**DEEP LEARNING笔记**

|  |  |
| --- | --- |
| **编 写** |  |
| **校 对** |  |
| **审 核** |  |
| **批 准** |  |

**北京树优科技有限公司**

**基于MBSE的新一代ZLDD领域专业模型适应性改造技术方案**

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

**北京树优科技有限公司**

目 录

[1. 模型顶层设计 3](#_Toc500922490)

[2. 攻防对抗体系的Sysml图形表示 7](#_Toc500922491)

[2.1. 攻防对抗体系图形描述 7](#_Toc500922492)

[2.1.1. 类图 7](#_Toc500922493)

[2.1.2. 需求图 7](#_Toc500922494)

[2.1.3. 用例图 8](#_Toc500922495)

[2.1.4. 时间图 9](#_Toc500922496)

[2.1.5. 顺序图 9](#_Toc500922497)

[2.2. 新一代战略导弹图形描述 9](#_Toc500922498)

[2.2.1. 类图 9](#_Toc500922499)

[2.2.2. 需求图 10](#_Toc500922500)

[2.2.3. 用例图 11](#_Toc500922501)

[2.2.4. 装配图 11](#_Toc500922502)

[2.2.5. 时间图 12](#_Toc500922503)

[2.2.6. 顺序图 12](#_Toc500922504)

[2.3. 突防模型图形描述 12](#_Toc500922505)

[2.3.1. 类图 12](#_Toc500922506)

[2.3.2. 需求图 13](#_Toc500922507)

[2.3.3. 用例图 13](#_Toc500922508)

[2.3.4. 装配图 14](#_Toc500922509)

[2.3.5. 时间图 14](#_Toc500922510)

[2.3.6. 顺序图 15](#_Toc500922511)

[3. 攻防对抗体系分解描述 15](#_Toc500922512)

[3.1. 体系组成 15](#_Toc500922513)

[3.2. 关联关系 16](#_Toc500922514)

[3.3. 攻防对抗进行流程 17](#_Toc500922515)

