**问题：**

1.上次分享提到的前置知识空间域卷积，谱图卷积，傅里叶变换，Laplacian算子作用是什么？

2.上次分享提到的机器学习框架，说重点不是语言是对框架的了解，大家有没有对相关知识进行调研？

1. **关于数据结构-图**

图的数据可以用矩阵表示

<https://blog.csdn.net/luzaijiaoxia0618/article/details/104718146/>

度矩阵（Degree）

度矩阵是对角阵，对角上的元素为各个顶点的度。顶点vi的度表示和该顶点相关联的边的数量。

无向图中顶点vi的度d(vi)=N(i)。

邻接矩阵（Adjacency）

邻接矩阵表示顶点间关系，是n阶方阵（n为顶点数量）。

邻接矩阵分为有向图邻接矩阵和无向图邻接矩阵。无向图邻接矩阵是对称矩阵，而有向图的邻接矩阵不一定对称。

1. **拉普拉斯矩阵**

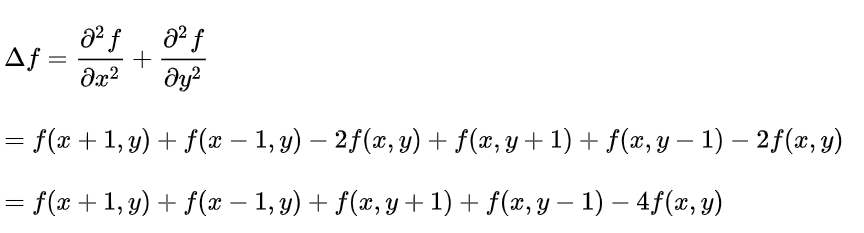
L=D（度矩阵）-A(邻接矩阵）

拉普拉斯矩阵的意义

普拉斯矩阵的第 i 行其实是第i个节点在产生扰动时对其他节点产生的收益累积。

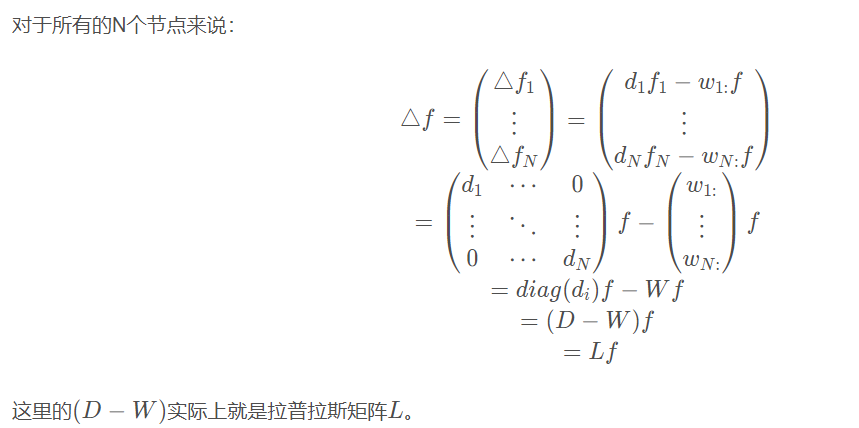
推倒过程

二维函数为例，上面考虑的都是连续状态下，在离散状态下，拉普拉斯算子的形式为：



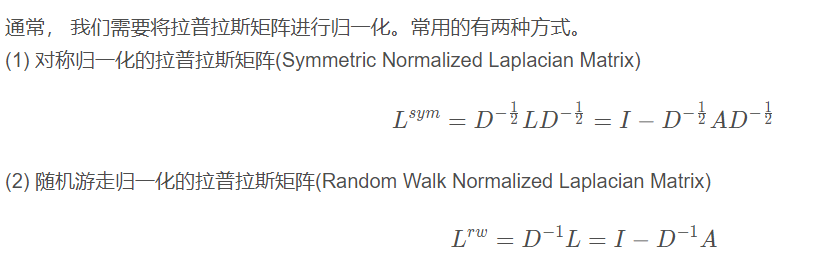
拉普拉斯矩阵是二维函数在图上面的推广

推倒过程：



参考：<https://blog.csdn.net/qq_45448654/article/details/122692366>

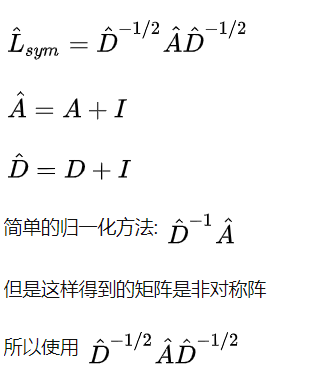
1. **什么是拉普拉斯矩阵归一化，为什么要归一化**



为什么要归一化：

采用加法规则时，对于度大的节点特征越来越大，而对于度小的节点却相反，这可能导致网络训练过程中梯度爆炸或者消失的问题

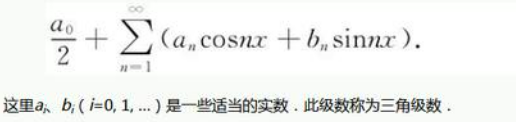
如何归一化：



1. **什么是傅里叶变换，为什么需要用傅里叶变换**

什么是傅里叶变换？

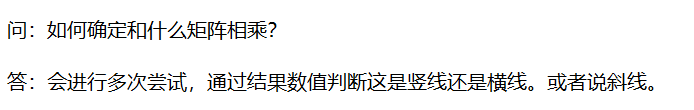
在有限闭区间上由任意图形定义的任何函数都可以分解为单纯的正弦和余弦函数的和。特别定义在（-π, π）上的函数总可表示为



类似于泰勒展开式，可以分解函数；傅里叶变换是用三角函数分解图形。

为什么需要用傅里叶变换？

上次分享提到了：



当时说的就是特征提取工程。

这样做有没有问题

没问题，只是效率低，需要一个个提取特征。人类通过进化固定了一些特征。但对于应对不同情况的机器，需要更高效率的自动获取特征。

