为例，N 代表总体样本数， $\hat{y}_i$  表示模型对样本的预测概率值。

在分类问题中，假设样本属于 K 个类中的某一类别，则可以应用 Softmax 函数将多个神经元输出映射到(0, 1)区间内，转换为每个类别的概率：

$$\text{Softmax}(z_i) = \frac{\exp(z_i)}{\sum_j \exp(z_j)},$$

此时在 Softmax 层的损失函数为：

$$L_{CE} = -\log \frac{e^{k_y}}{\sum_j^k e^{k_j}}, \quad (2-3)$$

公式（2-3）使得模型期望输出的分数在正确标签真实分数上尽可能大。

网络的典型优化方法有随机梯度下降(Stochastic Gradient Descent, SGD)，它是随机选择一个计算梯度样本优化函数任务，加快了训练速度，参数更新公式如下：

$$w^+ = w - \eta \cdot \frac{\partial L}{\partial w}, \quad (2-4)$$

$$b^+ = b - \eta \cdot \frac{\partial L}{\partial b}, \quad (2-5)$$

公式（2-4）、（2-5）中，w 表示权重，b 表示偏置， $\eta$  是学习率。

## 2.2 ReduNet 的预备理论

Claude Shannon<sup>[12]</sup>在 1956 年提出的率失真函数在MCR<sup>2</sup>和 ReduNet 的推导和构造中占据重要的位置, 反馈思想为编码速率降低提供了纠错空间, 高维数据在流形上的映射增强了神经网络的表达能力。

### 2.2.1 香农信息编码理论

在信息编码过程中, 常常需要将高维数据映射到低维空间中, 以实现数据的压缩, 当数据压缩到一定失真界限时的最低比特率(码率), 在一定的失真度量下, 存在一个最小的码率<sup>[13]</sup>。对于给定一个码率, 对应能达到的最小期望失真是多少, 这是简要的率失真理论(Rate-Distortion Theory), 既需要尽可能地去压缩数据, 也需要尽可能保留数据的重要信息。

信息论中, 通常通过一个失真度量(Distortion Measure)来衡量一个随机变量以及它的表示之间的距离。失真函数指的是从信源空间与再生空间的乘积空间到非负实数集上的映射:

$$d: \mathcal{X} \times \hat{\mathcal{X}} \rightarrow \mathcal{R}^+,$$

$d(x, \hat{x})$ 来描述使用 $\hat{x}$ 表示 $x$ 时的代价度量。若给定了这样一个失真度 $D$ , 信息的率失真函数 $R(D)$ 定义为:

$$R(D) = \frac{1}{n} \sum_{i=1}^n \min_{\hat{x}_i} \{D(x_i, \hat{x}_i)\},$$

描述了在一定的码率下, 达到一定的失真所需要的最小平均码长。从而得出率失真编码定理, 即 Shannon 第三定理<sup>[12]</sup>:

对于给定的离散无记忆信源  $S$ , 无损编码后的长度为  $L$ , 解码后的数据失真度量为  $D$ , 当  $L$  趋于无穷大时, 失真  $D$  也趋于 0, 也就是压缩数据的长度接近于信源的信息熵, 就可以达到无失真压缩的效果。

### 2.2.2 反馈控制论基本思想

为了使得高维数据降维后得到的低维数据误差更小, 利用反馈来进行纠错, 使得高维数据降维取得更好的效果。Wiener 在 1948 年<sup>[14]</sup>提出: 为了改善某个或某些受控对象的功能或发展, 需要获得并使用信息, 以这种信息为基础而进行通信并作用于对象。其基本思想是将系统的输出信号与期望输出信号进行比较, 得到误差信号后再将误差信息作为反馈信号送入系统, 通过调节系统的控制量使得误差信号趋于零, 从而达到控制系统输出的目的。这样使得控制系统具有较好的自适应性, 提高了抗干扰能力。在神经网络中可以利用反馈控制论的基本思想, 利用输出结果来对网络进行调整, 使得网络的输出更加接近于期望输出, 有助于提高网络的鲁棒性和稳定性。

## 2.3 本章小结

本章根据文献调研的成果, 对后续论文阐述MCR<sup>2</sup>、ReduNet 做了部分综述。首先, 简要说明 PCA、ICA、LDA 的简要思想, 总结其缺点, 这些方法可以更好地处理实验数据, 得到网络更好的性能; 其次, 介绍机器学习中的交叉熵损失函数、Softmax 激活函数、SDG 优化算法; 最后, 香农信息理论、维纳控制论是MCR<sup>2</sup>、ReduNet 模型的出发理论, 从而得到有效信息的表示方法。

## 3 基于 ReduNet 的高维数据低维特征抽取

本章从数据压缩角度实现分类和聚类任务出发, 阐释最大化编码速率降低(MCR<sup>2</sup>)

原理并用其进行信息的表示，然后通过优化MCR<sup>2</sup>目标构造白盒模型—ReduNet，实现从香农信息理论和维纳控制理论的角度寻求信息的有效表示。

### 3.1 最大化编码速率降低原理

最大编码速率降低原理（MCR<sup>2</sup>）的提出可以最大限度地提高整个数据集之间的编码率差异和每个类别的总和<sup>[8]</sup>，可使用线性判别表示（LDR）、率失真理论等进行解释和推导，从混合有大量非线性低维结构的高维数据中<sup>[11]</sup>（High-Dim data with Mixed Low-Dim Structures）抽取低维线性特征。

#### 3.1.1 通过压缩实现分类和聚类

聚类和分类是两种主要的数据学习任务，通常是在寻找高维目标的低秩结构，如图 3-1 所示。但是深度网络在过去十年的演变发展过程中，一定程度上忽略了数据内部结构。通常是直接用经验设计深度网络在数据输入与输出之间的一个映射，通过这样的映射，深度网络可以预测新的样本，但是很难保证可靠性和解释性。

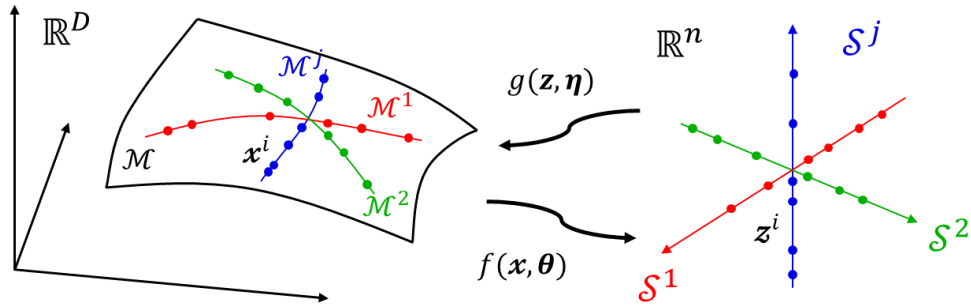


图 3-1 高维空间寻找低秩结构

如图 3-1 所示，如何寻找从混合低维子流形（A mixture of low-dim submanifolds）到混合不相干的子空间（A mixture of incoherent subspaces）的映射，定义  $f(x, \theta)$  这样一个非线性映射，将具有复杂低维结构的高维数据映射到紧凑的结构化表示， $\mathbb{R}^D$  指的是高维空间， $\mathbb{R}^n$  指的是线性独立子空间。

如果在一个数据压缩框架下，将高维数据的分类和聚类任务放入其中，在这个框架下实现数据的小体积划分，然后利用信息论，使用有损编码理论测量高维数据分布的维度和体积<sup>[9]</sup>。从而实现对于高维数据样本不同特征（如图 3-2）的压缩，描述聚类划分的现象。

