CSCI-GA.2560-001, Artificial Intelligence April 18, 2022

Solutions to Problem 1 of Homework 10 (5 Points)

Name: Anav Prasad (ap7152) *Due: 5PM on Monday, April 18*

Collaborators:

1. [5 pts.] Suppose that a classifier computes a numeric score to an item based on the classifier's "confidence" that the item is a member of the target category. In using the classifier, you set a threshold, accept the items whose score is higher than the threshold, and reject items whose score is lower. For instance, suppose you have the following training set and outputs for a binary classifier:

Name	a	b	c	d	e	f	g	h	i	j
Label	T	T	F	T	F	T	T	F	F	F
Score	.95	.92	.85	.84	.81	.75	.71	.69	.62	.56

Name	k	l	m	n	o	p	q	r	s	t
Label	F	F	T	F	T	F	F	F	T	F
Score	.51	.48	.43	.42	.32	.25	.21	.15	.08	.01

If you then set the threshold at 0.50, the classifier will accept items a-k and reject items l-t. However only the 'T' items should have been accepted.

Compute precision, recall, and F-score for the following thresholds: 0.9, 0.8, 0.7, 0.4, 0.2.

Solution:
For each threshold:

- **threshold = 0.9**:

$$\begin{aligned}\text{True Positives (TP)} &= 2\\ \text{False Positives (FP)} &= 0\\ \text{True Negatives (TN)} &= 12\\ \text{False Negatives (FN)} &= 6\end{aligned}$$

Therefore,

$$\begin{aligned}\text{Precision} &= \frac{TP}{TP+FP}\\ &= \frac{2}{2+0}\\ \therefore \textbf{Precision} &= \textbf{1.0}\end{aligned}$$

$$\begin{aligned}\text{Recall} &= \frac{TP}{TP+FN} \\ &= \frac{2}{2+6} \\ \therefore \textbf{Recall} &= \frac{1}{4} = \textbf{0.25}\end{aligned}$$

Thus,

$$\begin{aligned}\text{F-Score} &= \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \\ &= \frac{2 \cdot 1.0 \cdot (1/4)}{1.0 + (1/4)} \\ &= \frac{1/2}{5/4} \\ \therefore \textbf{F-Score} &= \frac{2}{5} = \textbf{0.4}\end{aligned}$$

- **threshold = 0.8**:

$$\begin{aligned}\text{True Positives (TP)} &= 3 \\ \text{False Positives (FP)} &= 2 \\ \text{True Negatives (TN)} &= 10 \\ \text{False Negatives (FN)} &= 5\end{aligned}$$

Therefore,

$$\begin{aligned}\text{Precision} &= \frac{TP}{TP+FP} \\ &= \frac{3}{3+2} \\ \therefore \textbf{Precision} &= \frac{3}{5} = \textbf{0.6} \\ \text{Recall} &= \frac{TP}{TP+FN} \\ &= \frac{3}{3+5} \\ \therefore \textbf{Recall} &= \frac{3}{8} = \textbf{0.375}\end{aligned}$$

Thus,

$$\begin{aligned}\text{F-Score} &= \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \\ &= \frac{2 \cdot (3/5) \cdot (3/8)}{3/5 + 3/8} \\ &= \frac{9/20}{39/40} \\ \therefore \textbf{F-Score} &= \frac{6}{13} \approx \textbf{0.462}\end{aligned}$$

- **threshold = 0.7**:

$$\begin{aligned}\text{True Positives (TP)} &= 5\\ \text{False Positives (FP)} &= 2\\ \text{True Negatives (TN)} &= 10\\ \text{False Negatives (FN)} &= 3\end{aligned}$$

Therefore,

$$\begin{aligned}\text{Precision} &= \frac{TP}{TP+FP}\\ &= \frac{5}{5+2}\\ \therefore \textbf{Precision} &= \frac{5}{7} \approx \textbf{0.714}\\ \text{Recall} &= \frac{TP}{TP+FN}\\ &= \frac{5}{5+3}\\ \therefore \textbf{Recall} &= \frac{5}{8} = \textbf{0.625}\end{aligned}$$

Thus,

$$\begin{aligned}\text{F-Score} &= \frac{2 \cdot Precision \cdot Recall}{Precision + Recall}\\ &= \frac{2 \cdot (5/7) \cdot (5/8)}{5/7 + 5/8}\\ &= \frac{25/28}{75/56}\\ \therefore \textbf{F-Score} &= \frac{2}{3} \approx \textbf{0.667}\end{aligned}$$

- **threshold = 0.4**:

$$\begin{aligned}\text{True Positives (TP)} &= 6\\ \text{False Positives (FP)} &= 8\\ \text{True Negatives (TN)} &= 4\\ \text{False Negatives (FN)} &= 2\end{aligned}$$

Therefore,

$$\begin{aligned}\text{Precision} &= \frac{TP}{TP+FP}\\ &= \frac{6}{6+8}\\ \therefore \textbf{Precision} &= \frac{3}{7} \approx \textbf{0.429}\end{aligned}$$

$$\begin{aligned}\text{Recall} &= \frac{TP}{TP+FN} \\ &= \frac{6}{6+2} \\ \therefore \textbf{Recall} &= \frac{3}{4} = \textbf{0.75}\end{aligned}$$

