CS620 HW5
@averm004@odu.edu

CS 620–Introduction to Data Science and Analytics, HW5

1) The following table summarizes the performance of a classification algorithm for identifying cancerous (Malignant) cells.

a. (5 pts) Calculate the accuracy of the classifier.

b. (10 pts) Calculate the precision and recall of the classifier.

c. (10 pts) Calculate the F1 measure of the classifier.

d. (10 pts) Calculate the Fβ measure of the classifier in which weights recall twice as much as

precision. Note: Fβ https://en.wikipedia.org/wiki/F-score

Classifier		Actual	
		Malignant	Benign
Prediction	Malignant	60	15
	Benign	5	35

Solution 1: Give the above confusion matrix since the positions of Actual and Predictions are inter changed the positions of TP and TN remains same but positions of FN and FP are interchanged.

TP = 60

FP = 15

FN = 5

TN = 35

Accuracy = (TP + TN) / (TP + FP + FN + TN) = 60+35/60+15+5+25 = 95/115 =**0.826**

Precision = TP / (TP + FP) = 60/60+15 = 60/75=**0.8**

Recall = TP / (TP + FN)=60/60+5 = 60/65=**0.92**

F1_Measure = 2 * (precision * recall) / (precision + recall) = 2*(0.8*0.92)/0.8+0.92 = 1.472/1.72 = **0.857**

CS620 HW5
@averm004@odu.edu

`F_beta` = (1 + beta**2) * (precision * recall) / ((beta**2 * precision) + recall)

Two commonly used values of beta are 2 and 0.5

beta_2 = 2
beta_0_5 = 0.5

F_beta_2_score = (1 + beta_2**2) * (precision * recall) / ((beta_2**2 * precision) + recall) = **0.895**

F__beta_0_5_score = (1 + beta_0_5**2) * (precision * recall) / ((beta_0_5**2 * precision) + recall) =**0.821**

2) Consider the 2 ranking algorithms in the figure below.

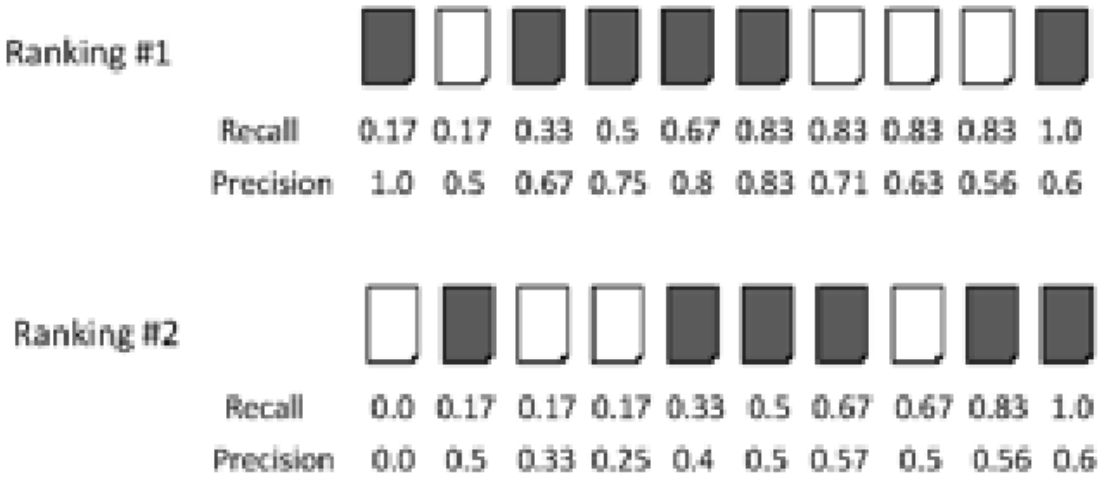
a. (10 pts) Calculate the confusion matrix values (TP, FP, TN, FN) for position 6 (1st position is in the leftmost) in each ranking method.

b. (10 pts) Using the confusion matrix calculated in part a.), compute the Accuracy and Harmonic Mean (F1 measure) at position 6 for both ranking methods.

c. (10 pts) Calculate the Mean Average Precision (MAP) for both ranking methods (for all the retrieved documents).

d. (15 pts) Interpolation defines precision at any recall level as the maximum precision observed in any recall-precision point at a higher recall level. Calculate the Interpolated Precision for each standard recall value (0.0,0.1….1.0) and generate the Recall-Precision graph for Ranking algorithms #1 and #2.

CS620 HW5
@averm004@odu.edu



Solution 2:

Number of relevant documents = 6

Ranking #1 Confusion Matrix:

	Relevant	Nonrelevant
Retrieved	tp = 5	fp = 1
Not retrieved	fn = 1	tn = 3

Ranking #2 Confusion Matrix:

	Relevant	Nonrelevant
Retrieved	tp = 3	fp = 3
Not retrieved	fn = 3	tn = 1

CS620 HW5
@averm004@odu.edu

Ranking 1 Accuracy and F1 Measure

Accuracy = (5+3)/(5+1+1+3)= **0.8**

F Measure =

2∗(0.83∗0.83)(0.83+0.83)2∗(0.83∗0.83)(0.83+0.83) = **0.83**

Ranking2 Accuracy and F1 Measure

Accuracy =(3+1)/(3+3+3+1) = **0.4**

F Measure = 2∗(0.5∗0.5)(0.5+0.5)2∗(0.5∗0.5)(0.5+0.5) = **0.5**

Ranking 1 MAP = (1.0+0.67+0.75+0.8+0.83+0.6)/6 = **0.775**

Ranking 2 MAP = (0.4+0.5+0.5+0.57+0.56+0.6)/6 =**0.521**

MAP for both ranking = (0.775+0.521)/2 = **0.648**

Interpolation

Recall # 1	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Interpolation of #1	1.0	1.0	0.83	0.83	0.83	0.83	0.83	0.83	0.6	0.6