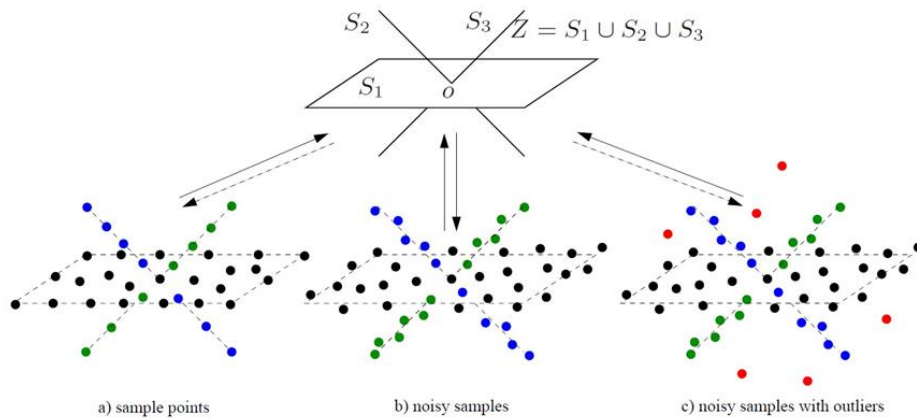


图 3-2 高维空间数据样本特征<sup>[1]</sup>

对于分布在一个或多个子流形上  $X \subset \cup_j^k \mathcal{M}_j$ ，在高维空间数据  $X = [x_1, x_2, \dots, x_m] \in$



$\mathbb{R}^{D \times m}$ , 满足精度  $\|x - \hat{x}\|_2 \leq \epsilon$  并受到如下约束<sup>[15]</sup>:

$$L(\mathbf{X}, \epsilon) \doteq \left( \frac{m+D}{2} \right) \log \det \left( \mathbf{I} + \frac{D}{m\epsilon^2} \mathbf{X}\mathbf{X}^T \right), \quad (3-1)$$

公式 (3-1) 是定义了分段数据的编码长度, 为了在给定失真条件下, 最大限度地减少分段数据的总编码长度。 $m$  和  $D$  是矩阵  $\mathbf{X}$  的维度, 前者是数据的特征数量, 后者是样本的数量, 矩阵  $\mathbf{I}$  是  $m$  大小的单位阵,  $\epsilon$  是允许的最小失真参数。从数据压缩有损编码和率失真理论角度说明了最小编码长度可以在信息量传输的可接受失真水平下可以权衡。

在分类任务中, 对新样本计算与各个类之间的距离, 若哪类数据可以用最少的编码增量来解释新样本, 则该样本就属于哪一类别。相比于传统方法, 最小增加编码长度<sup>[16]</sup> (Minimum Incremental Coding Length, MICL) 可以实现数据低维空间的表示, 得到一个真正的利用外插的分类器, 就是利用已知样本的归类来判别未知样本的类别。

$$\hat{y}(x) \doteq \arg \min_j \delta L_\epsilon(x, j), \quad (3-2)$$

公式 (3-2) 描述了对测试样本进行编码所需的最小额外位数, 对测试样本与  $j$  类之间的差异,  $\delta$  代表的是指标函数, 若后面内容条件为真取 1, 反之则取 0。MICL 分类器可以很好地用于编码函数长度式 (3-1) 高斯近似。实现通过有损编码和压缩来聚类, 通过最小增量编码长度来分类的效果。

### 3.1.2 线性判别表示

为了在得到数据特征表示, 如图 3-1 通过非线性映射  $f = (x, \theta)$  将数据变为线性结构, 最大限度地将数据实现聚类和分类。让同类数据尽可压缩在一个低维的线性子空间上面; 让不同类数据尽量分开。实现上述要求就是要寻求一个线性判别表示(LDR),  $\mathbf{Z} = [z_1, \dots, z_m] \in \mathbb{R}^d$  ( $d \ll D$ ) 对于数据  $\mathbf{X} = [x_1, \dots, x_m] \in \mathbb{R}^D$  满足:

$$\mathbf{X} \subset \mathbb{R}^D \xrightarrow{f(x, \theta)} \mathbf{Z} \subset \mathbb{R}^d \xrightarrow{g(z, \eta)} \hat{\mathbf{X}} \approx \mathbf{X} \in \mathbb{R}^D,$$

映射高维数据  $\mathbf{X}$  到低维特征表示  $\mathbf{Z}$ , 让杂乱无章但有结构的数据, 使用紧凑编码来表示。

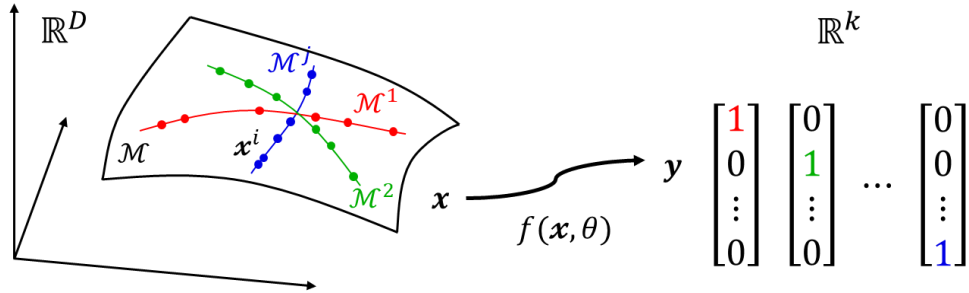


图 3-3 映射转换为低秩结构

LDR 模型指的是将可能位于高维空间中一系列低维子流形上的数据分别转换为一系列独立的低维线性子空间, 并用一系列特征值来表示, 如图 3-3 所示。

### 3.1.3 基于香农信息论的解释

从分类和聚类的学习任务上, 可基于两个优化目标构造: 一是要尽可能使得所有特征占据的空间足够张开; 二是尽可能使同类的特征空间收缩得足够小。具体地说, 要实现数据的表示, 当其符合某种低秩结构时, 样本表达能力最大; 同时, 将不同结构数据间的差异尽可能清晰地体现出来。从信息论的角度, 编码长度函数为最优表达提供了度量, 当样本表达最优时, 其表达所占的空间能够最大化, 即整体的样本拥有最大的编码长度, 使其能够让不同结构的样本相互靠近。

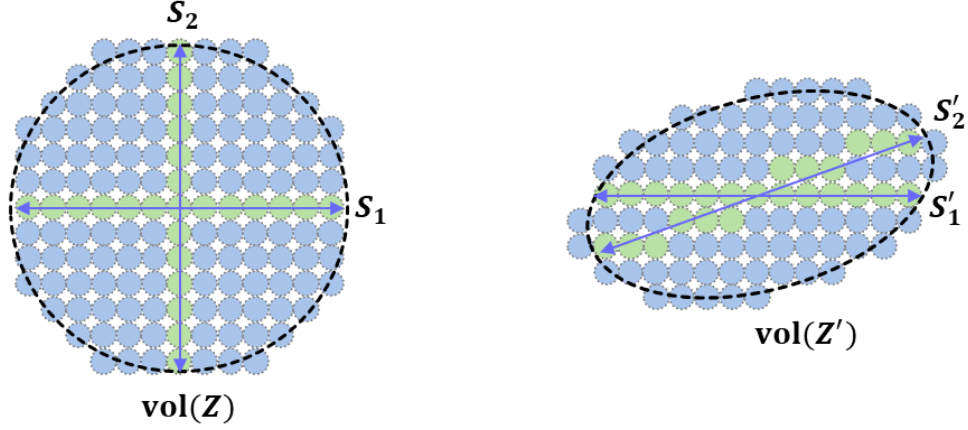


图 3-4 整体编码率尽可能大于类内编码率总和

图 3-4 使用圆球位置来表述信息的多少，通过对 $\mathbf{Z}$ 和 $\mathbf{Z}'$ 的比较， $R$ 是联合分布中填充球数量， $R^c$ 是所有子空间球数量之和， $\Delta R$ 是两者的差值。整体和部分的率失真差异称为速率降低，公式（3-3）所示，整体要最大限度地大于部分的总和如公式（3-4）所示：

$$\Delta R(\mathbf{Z}, \mathbf{\Pi}, \epsilon) = \underbrace{\frac{1}{2} \log \det \left( \mathbf{I} + \frac{d}{m\epsilon^2} \mathbf{Z}\mathbf{Z}^T \right)}_{R(\mathbf{Z})} - \underbrace{\sum_{j=1}^k \frac{\text{tr}(\mathbf{\Pi}_j)}{2m} \log \det \left( \mathbf{I} + \frac{d}{\text{tr}(\mathbf{\Pi}_j)\epsilon^2} \mathbf{Z}\mathbf{\Pi}_j\mathbf{Z}^T \right)}_{R^c(\mathbf{Z})}, \quad (3-3)$$

$$\max_{\mathbf{Z}} \Delta R(\mathbf{Z}(\theta), \mathbf{\Pi}, \epsilon) = R(\mathbf{Z}(\theta)) - R^c(\mathbf{Z}(\theta)|\mathbf{\Pi}, \epsilon), \quad \text{s. t. } \mathbf{Z} \in \mathcal{S}^{d-1}, \quad (3-4)$$