Thus,

$$\begin{aligned}\text{F-Score} &= \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \\ &= \frac{2 \cdot (3/7) \cdot (3/4)}{3/7 + 3/4} \\ &= \frac{9/14}{33/28} \\ \therefore \textbf{F-Score} &= \frac{6}{11} \approx \textbf{0.545}\end{aligned}$$

- **threshold = 0.2**:

$$\begin{aligned}\text{True Positives (TP)} &= 7 \\ \text{False Positives (FP)} &= 10 \\ \text{True Negatives (TN)} &= 2 \\ \text{False Negatives (FN)} &= 1\end{aligned}$$

Therefore,

$$\begin{aligned}\text{Precision} &= \frac{TP}{TP+FP} \\ &= \frac{7}{7+10} \\ \therefore \textbf{Precision} &= \frac{7}{17} \approx \textbf{0.412} \\ \text{Recall} &= \frac{TP}{TP+FN} \\ &= \frac{7}{7+1} \\ \therefore \textbf{Recall} &= \frac{7}{8} = \textbf{0.875} \\ \text{F-Score} &= \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \\ &= \frac{2 \cdot (7/17) \cdot (7/8)}{7/17 + 7/8} \\ &= \frac{49/68}{175/136} \\ \therefore \textbf{F-Score} &= \frac{14}{25} = \textbf{0.56}\end{aligned}$$

□

CSCI-GA.2560-001, Artificial Intelligence April 18, 2022

Solutions to Problem 2 of Homework 10 (5 Points)

Name: Anav Prasad (ap7152) *Due: 5PM on Monday, April 18*

Collaborators:

2. [5 pts.] Using the following training set with k-nearest neighbors

A1	**A2**	**C**
3	3	A
4	5	B
2	2	A
10	9	A
14	12	B
17	19	B
8	13	B

Indicate the predicted label for each element of the following test set, also provide precision and recall.

A1	**A2**	**C**
3	6	A
20	5	B
6	6	A
19	9	A
17	19	B

Use Euclidean distance squared, unit votes, and $K = 3$.

Solution:

For precision and recall measurement, let A represent the "positive" classification and B represent the "negative" classification.

Classification of the test set elements:

A1	**A2**	**C**
3	6	A

  Computing the euclidean distance squared from this test element to every other element:

A1	A2	Distance	Vote
3	3	9	A
4	5	2	B
2	2	17	A
10	9	58	A
14	12	157	B
17	19	365	B
8	13	74	B

Counting the votes of the K nearest neighbors:

Class	Vote
A	2
B	1

Thus, the kNN classification for this element is A.

A1	A2	C
20	5	B

Computing the euclidean distance squared from this test element to every other element:

A1	A2	Distance	Vote
3	3	293	A
4	5	256	B
2	2	333	A
10	9	116	A
14	12	85	B
17	19	205	B
8	13	208	B

Counting the votes of the K nearest neighbors:

Class	Vote
A	1
B	2

Thus, the kNN classification for this element is B.

A1	A2	C
6	6	A

Computing the euclidean distance squared from this test element to every other element:

A1	A2	Distance	Vote
3	3	18	A
4	5	5	B
2	2	32	A
10	9	25	A
14	12	100	B
17	19	290	B
8	13	53	B

Counting the votes of the K nearest neighbors:

Class	Vote
A	2
B	1

Thus, the kNN classification for this element is A.

A1	A2	C
19	9	A

Computing the euclidean distance squared from this test element to every other element:

A1	A2	Distance	Vote
3	3	292	A
4	5	241	B
2	2	338	A
10	9	81	A
14	12	34	B
17	19	104	B
8	13	137	B

Counting the votes of the K nearest neighbors:

Class	Vote
A	1
B	2

Thus, the kNN classification for this element is B.

A1	A2	C
17	19	B

Computing the euclidean distance squared from this test element to every other element:

A1	A2	Distance	Vote
3	3	452	A
4	5	365	B
2	2	514	A
10	9	149	A
14	12	58	B
17	19	0	B
8	13	117	B

Counting the votes of the K nearest neighbors:

Class	Vote
A	0
B	3

Thus, the kNN classification for this element is B.

Therefore, the compilated results of kNN on the test set are as follows:

A1	A2	Given Class	Predicted Class
3	6	A	A
20	5	B	B
6	6	A	A
19	9	A	B
17	19	B	B

Therefore,

$$\begin{aligned}\text{True Positives (TP)} &= 2\\ \text{False Positives (FP)} &= 0\\ \text{True Negatives (TN)} &= 2\\ \text{False Negatives (FN)} &= 1\end{aligned}$$

Thus,

$$\begin{aligned}\text{Precision} &= \frac{TP}{TP+FP}\\ \therefore \textbf{Precision} &= \frac{2}{2+0} = \mathbf{1.0}\\ \text{Recall} &= \frac{TP}{TP+FN}\\ &= \frac{2}{2+1}\\ \therefore \textbf{Recall} &= \frac{2}{3} \approx \mathbf{0.667}\end{aligned}$$

□