# 机器学习的步骤

## 数据预处理

### 导入数据

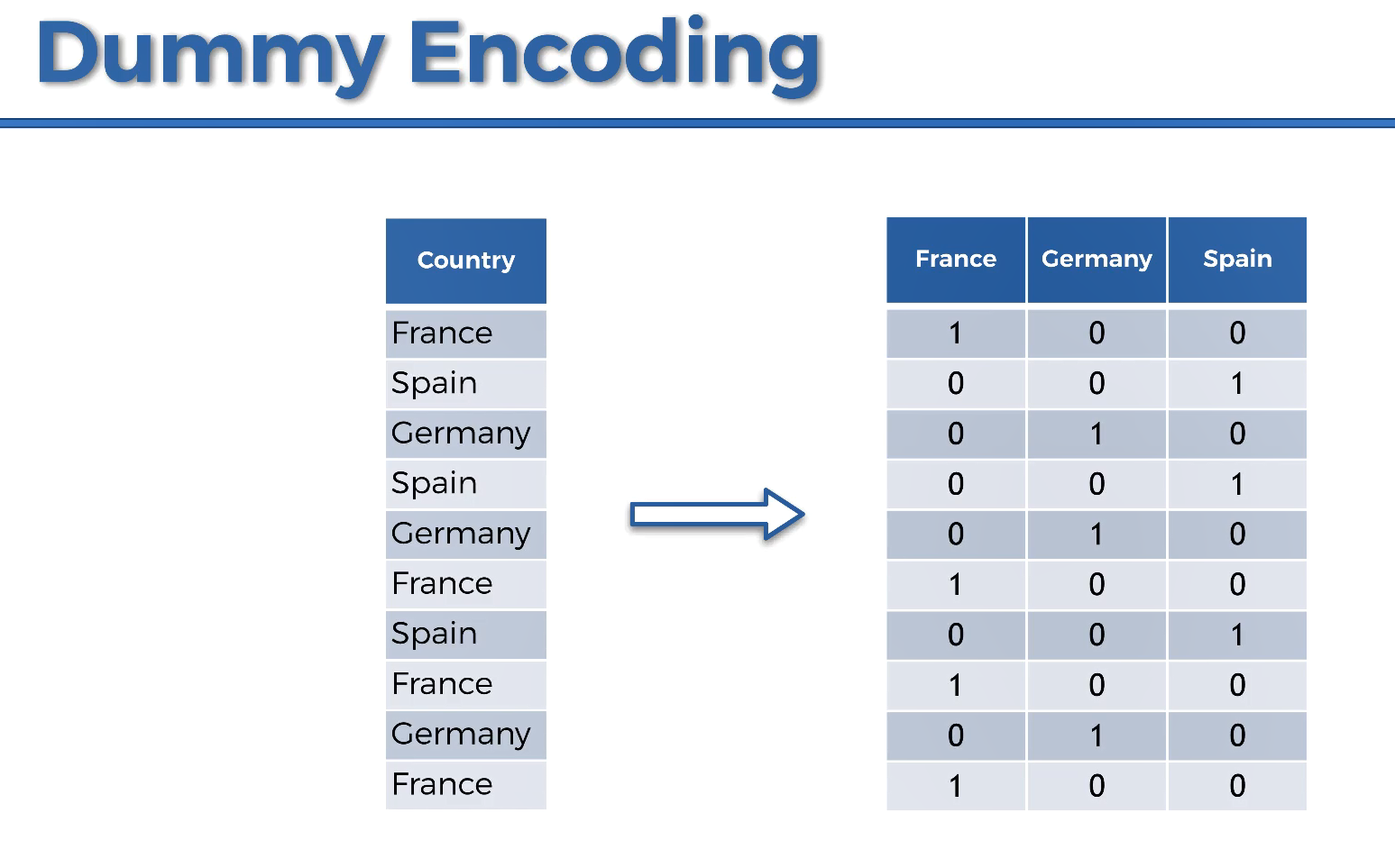
一般数据以csv格式导入，行列排列。

### 对缺失数据进行补充

不要因为有确实的数据就直接删除整个样本，用Imputer进行数据补齐，如采用平均值，中间值等方法。

### 对数据进行分类处理Categorical Data

将所有数据都转换成为数字形式，对字符串分类的数据（如国籍等）要进行编码处理。注意为了防止出现数字分类后因为你数字大小会影响计算，我们需要采取Dummy Encoding的方法