$R(\mathbf{Z})$ 是总体空间的编码率， $R^c(\mathbf{Z})$ 是类内空间编码率<sup>[9]</sup>，后者指的是数据进行降维后，同一类的样本表示的空间分布特征，从数学角度来说，是同一类样本平均距离相对整体的低维空间平均距离比值。

第一项要求原始输入 $\mathbf{X}$ 与转换特征 $\mathbf{Z}$ 之间的互信息尽可能比较大，也就是信息熵尽可能大，使得展开的空间体积比较大。第二项则是转换 $\mathbf{Z}$ 与标签 $\mathbf{Y}$ 之间互信息尽可能小，也就是类内的距离尽可能得小。然后使用投影梯度上升法（Projected Gradient Ascent, PGA）逐步更新 $\mathbf{Z}$ ，如图 3-5 所示：

$$\mathbf{Z}_{\ell+1} \propto \mathbf{Z}_{\ell} + \eta \cdot \left. \frac{\partial \Delta R}{\partial \mathbf{Z}} \right|_{\mathbf{Z}_{\ell}} \quad \text{subject to } \|\mathbf{Z}_{\ell+1}\| = 1. \quad (3-5)$$

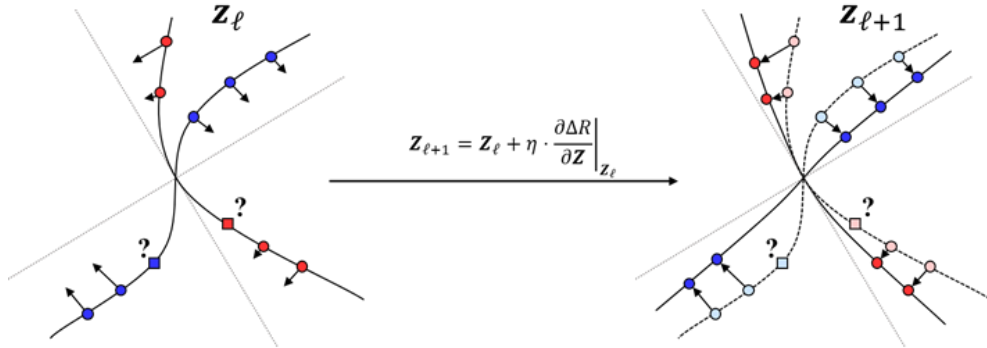


图 3-5 通过特征变换映射到低维

通过不断优化MCR<sup>2</sup>目标，当数据特征表达达到最优时，同类压缩更为紧凑，非同类间距拉大，自然导致 ReduNet 多层网络的生成。

### 3.2 基于 ReduNet 的高维数据低维特征最紧凑的表示

MCR<sup>2</sup>可以在深度网络的启发地解释分类和聚类，使得深度网络更好地设计参数和架构。通过 PGA 不断优化MCR<sup>2</sup>目标实现高维数据低维特征最紧凑的表示。

#### 3.2.1 ReduNet 模型

ReduNet 模型从第一性原理出发解释，通过数学公式推导优化构造，是一个基于深度神经网络的高维数据特征抽取方法，通过推导最大编码速率降低MCR<sup>2</sup>的梯度自然构建目标。此外，通过对网络施加移位不变属性，可以仅使用数据和MCR<sup>2</sup>推导出卷积运算符目标函数，从而使网络设计具有原则性和可解释性。

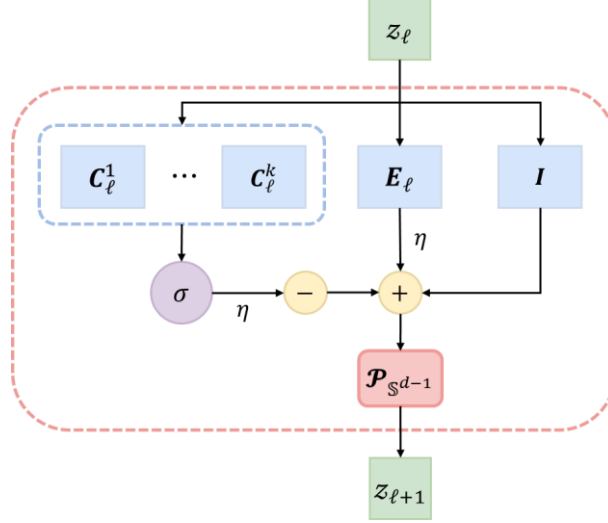


图 3-6 ReduNet 的实现原理

如图 3-6 所示的网络可表述梯度下降的过程。其中，每个不同的  $C$  通道对应一个类别的分类器， $E$  是扩张整个特征集的操作矩阵，梯度下降的更新过程是在当前特征加上得到的梯度。

$$\left. \frac{\partial R(\mathbf{Z})}{\partial \mathbf{Z}} \right|_{\mathbf{z}_\ell} = \alpha (\mathbf{Z} + \alpha \mathbf{Z}_\ell \mathbf{Z}_\ell^*)^{-1} \mathbf{Z}_\ell \doteq \mathbf{E}_\ell \mathbf{Z}_\ell \approx \alpha [\mathbf{Z}_\ell - \alpha \mathbf{Z}_\ell^* (\mathbf{Z}_\ell^* \mathbf{Z}_\ell)], \quad (3-6)$$

$$\mathbf{z}_{\ell+1} \propto \mathbf{z}_\ell + \eta \cdot \underbrace{[\mathbf{E}_\ell \mathbf{z}_\ell + \sigma([C_\ell^1 \mathbf{z}_\ell, \dots, C_\ell^k \mathbf{z}_\ell])]}_{g(\mathbf{z}_\ell, \theta_\ell)}, \quad \text{s.t. } \mathbf{z}_{\ell+1} \in \mathbb{S}^{d-1}, \quad (3-7)$$

$$f(\mathbf{x}, \theta) = \phi^L \circ \phi^{L-1} \circ \dots \circ \phi^0(\mathbf{x}), \text{ with } \phi^\ell(\mathbf{z}_\ell, \theta_\ell) \doteq \mathcal{P}_{\mathbb{S}^{d-1}}[\mathbf{z}_\ell + \eta \cdot g(\mathbf{z}_\ell, \theta_\ell)]. \quad (3-8)$$

公式 (3-6)、(3-7)、(3-8) 描述了用于 ReduNet 网络中速率降低目标迭代梯度上升方案，包含当前层输出、误差项和非线性激活函数的线性组合， $g(\mathbf{z}_\ell, \theta_\ell)$  包括非线性激活函数和卷积滤波器， $\mathbf{z}_\ell$  是当前输出层， $\mathbf{z}_{\ell+1}$  是下一层输出， $\mathbf{E}_\ell$  是误差项， $C_\ell^1$  是当前层卷积滤波器， $\sigma$  表示非线性激活函数，可以是 ReLU (Rectified Linear Unit, 校正线性单元) 或者 Sigmoid 函数。网络中每一层重复此过程，构建起 ReduNet 模型。

公式 (3-7) 中残差  $\|C_\ell^j \mathbf{z}_\ell\|$  使用 Softmax 函数评估样本所属该关系  $\pi^j(\mathbf{z}_\ell)$ :

$$\pi^j(\mathbf{z}_\ell) \approx \hat{\pi}^j(\mathbf{z}_\ell) \doteq \frac{\exp(-\lambda \|C_\ell^j \mathbf{z}_\ell\|)}{\sum_{j=1}^k \exp(-\lambda \|C_\ell^j \mathbf{z}_\ell\|)} \in [0, 1], \quad (3-9)$$