方法：傅里叶变换。

另外工程上，为了效率，使用的是快速傅里叶变换。

1. **机器学习框架**

框架的作用是什么？

类似java框架，做用就是提高开发效率，底层逻辑，比如傅里叶变换，我不需要手写，只需要传入参数，机器自动取实现。（这也是机器学习俗称调参侠的由来）

有哪些框架？

Tensorflow、 Caffe、 Pytorch、MXNet、CNTK、 Paddlepaddle、Darknet

都有哪些特点？

Tensorflow是谷歌开发的

Pytorch是Facebook开发的。

Caffe是由伯克利人工智能研究小组和伯克利视觉和学习中心开发

其他的用的少不提了。

Tensorflow更繁琐，Pytorch更简单。

而且根据调查，之前Tensorflow更热门，最近这两年都在用Pytorch。

而且Pytorch支持Java语言。

推荐用哪个？

Pytorch

金融机构更多会考虑安全问题，Java因为熟悉，所以安全度高。能支持Java自然最好。之前也在某次会议上听到公司这边高层是希望有Java开发的机器学习。

不过我们开发还是用python实现，理由上次提到了：

1. python开发效率更高。
2. Python我们只需要学习简单的if、循环、集合语句，重点是框架的使用而不是语言，学习成本很低，几小时能学完够用的python语法。

时间成本<时间花费，所以选择python前期开发。

1. Python写好之后，迁移到其他语言，成本很低。开发语言不代表运行语言
2. **硬件和软件安装**

硬件要求：Nvidia显卡

首先机器学习一定要显卡吗？

可以不用，用显卡是为了效率。大数据的并行处理。所以应该用

其次显卡一定要Nvidia的吗？

可以不用，但是现在主流框架都是支持英伟达的CUDA运算平台，软件生态的问题，所以用英伟达的显卡更适合。

安装流程（必须有Nvidia显卡）

1. 安装Anaconda

<https://www.anaconda.com/products/distribution/download-success-2>

通过conda -v确定是否安装成功

如果环境变量没配置

请在安装成功后配置环境变量

<https://blog.csdn.net/yinjun3215/article/details/123705879>

1. 安装CUDA

<https://developer.nvidia.com/cuda-downloads?target_os=Windows&target_arch=x86_64&target_version=10&target_type=exe_local>

