

# Assignment #2: 深度学习与大语言模型

Updated 2204 GMT+8 Feb 25, 2025

2025 spring, Compiled by 王梓航, 物理学院

## 作业的各项评分细则及对应的得分

标准	等级	得分
按时提交	完全按时提交：1分 提交有请假说明：0.5分 未提交：0分	1 分
源码、耗时（可选）、解题思路（可选）	提交了4个或更多题目且包含所有必要信息：1分 提交了2个或以上题目但不足4个：0.5分 少于2个：0分	1 分
AC代码截图	提交了4个或更多题目且包含所有必要信息：1分 提交了2个或以上题目但不足4个：0.5分 少于：0分	1 分
清晰头像、PDF文件、MD/DOC附件	包含清晰的Canvas头像、PDF文件以及MD或DOC格式的附件：1分 缺少上述三项中的任意一项：0.5分 缺失两项或以上：0分	1 分
学习总结和个人收获	提交了学习总结和个人收获：1分 未提交学习总结或内容不详：0分	1 分
总得分： 5	总分满分： 5分	

### 说明：

#### 1. 解题与记录：

- 对于每一个题目，请提供其解题思路（可选），并附上使用Python或C++编写的源代码（确保已在OpenJudge，Codeforces，LeetCode等平台上获得Accepted）。请将这些信息连同显示“Accepted”的截图一起填写到下方的作业模板中。（推荐使用Typora <https://typoraio.cn> 进行编辑，当然你也可以选择Word。）无论题目是否已通过，请标明每个题目大致花费的时间。

#### 2. 课程平台与提交安排：

- 我们的课程网站位于Canvas平台（<https://pku.instructure.com>）。该平台将在第2周选课结束后正式启用。在平台启用前，请先完成作业并将作业妥善保存。待Canvas平台激活后，再上传你的作业。
  - 提交时，请首先上传PDF格式的文件，并将.md或.doc格式的文件作为附件上传至右侧的“作业评论”区。确保你的Canvas账户有一个清晰可见的头像，提交的文件为PDF格式，并且“作业评论”区包含上传的.md或.doc附件。

### 3. 延迟提交:

- 如果你预计无法在截止日期前提交作业，请提前告知具体原因。这有助于我们了解情况并可能为你提供适当的延期或其他帮助。

请按照上述指导认真准备和提交作业，以保证顺利完成课程要求。

## 1. 题目

### 18161: 矩阵运算

matrices, <http://cs101.openjudge.cn/practice/18161>

思路：正常按照要求求和即可

代码：

```
a = []
b = []
for _ in range(3):
    x,y=map(int,input().split())
    a+= [x,y]
    b.append([list(map(int,input().split())) for _ in range(x)])
if a[0]==a[4] and a[3]==a[5] and a[1]==a[2]:
    c = [[sum([b[0][i][j]*b[1][j][k] for j in range(a[1])])+b[2][i][k] for k in
range(a[5])] for i in range(a[0])]
    for index in c:
        print(*index)
else:
    print('Error!')
```

代码运行截图 (至少包含有"Accepted")

状态: Accepted

源代码

```
a = []
b = []
for _ in range(3):
    x,y=map(int,input().split())
    a+= [x,y]
    b.append([list(map(int,input().split())) for _ in range(x)])
if a[0]==a[4] and a[3]==a[5] and a[1]==a[2]:
    c = [[sum([b[0][i][j]*b[1][j][k] for j in range(a[1])])+b[2][i][k] :
```

基本信息

#: 48364645  
题目: 18161  
提交人: 24n2400011481  
内存: 4196kB  
时间: 83ms  
语言: Python3  
提交时间: 2025-02-26 13:52:50

# 19942: 二维矩阵上的卷积运算

matrices, <http://cs101.openjudge.cn/practice/19942/>

思路：正常按照要求处理即可

代码：

```
i1,j1,i2,j2=map(int,input().split())
import array
a = [array.array('b',list(map(int,input().split())) for i in range(i1)]
b = [array.array('b',list(map(int,input().split())) for i in range(i2)]
for i in range(i1-i2+1):
    g=[]
    for j in range(j1-j2+1):
        sum=0
        for ip,p in enumerate(range(i,i+i2)):
            for iq,q in enumerate(range(j,j+j2)):
                sum+=a[p][q]*b[ip][iq]
        g.append(sum)
    print(*g)
```