公式 (3-9) 确保函数输出是所有输出类的概率分布，而对于残差的权重约束如下：

$$\sigma([C_\ell^1 z_\ell, \dots, C_\ell^k z_\ell]) \doteq \sum_{j=1}^k \gamma_j C_\ell^j z_\ell \cdot \hat{\pi}^j(z_\ell) \in \mathbb{R}^d,$$

在其他方案中，可使用 ReLU 强制使所有特征在第一象限，公式（3-10）中 $\sigma$ 的几何意义是根据每个特征所属的子空间计算每个特征的残差：

$$\sigma(z_\ell) \approx z_\ell - \sum_{j=1}^k \text{ReLU}(P_\ell^j z_\ell), \quad (3-10)$$

ReduNet 可以实现识别的平移不变性，同时在深度网络使用循环矩阵表示样本，如图 3-7 将循环移位后的样本视为同一组，将一组样本编码到不同的低秩空间，因此与 $\mathbf{E}$ 、 $\mathbf{C}$ 矩阵进行矩阵相乘时，网络的操作自然等价于循环矩阵， $\mathbf{E}$ 、 $\mathbf{C}$ 在考虑不变性后，自动产生了卷积效果，可用来解释数据的稀疏性和不变性之间的关系。

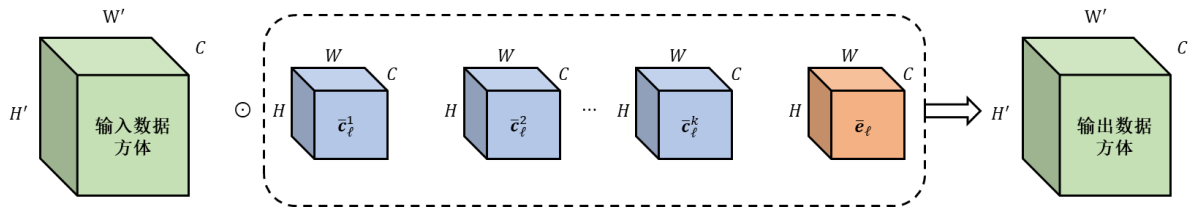


图 3-7 扩张矩阵和压缩矩阵经平移不变性成为卷积算子

ReduNet 网络中的结构参数可以通过正向传播(Forward Propagation, FP)计算得到，并不需要经过反向传播(Back Propagation, BP)计算，由此获得的所有的算子和参数具备精确的基于几何、统计和优化的意义。由于 ReduNet 是从优化一个最基本的信息量直接严格推导而来，没有任何主观、经验的因素，所以可认为这是基于第一性原理得来的神经网络。

ReduNet 网络的结构与目前通用的网络非常相似，以往的深度网络的架构和运算符主要是根据经验提出设计的，并通过反向传播算法作为一个黑盒子训练。下表 3-1 是基于数据压缩原理的深度网络与传统的深度网络进行了对比，左栏是深度网络传统的特点，右栏是基于新框架下的特点：

表 3-1 传统网络与 ReduNets 的比较

|       | 传统 DNNs | 压缩 ReduNets   |
|-------|---------|---------------|
| 优化目标  | 输入输出拟合  | 速率降低          |
| 深度架构  | 检验和误差   | 迭代优化          |
| 网络层算子 | 经验      | 投影梯度          |
| 移位不变性 | CNNs+增强 | 不变 ReduNets   |
| 初始化   | 随机/预先设计 | 前向计算          |
| 训练方式  | 后向传播/固定 | 正向/后向传播       |
| 可解释性  | 黑盒      | 白盒            |
| 表示    | 不明确     | 不相干的子空间 (LDR) |

### 3.2.2 压缩闭环转录

自洽原则指的是自主智能系统通过最小化输入的信息和经过处理的信息之间的内部差异，寻求一种最自洽的外部世界观察模型<sup>[10]</sup>。而 ReduNet 正基于这种原则构造的模型，其蕴含着反馈控制的思想，具体表现为通过对模型的输出与标签之间的误差进行反馈调整，实现模型参数的优化。ReduNet 在训练过程中，将输出与标签之间的误差作为反馈信号，传递回网络的隐藏层，对模型参数进行调整，从而达到提高模型性能的目的。

若要对外界信息做出完整的解释，由于深度网络可以学习特征复杂的映射关系，可以最大化信息增益，闭环转录则是自主自洽的基本学习单位<sup>[10]</sup>。基于自洽原则，对于特征 $\mathbf{Z}$ 建立闭环反馈如图 3-8 所示，图中描述了数据分布在特征空间中的闭环误差，并且其映射关系如下：

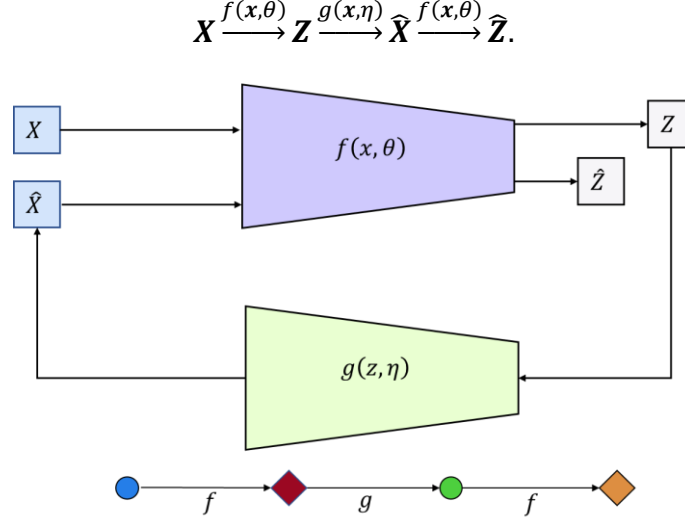


图 3-8 特征  $\mathbf{Z}$  在闭环反馈流向示意

解码控制器 $g$ 用来最小化 $\mathbf{X}$ 和 $\hat{\mathbf{X}}$ 之间的差异：

$$d(\mathbf{X}, \hat{\mathbf{X}}) \doteq \min_{\eta} \sum_{j=1}^k \Delta R(\mathbf{Z}_j, \hat{\mathbf{Z}}_j) = \min_{\eta} \sum_{j=1}^k \Delta R(\mathbf{Z}_j, f(g(\mathbf{Z}_j, \eta), \theta)).$$

编码感知器 $f$ 用来放大 $\mathbf{X}$ 和 $\hat{\mathbf{X}}$ 之间的差异：

$$d(\mathbf{X}, \hat{\mathbf{X}}) \doteq \max_{\theta} \sum_{j=1}^k \Delta R(\mathbf{Z}_j, \hat{\mathbf{Z}}_j) = \max_{\theta} \sum_{j=1}^k \Delta R(f(\mathbf{X}_j, \theta), f(\hat{\mathbf{X}}_j, \theta)).$$

整个在编码器 $f$ 和解码器 $g$ 的  $max-min$  博弈就是让 $f$ 最大化数据 $\mathbf{X}$ 中特征 $\mathbf{Z}$ 的速率降低、 $g$ 最小化数据 $\hat{\mathbf{X}}$ 中特征 $\hat{\mathbf{Z}}$ 的速率降低<sup>[17]</sup>。

一个对数据 $\mathbf{X} = \cup_{j=1}^k \mathbf{X}_j$ 学习多类 LDR  $max-min$  博弈环节由两个公式解释：

$$\max_{\theta} \min_{\eta} \underbrace{\Delta R(f(\mathbf{X}, \theta))}_{\text{Expansive encode}} + \underbrace{\Delta R(h(\mathbf{X}, \theta, \eta))}_{\text{Compressive decode}} + \sum_{j=1}^k \underbrace{\Delta R(f(\mathbf{X}_j, \theta), h(\mathbf{X}_j, \theta, \eta))}_{\text{Contrastive \& Contractive}}$$

with  $h(x) \doteq f \circ g \circ f(x)$ , 同样等于下式

$$\max_{\theta} \min_{\eta} \Delta R(\mathbf{Z}(\theta)) + \Delta R(\hat{\mathbf{Z}}(\theta, \eta)) + \sum_j \Delta R(\mathbf{Z}_j(\theta), \hat{\mathbf{Z}}_j(\theta, \eta)), \quad (3-11)$$