通过命令Nvcc -V 确定是否安装成功

1. 安装pytouch

<https://developer.nvidia.com/cuda-downloads?target_os=Windows&target_arch=x86_64&target_version=10&target_type=exe_local>

conda install pytorch torchvision torchaudio cudatoolkit=11.3 -c pytorch

(根据自己的CUDA版本安装）

1. **先写个CNN（卷积神经网络）做个图像识别练练手**

参考文件：

myCnn.py

我们可以总览一下这个代码。

首先我们可以看到Pytorch平台是提供了一些数据让我们可以实验。可以直接下载到本地。

其次，代码中并没有复杂的公式，没有我所说的快速傅里叶，没有复杂的python语法。我们只是设置参数，这是因为复杂的地方框架已经替我们做好了，我们要做的是理解背后的原理，然后调用对应的API，设置好参数。（如同炼丹）

1. **Pytorch函数简单介绍**

sequential函数：

如果你不用这个函数，你要实例化每个神经元，然后依次传参，如果用这个函数的话，啥都不用管了，直接把神经元的的名字和参数再这里面写好，然后这个函数会自动的把你写的所有神经元依次连接起来，就是代替你的工作了，不需要你去逐个连接了

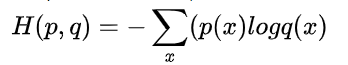
Adam函数：

Pytorch深度学习优化器的其中一种，其他内置算法包括SGD、RMSprop等。

作用：无需人工实现随机梯度下降算法，直接调用即可。

CrossEntropyLoss函数

Pytorch交叉熵损失函数



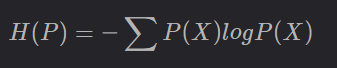
什么是交叉熵：

用于多分类的损失函数，熵越大模型越不确定，熵越小模型越确定，即优化模型目的是最小化交叉熵

交叉熵是从信息熵变换出来的。有兴趣的建议了解一下**香农公式**（非常重要的公式）

。

香农公式



表示表示的是随机变量的不确定性，某个事件发生的概率越小，其信息熵越大。（代表信息量很大）

举例子：今天我早上吃饭了。因为吃早饭是高概率事件，所以信息熵很小。

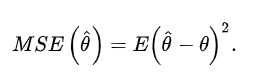
但是，我今天早上被走在路上被雷劈中，因为被雷劈中是低概率事件，信息熵很大，所以信息量很大。（听到这个信息立马激发别人的好奇心理）