### 将数据拆分为训练数据和测试数据

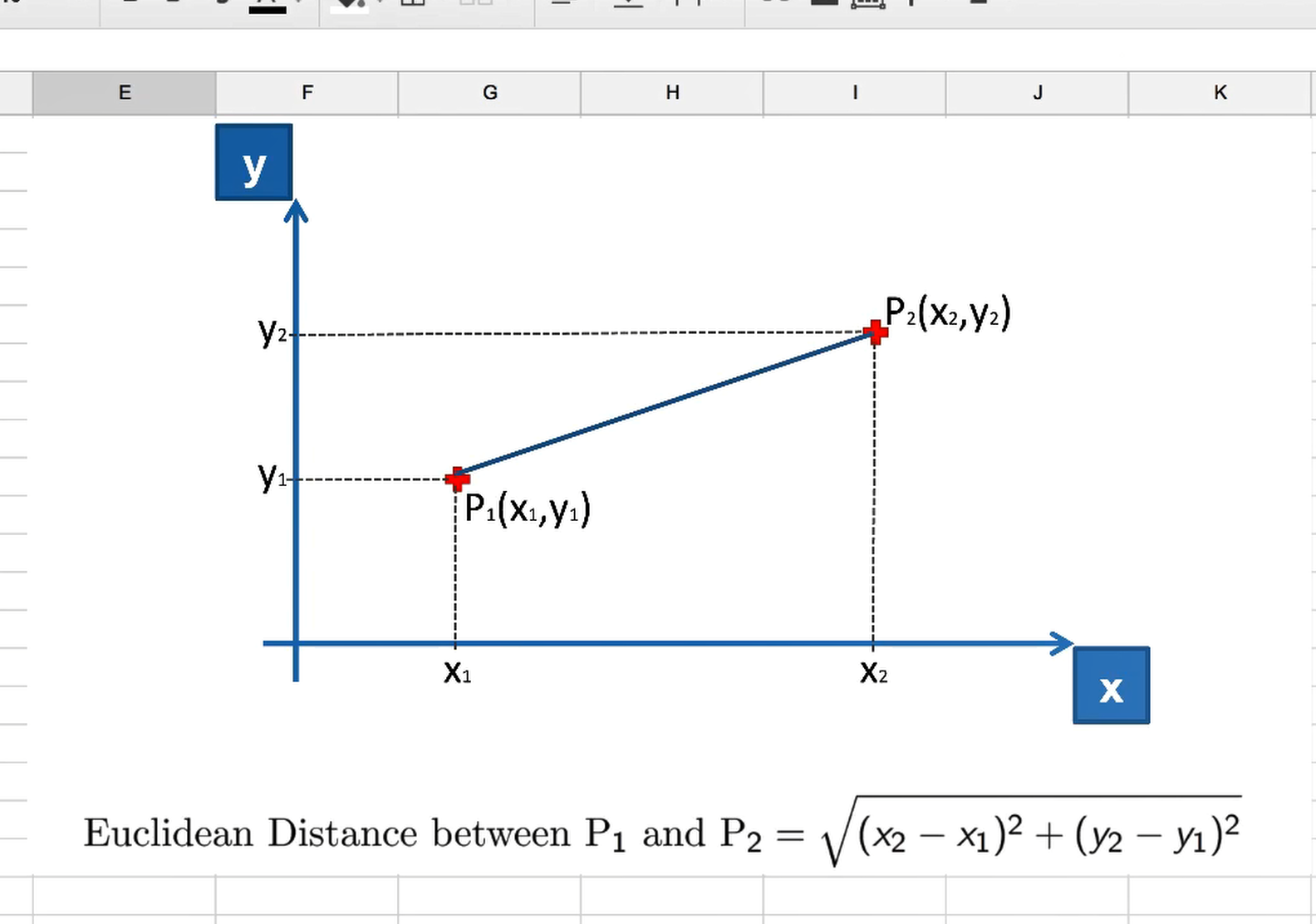
注意训练数据一般占比为80%左右，测试数据20%左右。

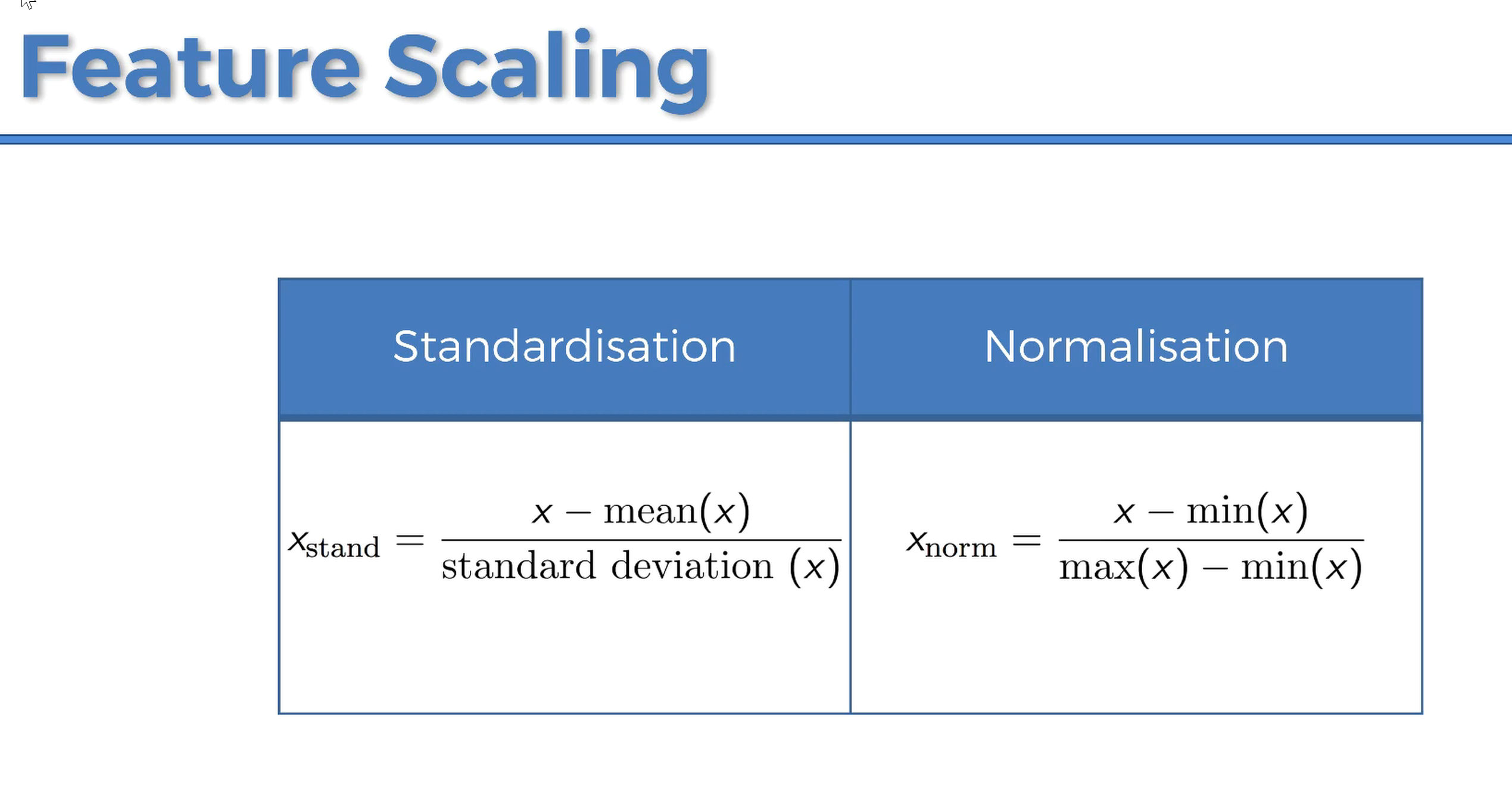
### 特征缩放（Feature Scaling）

欧氏距离

另一种是聚类分析法，采用离差平方和(Ward’s)法，利用欧氏距离(Euclidean)作为样品的聚类依据，以峰面积为样品谱图的特征值，将样品分为几类。

因此需要对样本的特征进行缩放处理





# 回归模型（Regression Model）

回归模型（线性或者非线性）是用来预测一个真值时使用的。如果独立变量是时间的话，你就可以用回归模型预测未来的值，否则就是用回归模型预测现在但是是不知道的值。回归技术途径涵盖了从线性回归到SVR回归和随机森林回归的全部。