公式 (3-11) 涉及了编码器和解码器之间的信息表示的  $max-min$  博弈，目标就是降低速率。公式包括了三个部分扩展编码（原始数据与编码数据之间的速率差异）、压缩解码（编码数据和解码数据之间的速率差异）、对比和收缩编码（每个类别编码和解码数据之间的差异），可以通过编码函数 $\mathbf{Z}$ 和解码后的 $\hat{\mathbf{Z}}$ 来简化目标。速率降低的目标函数将表示结构化，不同类别的映射到相应的独立空间。

在最新文献<sup>[17]</sup>中，通过闭环转录框架学习数据集结构化生成的模型，用于特征空间



中子空间的高斯混合之间的距离计算，避免在特征空间中评估最小化近似距离。公式(3-11)也统一了自动编码和 GAN(Generative Adversarial Network, 生成对抗网络)的优势，那在 LDR 框架下，学习的特征在内部结构变得透明，基于速率降低的目标可以在广泛的数据集合网络架构中稳定优化。

### 3.3 实验设计与模型测试

深度网络的层次结构就是在做迭代的优化，数据的压缩，逐层优化数据特征的表达。本实验部分将验证MCR<sup>2</sup>构造生成 ReduNet、ReduNet 在 $\mathbb{S}^2$ 高斯混合空间学习状况以及 ReduNet 的平移不变性。

以多层 ReduNet 模型为例，借助 MNIST 数据集，验证 ReduNet loss 快速收敛的效果；在 $\mathbb{S}^1$ 和 $\mathbb{S}^2$ 中学习高斯混合情况，观察数据分布聚类 and 分类效果，说明 ReduNet 的可解释性；验证基于 MNIST 数据集的 ReduNet 二维平移不变性，探究高维数据低维特征提取性能的变化规律并进行可视化。将其与传统的降维方法和深度网络进行比较，说明 ReduNet 模型的优势。

#### 3.3.1 实验环境配置与数据集

实验环境使用 Google Colab 平台环境、PyTorch 框架实现，会使用到 Numpy、Pandas、sklearn 库等科学数据分析库，Google Colab 可以提供 Nvidia Tesla T4 GPU，可以提升数据集的运算效率。在 Colab Cells 中使用 Shell 脚本导入项目，根据 Demo 的说明文档编码调参实现。实验中主要使用到 MNIST 数据集，这是一个经典的手写数字识别数据集，有 0-9 数字在黑白图片中的手写样本，每个图片的像素是 28×28，共有 70000 张图片，其中 60000 张为训练集，10000 张为测试集。MNIST 数据集因为其易于理解和相对简单的特点，在深度学习常用来测试算法的性能，也可以方便地与其他算法进行比较。MNIST 数据集相比其他 Coco 数据集等体积小，训练速度较快，可以节省训练时间。

#### 3.3.2 模型的设计

基本的模型搭建步骤如下图 3-9 所示：

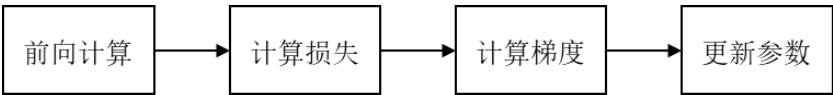


图 3-9 神经网络搭建过程

编码实现流程是 ReduNet 类首先通过 Vector 类（代码详见附录）搭建网络层结构，传入相应的参数  $layers$ 、 $\epsilon$ 、 $\eta$  等，并且定义了正向传播函数，传入 ReduLayer 形参，计算  $E$  和  $C$  矩阵，实现梯度更新，计算 MCR<sup>2</sup> 损失，使用 compute\_mcr2 函数。如图 3-10 编码实现 ReduNet 模型。

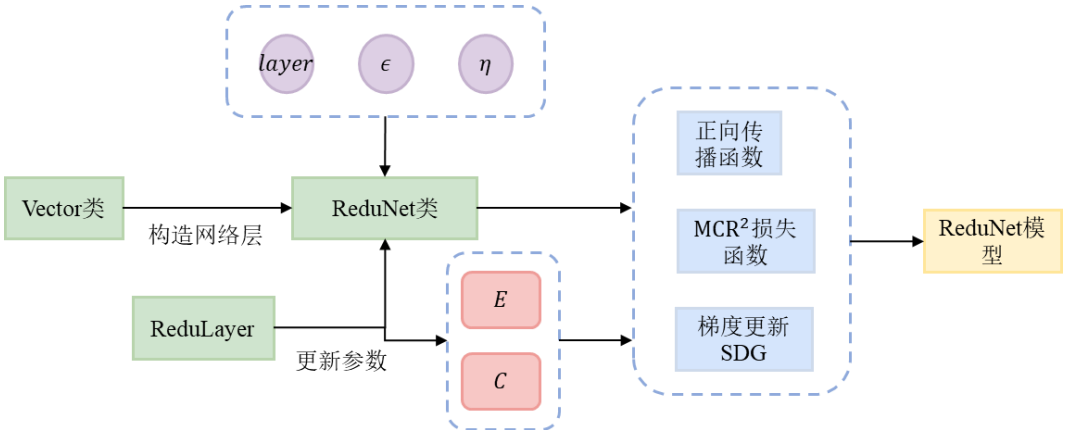


图 3-10 ReduNet 编码实现流程

### 3.3.3 模型的训练与测试

以下是在 Google Colab 中键入相应的命令行。

验证 ReduNet 正向构造的 $\Delta R$ 收敛效果：使用 MNIST 数据集，层数分别为 10 层和 50 层，1000 个样本观察曲线情况 [ `!python train_forward.py --data mnistvector --arch layers10 --samples 1000` ]。

分别在 $S^1$ 和 $S^2$ 中学习高斯混合：定义层数 layers = 2000 或 200 层，在步长 $\eta = 0.5$ 和精度 $\epsilon = 0.1$ ，分别使用两个不同的数据分布，将数据点投影到混合空间上 [ `!python gaussian3d.py --data 1 --noise 0.1 --samples 500 --layers 2000 --eta 0.5 --eps 0.1` ]。

验证 MNIST 数字的二维平移不变性 (2D Translation Invariance)：输入 10 类、10 样本、外通道 75、层数 25、步长 $\eta = 0.5$ 、精度 $\epsilon = 0.1$ ，评估得到的结果。Colab 训练了约 19 分钟 [ `!python mnist2d.py --classes 0 1 2 3 4 5 6 7 8 9 --samples 10 --layers 25 --outchannels 75 --ksize 9 --eps 0.1 --eta 0.5` ]。

```
[ ] from google.colab import drive
    drive.mount("/content/drive")

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

[ ] import os
    path="/content/drive/My Drive/ReduNet/examples"
    os.chdir(path)
    os.listdir(path)

['utils_example.py', '__pycache__', 'gaussian3d.ipynb', 'gaussian2d.ipynb']

[ ] import os
    os.chdir('../')

[ ] from redunet import ReduNetVector
    import examples.utils_example as ue
    import plot
```

