

## Assignment 2<sup>1</sup>

1.考虑如下文本集合:

- My dog has flea problems, please help.
- Maybe not take him to park.
- My dog is so cute and I love it.
- Stop posting stupid garbage.
- mr licks ate my steak.
- how to stop him.
- quit buying worthless dog food, stupid.

将所有大写字母变成小写，将每条文本去除停用词后进行分词(以空格为自然分隔符)，并计算每个词的tf-idf值(注意：英文的停用词表需要到网上搜索下载才能使用，文本库就是上述文本集合)。

---

<sup>1</sup>Due date: 6:30pm 4/21/2025

2.对 *Lecture 4: Keyword Extraction* 讲义41页的矩阵  $A$  进行如下两种截断奇异值分解:

- 截断列数  $k = 2$ ;
- 截断列数  $k = 1$ .

哪一种截断奇异值分解的近似效果更好?

3.编写一个Python自定义函数计算两个相同长度向量的余弦相似度。

4. 考虑 *Lecture 6: Artificial Neural Network* 讲义24页所示含双隐藏层的神经网络模型:

(i) 将预测值  $(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_k)$  写成关于输入  $(x_1, x_2, \dots, x_p)$  的函数形式.

(ii) 如果每个神经元之间的连接均带有偏置参数(bias), 计算该神经网络模型所包含的未知参数个数.

(iii) 假设输出层只有一个神经元( $K = 1$ ), 计算平方损失函数分别对两个隐藏层偏置参数的梯度.

5.证明*Lecture 6: Artificial Neural Network*中关于激活函数的几个结论:

(i)对S型函数

$$\Gamma(z) = \frac{e^z}{1 + e^z},$$

证明 $\Gamma'(z) = \Gamma(z)(1 - \Gamma(z))$ .

(ii)对双曲正切函数

$$\tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}},$$

证明 $\tanh(z) = 2\Gamma(2z) - 1$ .

6.考虑 *Lecture 6: Artificial Neural Network* 讲义57页所示神经网络:

(i)求解梯度  $\frac{\partial L}{\partial w_2}, \frac{\partial L}{\partial w_3}, \frac{\partial L}{\partial w_4}$ .

(ii)求解经过一次梯度下降迭代后损失函数  $L$  的值, 与迭代前  $L$  的值相比有何变化?