所以小说一般都是开篇发生低概率事件，这样会导致信息量巨大，而在机器学习中，我们通过香农公式或者其变形的公式来评估决策。万变不离其宗，所以要先了解香农公式。

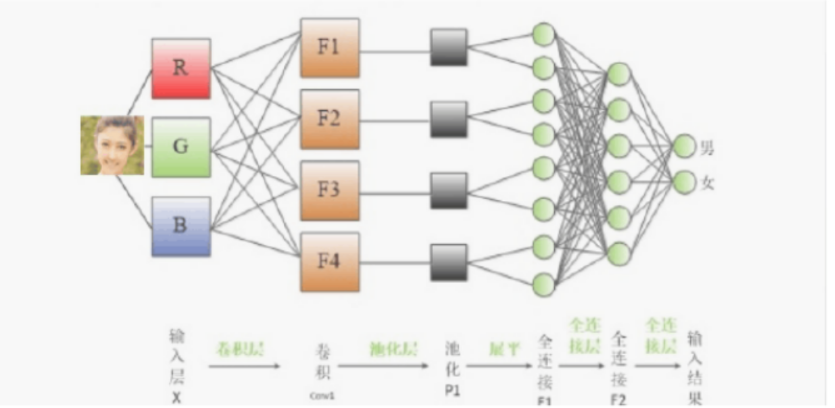
什么是损失函数

评估预测和实际的差距。

例如均方误差



1. **卷积神经网络详细介绍**



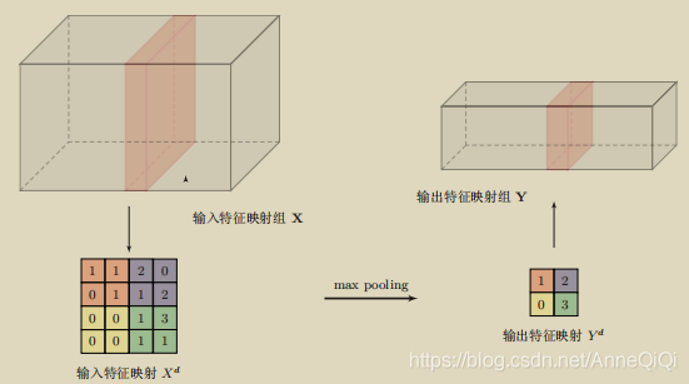
步骤：卷积—》池化—》卷积—》池化…—》全连接—》softmax

我们可以从刚刚的代码中sequential函数中看到我们设置了卷积核池化等参数。

卷积我上次分享说了原理了，就是通过矩阵变换来识别。

池化的作用是减少冗余信息，比如四个数字，取最大值，这样的方法。我们的CNN代码中使用的是最大值。

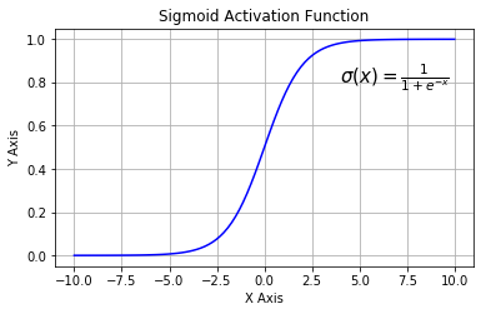
一般通过简单的最大值、最小值或平均值操作完成



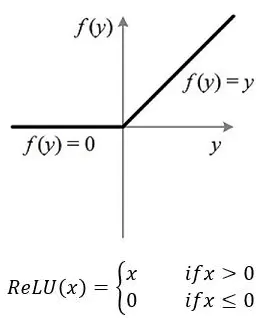
多次卷积和池化后，我们可以最终通过全连接层和输出层，把矩阵变换为一维数据，最终通过一维数据堵笔，得出结论，比如是男是女。

激活函数：

输入到输出的映射例如Sigmoid



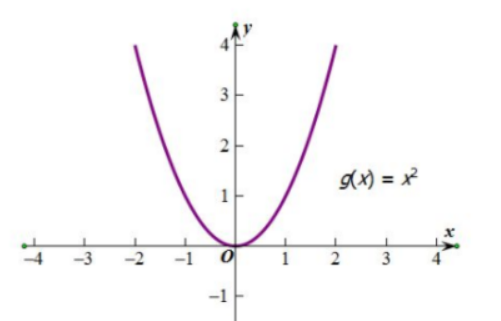
Demo中用的是ReLU激活函数



上一节优化器提到的梯度下降：

作用：梯度下降就是用来求某个函数最小值时自变量对应取值

举个例子，如果想求最低点在哪，我们可以通过不断地减少x步长，对比y的结果，得出最低点x和y大概在哪。



机器学习正常步骤：

训练-测试

通过训练让机器识别能力越来越强。

但是为什么需要测试：为了防止过拟合

什么是过拟合：杯弓蛇影，草木皆兵。把无关的东西也识别出来了，所以需要拿一部分训练之外的数据进行测试。防止过拟合。

为什么有了测试就可以防止过拟合？全部数据用于训练就不行。

因为如果全部数据都去训练了，数据会变得符合现有的数据，机器学习终究要去用于实际生产，实际生产都是未知数据，所以拿一部分数据作为未知数据，在训练好之后测试模型，判断是不是预测过度。

1. **下一步该做什么**

我们的目标是通过GCN实现反洗钱系统。

所以下一步是

1.继续深入了解原理，缺乏基础的建议了解和机器学习相关的线性代数、概率论；并阅读《机器学习实战》等相关书籍

2.实现图卷积神经网络开发Demo

3.试图获取数据，训练和测试Demo