(a) 设置相应的训练路径与包

```
!python train_forward.py --data mnistvector --arch layers50 --samples 1000
!python gaussian3d.py --data 1 --noise 0.1 --samples 500 --layers 2000 --eta 0.5 --eps 0.1
!python mnist2d.py --classes 0 1 2 3 4 5 6 7 8 9 --samples 10 --layers 10 --outchannels 25 --ksize 9 --eps 0.1 --eta 0.5
!python evaluate.py --knn
!python train_forward.py --data mnistvector --arch layers10 --samples 1000

./saved_models/forward/mnistvector+layers10/samples1000
0 | 85.465027 289.690002 204.224976
1 | 98.502640 301.192200 202.689560
2 | 113.203522 314.548859 201.345337
3 | 129.504486 329.748230 200.243744
4 | 147.278366 346.684692 199.406326
5 | 166.334991 365.163696 198.828705
6 | 186.429123 384.919250 198.490128
7 | 207.277084 405.639709 198.362625
8 | 228.579102 426.996063 198.416962
9 | 250.041702 448.666504 198.624802
Plot saved to: ./saved_models/forward/mnistvector+layers10/samples1000/figures/loss_mcr/train.png
./saved_models/forward/mnistvector+layers10/samples1000

[ ] !python evaluate_forward.py --model_dir ./saved_models/forward/mnistvector+layers10/samples1000

Loading checkpoint: ./saved_models/forward/mnistvector+layers10/samples1000/checkpoints/model.pt
train
test
`C
```

(b) 输入相应的命令行

图 3-11 模型测试与训练

基本的训练过程如图 3-11，设置相应程序的路径、导入应用包，然后根据图 3-10 设置对应的参数，得到 *loss* 值，评估训练得到 *loss* 曲线图，并使用了 *matplotlib* 库绘制训练、测试效果图。

### 3.4 实验结果分析

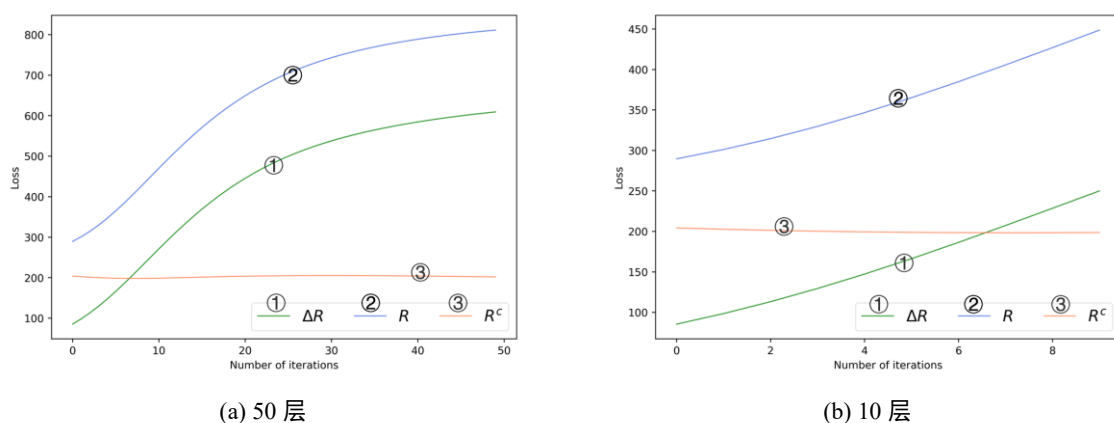


图 3-12 使用 MNIST 数据集构造 ReduNet

图 3-12 使用 MNIST 数据集在不同层数下的正向构造，可以看出 ReduNet 需要不断迭代优化达到  $R$ 、 $R^c$  两类编码率平稳收敛，说明该模型具有较好的计算成本，在训练过程具有较强的可解释性。