In this part, you will understand and learn how to implement the following Machine Learning Regression models:

Simple Linear Regression（线性回归）

Multiple Linear Regression（多线性回归）

Polynomial Regression（多项式回归）

Support Vector for Regression (SVR) 向量机回归

Decision Tree Classification 决策树分类

Random Forest Classification 随机森林分类

# 深度学习的概念

人的视觉系统的信息处理是分级的。从低级的V1区提取边缘特征，再到V2区的形状或者目标的部分等，再到更高层，整个目标、目标的行为等。也就是说高层的特征是低层特征的组合，从低层到高层的特征表示越来越抽象，越来越能表现语义或者意图。而抽象层面越高，存在的可能猜测就越少，就越利于分类。

特征表示：稀疏编码（Sparse Coding）

反向传播算法（也叫Back Propagation算法或者BP算法）

2006年，加拿大多伦多大学教授、机器学习领域的泰斗Geoffrey Hinton和他的学生Ruslan Salakhutdinov在《科学》上发表了一篇文章，开启了深度学习在学术界和工业界的浪潮。这篇文章有两个主要观点：

1）多隐层的人工神经网络具有优异的特征学习能力，学习得到的特征对数据有更本质的刻画，从而有利于可视化或分类；

2）深度神经网络在训练上的难度，可以通过“逐层初始化”（layer-wise pre-training）来有效克服，在这篇文章中，逐层初始化是通过无监督学习实现的。

深度学习的实质，是通过构建具有很多隐层的机器学习模型和海量的训练数据，来学习更有用的特征，从而最终提升分类或预测的准确性。因此，“深度模型”是手段，“特征学习”是目的。

区别于传统的浅层学习，深度学习的不同在于：

1）强调了模型结构的深度，通常有5层、6层，甚至10多层的隐层节点；

2）明确突出了特征学习的重要性，也就是说，通过逐层特征变换，将样本在原空间的特征表示变换到一个新特征空间，从而使分类或预测更加容易。与人工规则构造特征的方法相比，利用大数据来学习特征，更能够刻画数据的丰富内在信息。

优。  
将除最顶层的其它层间的权重变为双向的，这样最顶层仍然是一个单层神经网络，而  
其它层则变为了图模型。向上的权重用于“认知”，向下的权重用于“生成”。然后使用WakeSleep算法调整所有的权重。让认知和生成达成一致，也就是保证生成的最顶层表示能够  
尽可能正确的复原底层的结点。比如顶层的一个结点表示人脸，那么所有人脸的图像应该  
激活这个结点，并且这个结果向下生成的图像应该能够表现为一个大概的人脸图像。  
Wake-Sleep算法分为醒（wake）和睡（sleep）两个部分。

1）wake阶段：认知过程，通过外界的特征和向上的权重（认知权重）产生每一层的抽象  
表示（结点状态），并且使用梯度下降修改层间的下行权重（生成权重）。也就是“如果现  
实跟我想象的不一样，改变我的权重使得我想象的东西就是这样的”。  
2）sleep阶段：生成过程，通过顶层表示（醒时学得的概念）和向下权重，生成底层的状  
态，同时修改层间向上的权重。也就是“如果梦中的景象不是我脑中的相应概念，改变我的  
认知权重使得这种景象在我看来就是这个概念

## 攻防对抗体系图形描述

### 类图

上图采用了层次化的方法说明系统的功能需求。通过该需求模型，开发者可以详细了解该系统的设计目标，以及组成该目标的各个功能及其分解。该需求模型只采用了若干表达形式中的一种，即树形结构，它便于从功能的角度分析系统的需求。

### 用例图



图 2‑3 SysML用例图

典型的用例模型如图所示。图中的方框代表了该用例作为一个独立系统时的边界，位于边界外的矩形标志代表了系统的角色，位于左边的角色是系统的使用者，是位于系统边界以外的；位于右边的角色是实现系统功能的实体，是包含在系统设计以内的，和对应的用例间用点画线连接。边界内的椭圆代表了系统的用例。方框内带箭头实线表明用例间的调用关系，如用例“气动计算”使用用例“总体参数”。

### 时间图



图 2‑2 攻防对抗的时间图

### 顺序图



图 2‑3 攻防对抗的顺序图