Recall # 2	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Interpolation of #2	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6

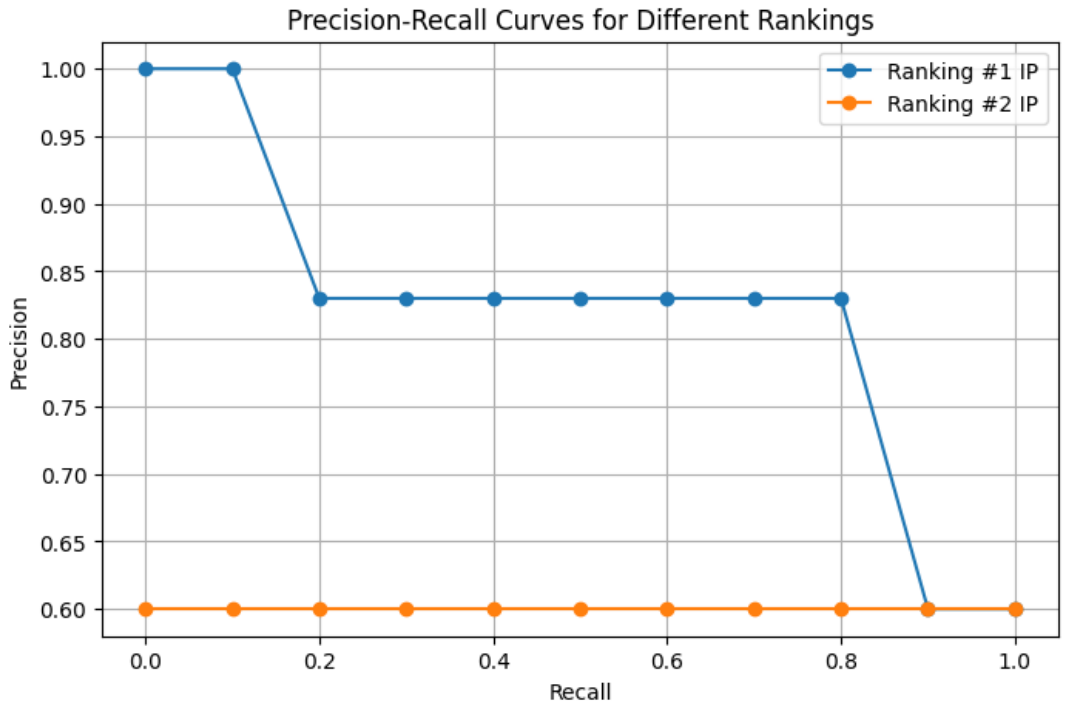
CS620 HW5
@averm004@odu.edu

```python
import matplotlib.pyplot as plt

# Provided recall and precision values
recall_values = [0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]
precision_values_ranking_1 = [1.0, 1.0, 0.83, 0.83, 0.83, 0.83, 0.83, 0.83, 0.83, 0.6, 0.6]
precision_values_ranking_2 = [0.6, 0.6, 0.6, 0.6, 0.6, 0.6, 0.6, 0.6, 0.6, 0.6, 0.6]

# Plotting
plt.figure(figsize=(8, 5))
plt.plot(recall_values, precision_values_ranking_1, marker='o', label='Ranking #1 IP')
plt.plot(recall_values, precision_values_ranking_2, marker='o', label='Ranking #2 IP')

# Add labels and title
plt.xlabel('Recall')
plt.ylabel('Precision')
plt.title('Precision-Recall Curves for Different Rankings')
plt.legend()
plt.grid(True)
plt.show()
```



+ Code + Markdown

CS620 HW5
@averm004@odu.edu

3)Describe your thoughts about what you think it means to work as a data scientist. You may therefore – if you like – be very personal and describe your plans and fears for your future career, criticism (or appreciation) for your education, and skills you need to develop further, and soon. This question is intended to encourage you to reflect on yourself and your future career and will therefore be graded generously!

Solution 3: As an extensive experience in data engineering background, I wanted dig one step deeper into data but due to limitations of opportunity in at my work station, I always wanted to move to data science stream, to look more closely at every aspect of data, get deep insights from it and make meaningful inference which can be used for betterment of society, environment, medical, financial and other aspects of day-to-day life.

I am always fascinated by the fact that they way our brain works similar way we are able to mold machine learning and deep neural networks. The challenging part is to understand various flavors of machine and deep learning algorithm. There 4 major deep learning giants we have namely Tensorflow,Pytorch, Hugging face and OpenNN. Learning one framework will not always solve problem as different industry used different deep learning library as per the use case. I have done data project in both Tensorflow and Pytorch hopefully will learn more deeply Tensorflow as its used in current industry during the holiday seasons.

I always wanted to peruse PhD in Machine Learning inspired by **Sebastian Raschka** but due to personal commitment there might be some delay to start PhD path.

After I have completed Masters, my goal is to get a job either in Data science or Machine Learning field and ODU is really helping in setting the mind in right directions.

This master is dedicated to my parents whom I lost during Covid in 2021.

Due to this I wanted to utilize my deep learning techniques in medical research programs to help people to have longer life span and better living.

CS620 HW5
@averm004@odu.edu