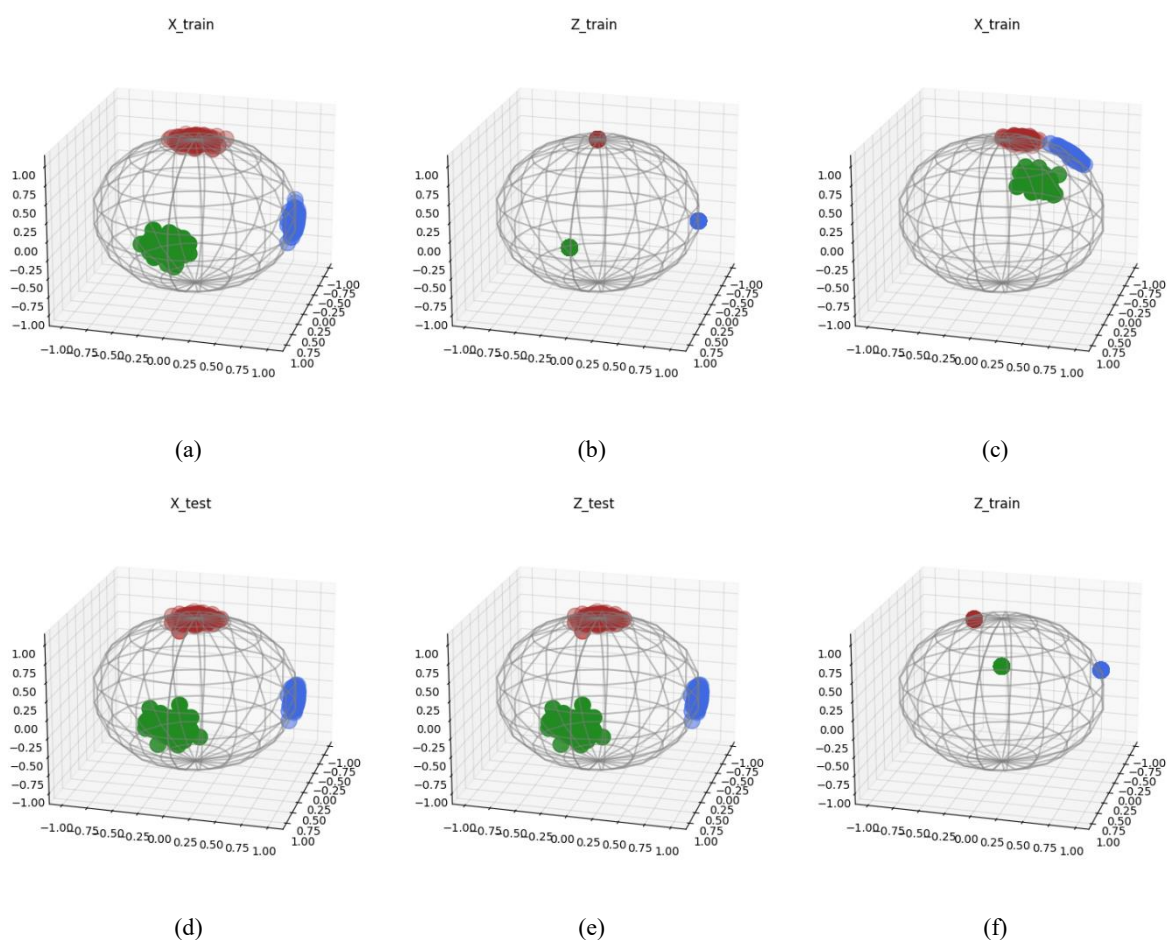


图 3-13 3D 高斯混合的原始样本和学习表示，散点图可视化数据 X 和样本特征 Z



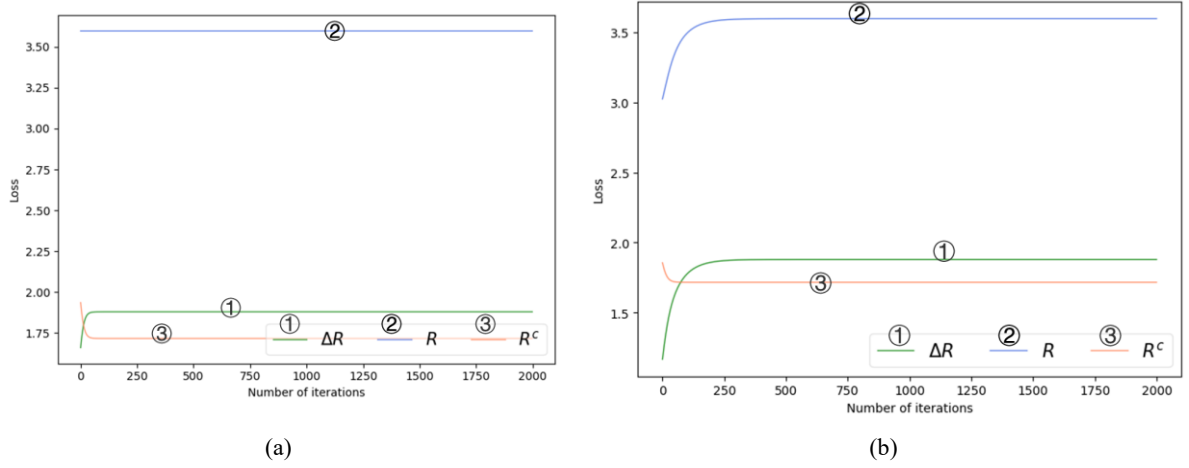


图 3-14 3D 高斯混合原始样本学习表示的  $loss$  曲线

在图 3-13 组图中， $\mathbb{S}^2$  空间三种颜色代表不同类别样本，可以观察到经过映射  $f(x, \theta)$  [见 3.1.1 节]，来自每一类的数据收敛到同一簇，角度呈三维正交，正如图 3-13 (a-b) 所示。构建的网络能够最大化  $MCR^2$  的损失并且收敛稳定，同一类样本始终在训练样本的相同聚类中心。图 3-14 (a) 展示了采用正交角度数据分布在 2000 层训练的效果， $\Delta R$  比较稳定。而修改生成数据分布的角度后，如图 3-13 (c, f) 所示，两两之间聚类的角度是  $\pi/3$ ，图 3-14 (b) 是该训练下的  $loss$  曲线，在起始段  $R$  并没有快速收敛，这与  $MCR^2$  有损效果是一致的，在 250 层迭代后收敛效果稳定。

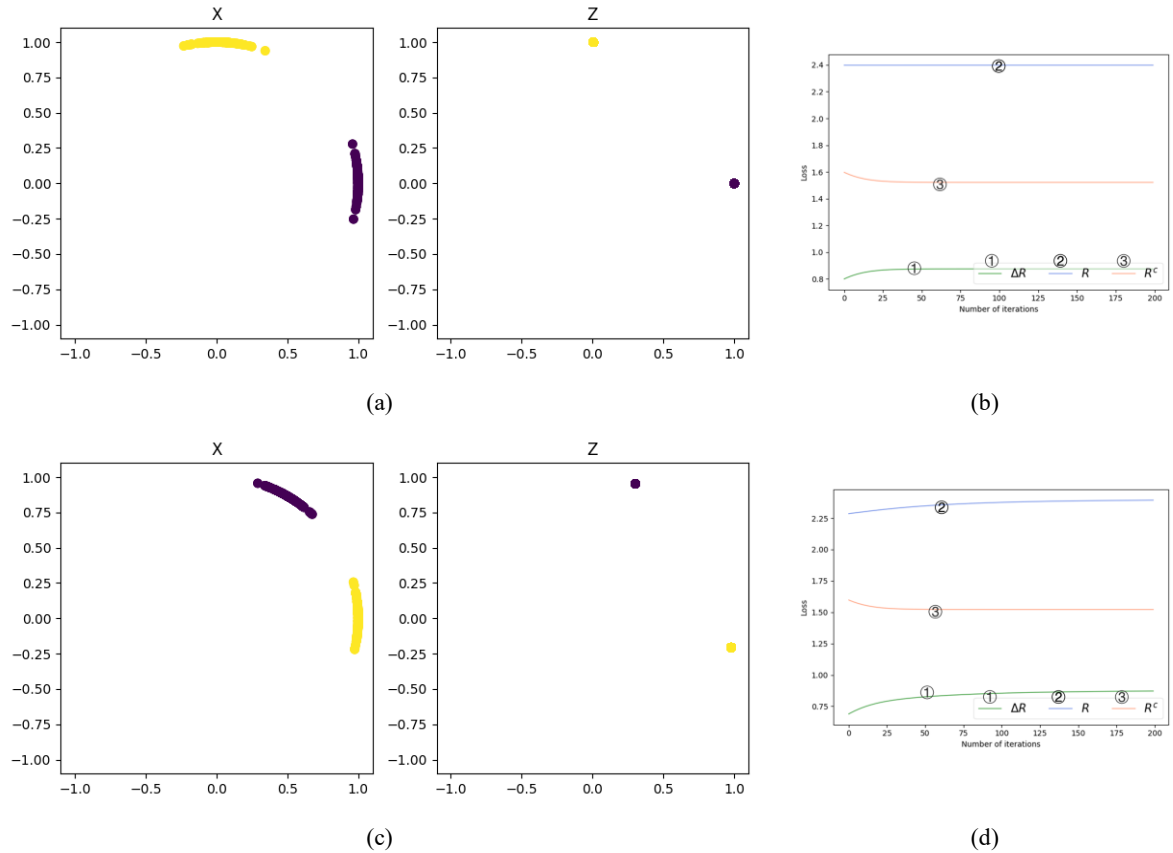


图 3-15 2D 高斯混合的原始样本和学习表示，散点图可视化数据  $X$  和样本特征  $Z$

在图 3-15 (a) 是在同一样本聚类角度为  $\pi/2$  的训练效果，图 3-15 (b)  $loss$  曲线正如 3D 一样训练稳定。图 3-15 (c) 是在统一样本聚类后角度  $\pi/3$  的效果。通过以上可以看到 ReduNet 可以将不同类别的数据点映射到正交子空间上。 $MCR^2$  促进了将数据嵌入

