Part2 数学基础

22.给定两个矩阵，怎么计算它们之间的相乘？ 怎么计算一个矩阵 inverse?

A = [[1,3],[2,5]], B[[4,1],[2,4]]。A\*B=? A^-1=?

解答: 矩阵相乘：各个位置元素相乘。A\*B=[[1\*4,3\*1],[2\*2,5\*4]]

2X2矩阵求逆：假设A=[[a,b],[c,d]],调换a和d的位置，把负号放在b和c前面，然后各个元素除以A矩阵的行列式 （ad-bc）

A^-1=-1\*[[5,-3],[-2,1]]=[[-5,3],[2,-1]]

23.怎么计算一个向量的 norm? a = (3,1,5,1)， |a| = ?

一范式：是向量a的一范式，即向量中绝对值之和 ||a||1 =10

二范式：向量元素绝对值的平方和再开方 ||a||2 =6

1. 什么是 Frobenius norm？ 给定 A = [[1,3],[2,5]], 请计算

解答: F-范数:矩阵A各项元素的绝对值平方和开方。A的F-范数为6.244997998398398

Note：https://www.cnblogs.com/fengff/p/9885765.html 向量和矩阵是不一样的

25.什么叫矩阵的 determinant? 怎么计算⼀个矩阵的 determinant? A = [[1,3],[2,5]].

det(A)=?

解答: 矩阵行列式是指矩阵的全部元素构成的行列式，设A=(aij)是数域P上的一个n阶矩阵，则所有A=(aij)中的元素组成的行列式称为矩阵A的行列式，记为|A|或det(A)。

2X2矩阵的行列式det(A)= ad-bc= -1

26.什么叫 underflow, 什么叫 overflow? 对于很多的 AI 问题，如果出现很多概率的

相乘，我们通常都在最前⾯加 log， 为什么？ ⽐如 argmax p(x), 通常求解

argmax log p(x).

解答:

(1)实数在计算机内用二进制表示，所以不是一个精确值，当数值过小的时候，被四舍五入为0，这就是下溢出。此时如果对这个数再做某些运算（例如除以它）就会出问题。反之，当数值过大的时候，被视为（正负）无穷，情况就变成了上溢出。

(2)加log是为了将乘除运算转变为加减运算，简化运算。

另：防止计算中出现上下溢出的方法：<https://www.codelast.com/%E5%8E%9F%E5%88%9B-%E5%A6%82%E4%BD%95%E9%98%B2%E6%AD%A2softmax%E5%87%BD%E6%95%B0%E4%B8%8A%E6%BA%A2%E5%87%BAoverflow%E5%92%8C%E4%B8%8B%E6%BA%A2%E5%87%BAunderflow/>

27.什么叫信息熵？ 什么叫互信息？ 他们具体的含义是什么？

解答:

(1)信息熵：在信息论中，熵是接收的每条消息中包含的信息的平均量，又被称为信息熵、信源熵、平均自信息量。消息代表来自分布或数据流中的事件、样本或特征。在信息世界，熵可以理解为不确定性的量度，熵越高，则能传输越多的信息，不确定性高，那么正确估计其值的可能性就越小；熵越低，则意味着传输的信息越少，不确定性低，正确估计其值的可能性就越大。

参考链接：<https://blog.csdn.net/pipisorry/article/details/51695283>



(2)互信息：在概率论和信息论中，两个随机变量的互信息（Mutual Information，简称MI）或转移信息（transinformation）是变量间相互依赖性的量度。不同于相关系数，互信息并不局限于实值随机变量，它更加一般且决定着联合分布p(X,Y) 和分解的边缘分布的乘积p(X)p(Y) 的相似程度。

参考链接：<https://www.cnblogs.com/gatherstars/p/6004075.html>





(3)交叉熵:是用来高衡量在给定的真实分布下，使用非真实分布指定的策略消除系统的不确定性所需要付出努力的大小



28. 对于softmax 函数，我们去实现的时候怎么避免underflow 或者 overflow?

答案：

『1』什么是下溢出（underflow）和上溢出（overflow）

实数在计算机内用二进制表示，所以不是一个精确值，当数值过小的时候，被四舍五入为0，这就是下溢出。此时如果对这个数再做某些运算（例如除以它）就会出问题。

反之，当数值过大的时候，情况就变成了上溢出。

『2』softmax函数是什么







30. 利利⽤用chain rule来分解联合概率 p(x1, x2, x3, x4, x5) =

p(x1,x2,x3,x4,x5)=p(x5|x1,x2,x3,x4)p(x1,x2,x3,x4)

p(x1,x2,x3,x4)=p(x4|x1,x2,x3)p(x1,x2,x3)

p(x1,x2,x3)=p(x3|x1,x2)p(x1,x2)

p(x1,x2)=p(x2|x1)p(x1)

所以

p(x1,x2,x3,x4,x5)= p(x5|x1,x2,x3,x4) p(x4|x1,x2,x3) p(x3|x1,x2) p(x2|x1) p(x1)

31. 什什么是KL-Divergence, 写出它的数学细节，什什么时候需要⽤用到它？

(1)Kullback-Leibler Divergence，它衡量的是相同事件空间里的两个概率分布的差异情况。

(2)数学公式：





**(3)用途**：比较两个概率分布的接近程度。在统计应用中，我们经常需要用一个简单的，近似的概率分布 f∗f∗ 来描述

观察数据 DD 或者另一个复杂的概率分布 ff 。这个时候，我们需要一个量来衡量我们选择的近似分布 f∗f∗ 相比原分布 ff 究竟损失了多少信息量，这就是KL散度起作用的地方。

39. Linear Programming 的常⻅见的应⽤用场景有哪些？请说出⾄至少 2 种可以⽤用来解决 LP问题的算法名称。

定义：目标函数和限制条件都是线性函数的都算线性规划

(1)应用

线性规划是运筹学中研究较早、发展较快、应用广泛、方法较成熟的一个重要分支,它是辅助人们进行科学管理的一种数学方法。在经济管理、交通运输、工农业生产等经济活动中，提高经济效果是人们不可缺少的要求，而提高经济效果一般通过两种途径：一是技术方面的改进，例如改善生产工艺，使用新设备和新型原材料。二是生产组织与计划的改进，即合理安排人力物力资源.线性规划所研究的是：在一定条件下，合理安排人力物力等资源，使经济效果达到最好。

(2) 解法：

求解线性规划问题的基本方法是单纯形法，现在已有单纯形法的标准软件，可在电子计算机上求解约束条件和决策变量数达 10000个以上的线性规划问题。为了提高解题速度，又有改进单纯形法、对偶单纯形法、原始对偶方法、分解算法和各种多项式时间算法。对于只有两个变量的简单的线性规划问题，也可采用图解法求解。这种方法仅适用于只有两个变量的线性规划问题。它的特点是直观而易于理解，但实用价值不大。通过图解法求解可以理解线性规划的一些基本概念。

40. 请使⽤用Linear Programming来求解Maximum Flow Problem，写出详细的过程。

设有向网络N（V，A），V顶点，A是边，其中，Vs是源， Vt 是汇。网络中有通过边aij从源到汇的流，但边流量不超过cij ，问该网络的最大流量为多少？ 这个问题，称为最大流问题。它是网络流中最经典的问题之一。最大流问题广泛地应用在交通运输、供水、油管供油、邮电通讯，也可以用在生产安排，管理优化等实际问题上。





Ford-Fulkerson算法

参考算法导论

KKT

<https://zhuanlan.zhihu.com/p/38163970>