代码运行截图 (至少包含有"Accepted")

状态: Accepted

源代码

```
i1,j1,i2,j2=map(int,input().split())
import array
a = [array.array('b',list(map(int,input().split())) for i in range(i1)]
b = [array.array('b',list(map(int,input().split())) for i in range(i2)]
for i in range(i1-i2+1):
    g=[]
    for j in range(j1-j2+1):
        sum=0
```

基本信息

#: 46419841

题目: 19942

提交人: 24n2400011481

内存: 3640kB

时间: 22ms

语言: Python3

提交时间: 2024-10-11 09:42:53

# 04140: 方程求解

牛顿迭代法, <http://cs101.openjudge.cn/practice/04140/>

请用牛顿迭代法实现。

因为大语言模型的训练过程中涉及到了梯度下降（或其变种，如SGD、Adam等），用于优化模型参数以最小化损失函数。两种方法都是通过迭代的方式逐步接近最优解。每一次迭代都基于当前点的局部信息调整参数，试图找到一个比当前点更优的新点。理解牛顿迭代法有助于深入理解基于梯度的优化算法的工作原理，特别是它们如何利用导数信息进行决策。

牛顿迭代法

- **目的：**主要用于寻找一个函数  $f(x)$  的根，即找到满足  $f(x) = 0$  的  $x$  值。不过，通过适当变换目标函数，它也可以用于寻找函数的极值。
- **方法基础：**利用泰勒级数的一阶和二阶项来近似目标函数，在每次迭代中使用目标函数及其导数的信息来计算下一步的方向和步长。
- **迭代公式：** $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$  对于求极值问题，这可以转化为  $x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}$ ，这里  $f'(x)$  和  $f''(x)$  分别是目标函数的一阶导数和二阶导数。
- **特点：**牛顿法通常具有更快的收敛速度（尤其是对于二次可微函数），但是需要计算目标函数的二阶导数（Hessian矩阵在多维情况下），并且对初始点的选择较为敏感。

### 梯度下降法

- **目的：**直接用于寻找函数的最小值（也可以通过取负寻找最大值），尤其在机器学习领域应用广泛。
- **方法基础：**仅依赖于目标函数的一阶导数信息（即梯度），沿着梯度的反方向移动以达到减少函数值的目的。
- **迭代公式：** $x_{n+1} = x_n - \alpha \cdot \nabla f(x_n)$  这里  $\alpha$  是学习率， $\nabla f(x_n)$  表示目标函数在  $x_n$  点的梯度。
- **特点：**梯度下降不需要计算复杂的二阶导数，因此在高维空间中相对容易实现。然而，它的收敛速度通常较慢，特别是当目标函数的等高线呈现出椭圆而非圆形时（即存在条件数大的情况）。

### 相同与不同

- **相同点：**两者都可用于优化问题，试图找到函数的极小值点；都需要目标函数至少一阶可导。
- **不同点：**
  - 牛顿法使用了更多的局部信息（即二阶导数），因此理论上收敛速度更快，但在实际应用中可能会遇到计算成本高、难以处理大规模数据集等问题。
  - 梯度下降则更为简单，易于实现，特别是在高维空间中，但由于只使用了一阶导数信息，其收敛速度可能较慢，尤其是在接近极值点时。

代码：

```
x = 4
while True:
    y = x - (x**3 - 5*x**2 + 10*x - 80) / (3*x**2 - 10*x + 10)
    if abs(y - x) <= 10**(-10):
        print('{:.9f}'.format(y))
        exit()
    else:
        x = y
```

代码运行截图 (至少包含有"Accepted")

状态: Accepted

源代码

```
x = 4
while True:
    y = x-(x**3-5*x**2+10*x-80)/(3*x**2-10*x+10)
    if abs(y-x)<=10**(-10):
        print('{:.9f}'.format(y))
        exit()
    else:
        x = y
```

基本信息

#: 48364870  
题目: 04140  
提交人: 24n2400011481  
内存: 3932kB  
时间: 30ms  
语言: Python3  
提交时间: 2025-02-26 14:08:42