到多个独立子空间，使得特征根各向同性地分布在每个子空间<sup>[11]</sup>。

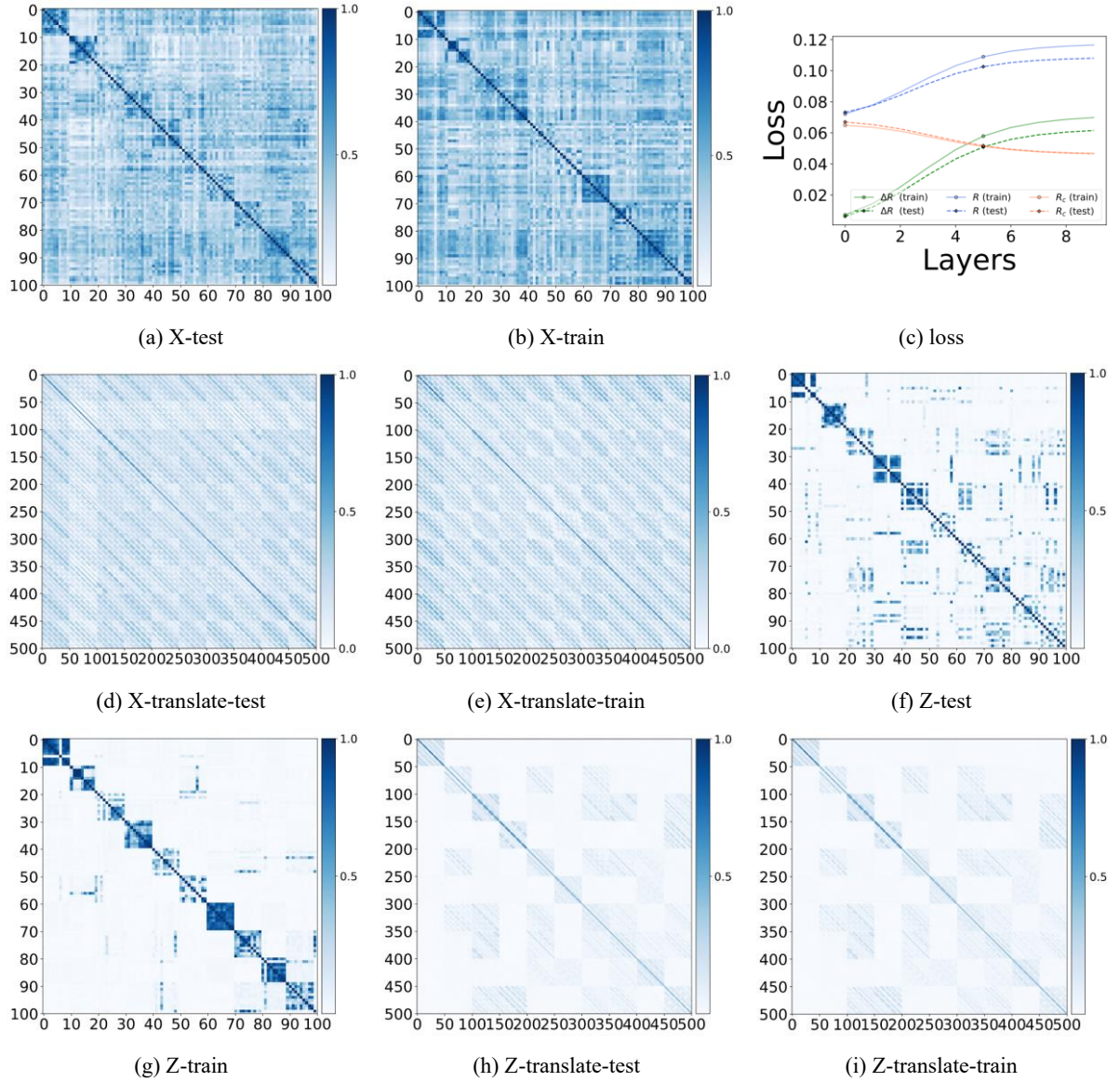


图 3-16 ReduNet 的平移不变性

如图 3-16 所示, (a,b,e-i) 是移位训练测试数据的 *cosine* 相似热图, 可以看到 ReduNet 能够将来自不同类别的训练数据映射到 10 个几乎正交的子空间, 如图 (h-i), 也就是说学习子空间的对极角的移位变换是不变的。如图 (c) 可以发现 ReduNet 能够最大化  $MCR^2$  损失并测试数据。

### 3.5 本章小结

本章从介绍  $MCR^2$  原理开始, 通过聚类 and 分类任务, 利用香农信息理论寻求信息的有效表示。然后在数据压缩和线性判别表示框架下, 通过优化  $MCR^2$  目标, 自然推导生成 ReduNet 模型。不同于通用的深度网络, ReduNet 是白盒方法, 具有可解释性和原则性。然后总结压缩闭环转录框架, 基于反馈、博弈论思想, 寻求信息的有效表达。最后通过编码验证相关结论。

## 4 结论与展望

在基于数据压缩和线性判别表示框架下, 学习给定数据的低维 LDR 目标中推导出

深度卷积网络作为白盒——ReduNet，它是基于梯度的迭代方案来优化速率降低目标，几乎共享现代深度网络的所有主要结构特征，它的架构、线性和非线性运算符，都具有精确的几何解释和统计解释，每一层线性算子具有数据自回归解释，提高了网络的可解释性和稳定性，用于分类任务（例如对象检测或识别）时数据可较好地表示，具有在图像分类、人脸识别、语音文本规模应用的潜力。

ReduNet 模型中所有层次结构都能利用前向传播算法显式构造得出，无需反向传播算法进行学习，并直接推导出多通道卷积的结构，而反向传播可以对网络进行进一步细调。所以 ReduNet 网络本身的前向传播是在对编码率进行优化，而反向传播可以看成是对正向的优化路径进行进一步变分优化，从而使用前向构建网络的初始化比随机初始化具有优势。

ReduNet 还在建模不变性中，为多通道卷积在每层作用提供了新的拓展，其可作为增量算子来压缩或扩展用于学习数据的 LDR 的所有等变实例。基于压缩闭环转录新的框架，还揭示了稀疏性和不变性之间的权衡关系，即施加给每个类中的信号被稀疏生成。另外，ReduNet 模型在多模态具有高度非线性几何和拓扑中的实践意义，但需要更复杂的策略来控制压缩和扩展过程。在线性动态化处理附加结构、计算机图形处理上拥有较大的研究价值。

ReduNet 综合了香农信息理论、反馈控制论、博弈论等思想，蕴含了简约、自洽的原则，反映了自然智能系统能准确、有效地表示外部信息，具有较强的表达能力和泛化能力，而基于压缩闭环转录框架，则为 AI 感知、决策和智能提供了新的研究视角。

## 致谢

首先，感谢张明强老师的倾心指导，张老师是一位十分负责的老师，从他为我论文指导的过程中看到了他严谨治学的态度，启示了我在未来读研的学习工作中应该保持对学术研究的敬重。也感谢导师指定论文的选题让我接触到了新领域，跟踪到 AI 的前沿话题，拓展了新的视野，让我掌握了初步的科研工作方法，训练了批判性思维能力。

其次，感谢四年来各位老师倾心地传授学业知识，让我在曲园这个舞台不断取得进步，使我培养了自学自律能力，树立起了正确的人生观、价值观。也感谢四年来朝夕相处的同学们，努力的路上不总是孤独的，互相学习，互相进步，愿他们的未来前程似锦。

最后，感谢爸爸妈妈在我艰难的时刻不断鼓励我，让我在拼搏和奋斗的路上不感到疲倦。四年的故事即将画上句号，但这又是一个新的开始，愿我在未来学习工作中一切顺利。

## 参考文献

- [1] Wright J, Ma Y. High-Dimensional Data Analysis with Low-Dimensional Models: Principles, Computation, and Applications[M]. Cambridge University Press, 2022.
- [2] Nouinou S, El Afia A, El Fkihi S. Overview on Last Advances of Feature Selection[C]//Proceedings of the International Conference on Learning and Optimization Algorithms: Theory and Applications. 2018: 58.
- [3] Tian T S. Dimensionality Reduction for Classification with High-Dimensional Data[M]. University of Southern California, 2009.
- [4] Bellman R. Dynamic Programming and Lagrange Multipliers[J]. Proceedings of the National Academy

- of Sciences, 1956, 42(10): 767-769.
- [5] Pearson K. LIII. On Lines and Planes of Closest Fit to Systems of Points in Space[J]. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 1901, 2(11): 559-572.
  - [6] Fisher R A. The Use of Multiple Measurements in Taxonomic Problems[J]. Annals of Eugenics, 1936, 7(2): 179-188.
  - [7] Hérault, Jeanny et al. Détection de Grandeurs Primitives Dans un Message Composite par une Architecture de Calcul Neuromimétique en Apprentissage non Supervisé[J]. France, Nice, Proc.SSIP, 1985.
  - [8] Yu Y, Chan K H R, You C, et al. Learning Diverse and Discriminative Representations via the Principle of Maximal Coding Rate Reduction[J]. Advances in Neural Information Processing Systems, 2020, 33: 9422-9434.
  - [9] Chan K H R, Yu Y, You C, et al. Deep Networks from the Principle of Rate Reduction[J]. arXiv preprint arXiv:2010.14765, 2020.
  - [10] Ma Y, Tsao D, Shum H Y. On the Principles of Parsimony and Self-Consistency for the Emergence of Intelligence[J]. Frontiers of Information Technology & Electronic Engineering, 2022, 23(9): 1298-1323.
  - [11] Chan K, Yu Y, You C, et al. ReduNet: A White-Box Deep Network from the Principle of Maximizing Rate Reduction[J]. Journal of Machine Learning Research, 2022, 23(114).
  - [12] Shannon C E. Coding Theorems for a Discrete Source with a Fidelity Criterion[J]. IRE Nat. Conv. Rec, 1959, 4(142-163): 1.
  - [13] Cover T M, Thomas J A. Elements of Information Theory Second Edition Solutions to Problems[J]. Internet Access, 2006: 19-20.
  - [14] Wiener N. Cybernetics or Control and Communication in the Animal and the Machine[M]. MIT press, 2019.
  - [15] Ma Y, Derksen H, Hong W, et al. Segmentation of Multivariate Mixed Data via Lossy Data Coding and Compression[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(9): 1546-1562.
  - [16] Wright J, Tao Y, Lin Z, et al. Classification via Minimum Incremental Coding Length (MICL)[J]. Advances in Neural Information Processing Systems, 2007, 20.
  - [17] Dai X, Tong S, Li M, et al. CTRL: Closed-Loop Transcription to an LDR via Minimizing Rate Reduction[J]. Entropy, 2022, 24(4): 456.