## 06640: 倒排索引

data structures, <http://cs101.openjudge.cn/practice/06640/>

思路: 正常收集处理即可

代码:

```
m = int(input())
a = {}
for i in range(1,m+1):
    b = input().split()[1:]
    for index in b:
        a.setdefault(index,set())
        a[index].add(i)
n = int(input())
c = set(a.keys())
for _ in range(n):
    d = input()
    if d in c:
        print(*sorted(list(a[d])))
    else:
        print('NOT FOUND')
```

代码运行截图 (至少包含有"Accepted")

状态: Accepted

源代码

```
m = int(input())
a = {}
for i in range(1,m+1):
    b = input().split()[1:]
    for index in b:
        a.setdefault(index,set())
        a[index].add(i)
```

基本信息

#: 48337619  
题目: 06640  
提交人: 24n2400011481  
内存: 11384kB  
时间: 86ms  
语言: Python3  
提交时间: 2025-02-23 10:32:50

## 04093: 倒排索引查询

data structures, <http://cs101.openjudge.cn/practice/04093/>

思路：用集合处理

代码：

```
n = int(input())
a = []
b = set()
for j in range(n):
    x = set(list(map(int, input().split()))[1:])
    a.append(x)
    b |= x
m = int(input())
def key1(x):
    return x[1]
for _ in range(m):
    res = b.copy()
    c = list(map(int, input().split()))
    c = sorted(list(enumerate(c)), key=key1, reverse=True)
    for j, index in c:
        if index == 1:
            res &= a[j]
        elif index == -1:
            res -= a[j]
    if res:
        print(*sorted(list(res)))
    else:
        print('NOT FOUND')
```

代码运行截图 (至少包含有"Accepted")

状态: Accepted

源代码

```
n = int(input())
a = []
b = set()
for j in range(n):
    x = set(list(map(int, input().split()))[1:])
    a.append(x)
    b |= x
m = int(input())
def key1(x):
    return x[1]
for _ in range(m):
    res = b.copy()
    c = list(map(int, input().split()))
    c = sorted(list(enumerate(c)), key=key1, reverse=True)
    for j, index in c:
        if index == 1:
            res &= a[j]
        elif index == -1:
            res -= a[j]
    if res:
        print(*sorted(list(res)))
    else:
        print('NOT FOUND')
```

基本信息

#: 48345768  
题目: 04093  
提交人: 24n2400011481  
内存: 8892kB  
时间: 59ms  
语言: Python3  
提交时间: 2025-02-24 09:29:09

## Q6. Neural Network实现鸢尾花卉数据分类

在<http://clab.pku.edu.cn> 云端虚拟机，用Neural Network实现鸢尾花卉数据分类。

参考链接，[https://github.com/GMyhf/2025spring-cs201/blob/main/LLM/iris\\_neural\\_network.md](https://github.com/GMyhf/2025spring-cs201/blob/main/LLM/iris_neural_network.md)

```
[rocky@younameone ~]$ python 1.py
Epoch [10/100], Loss: 0.2850
Epoch [20/100], Loss: 0.1172
Epoch [30/100], Loss: 0.0734
Epoch [40/100], Loss: 0.0625
Epoch [50/100], Loss: 0.0521
Epoch [60/100], Loss: 0.0465
Epoch [70/100], Loss: 0.0452
Epoch [80/100], Loss: 0.0422
Epoch [90/100], Loss: 0.0387
Epoch [100/100], Loss: 0.0388
Test Accuracy: 96.67%

Sample prediction: True class 0, Predicted class 0
[rocky@younameone ~]$ python 1.py
Epoch [10/100], Loss: 0.2388
Epoch [20/100], Loss: 0.0873
Epoch [30/100], Loss: 0.0602
Epoch [40/100], Loss: 0.0513
Epoch [50/100], Loss: 0.0459
Epoch [60/100], Loss: 0.0417
Epoch [70/100], Loss: 0.0400
Epoch [80/100], Loss: 0.0381
Epoch [90/100], Loss: 0.0357
Epoch [100/100], Loss: 0.0306
Test Accuracy: 93.33%
```

用云虚拟机跑出来的结果，但是还是没有完全的理解。

## 2. 学习总结和个人收获

如果发现作业题目相对简单，有否寻找额外的练习题目，如“数算2025spring每日选做”、LeetCode、Codeforces、洛谷等网站上的题目。

对于大模型的训练有了些基本的感受。也看了老师提供的youtube视频，有了些大致的了解。每日选做一直在跟进。