HW 1

January 17, 2024

Pupipat Singkhorn

1 Homework 1 Clustering and Regression

1.1 Metrics

Model A	Predicted dog	Predicted cat
Actual dog	30	20
Actual cat	10	40

1.1.1 T1.

What is the accuracy of Model A?

[ ]:
```
tn = 30
fn = 10
fp = 20
tp = 40

accuracy = ( tp+tn ) / ( tp+tn+fp+fn )

print(f' Accuracy = {accuracy} ')
```

```
 Accuracy = 0.7
```

1.1.2 T2.

Consider cats as 'class 1' (positive) and dogs as 'class 0' (negative), calculate the precision, recall, and F1.

Model A	Predicted dog (-)	Predicted cat (+)
Actual dog	TN	FP
Actual cat	FN	TP

[ ]:
```
accuracy = ( tp+tn ) / ( tp+tn+fp+fn )
print(f' Accuracy = {accuracy} ')
```

```
Accuracy = 0.7
```

[ ]:
```
tn = 30
fn = 10
fp = 20
tp = 40

precision = tp / (tp+fp)
recall = tp / (tp+fn)
F1 = 2*tp / ( 2*tp + fp + fn )

print(f' Precision = {precision} ')
print(f' Recall = {recall} ')
print(f' F1 = {F1} ')
```

```
Precision = 0.6666666666666666
Recall = 0.8
F1 = 0.7272727272727273
```

1.1.3 T3.

Consider class cat as 'class 0' and class dog as 'class 1', calculate the precision, recall, and F1.

Model A	Predicted dog (+)	Predicted cat (-)
Actual dog	TP	FN
Actual cat	FP	TN

[ ]:
```
accuracy = ( tp+tn ) / ( tp+tn+fp+fn )

print(f' Accuracy = {accuracy} ')
```

```
Accuracy = 0.7
```

[ ]:
```
tp = 30
tn = 40
fp = 10
fn = 20

precision = tp / (tp+fp)
recall = tp / (tp+fn)
F1 = 2*tp / ( 2*tp + fp + fn )

print(f' Precision = {precision} ')
print(f' Recall = {recall} ')
print(f' F1 = {F1} ')
```

```
Precision = 0.75
Recall = 0.6
F1 = 0.6666666666666666
```

1.1.4 T4.

Now consider a lopsided population where there are 80% cats.
What is the accuracy of Model A?
Using dog as the positive class, what is the precision, recall, and F1?
Explain how and why these numbers change (or does not change) from the previous questions.

Model A	Predicted dog (+)	Predicted cat (-)
Actual dog	TP	FN
Actual cat	FP	TN

Model A	Predicted dog	Predicted cat
Actual dog	30	20
Actual cat	10	40

when 80% cats, Actual cat = 80 Actual dog = 20

Actual dog TP = 30 but from new constrain Actual dog(TP+FN) must not exceed 20
then, **TP will decrease**

From T2, T3 acccuracy does not change even data changes.
$Accuracy = 0.7$

$$tp + tn + fp + fn = 100$$
$$fp + tn = 80$$
$$0.7 = \frac{tp + tn}{tp + tn + fp + fn}$$

solve three eq. in form of tp

$$\therefore tn = 70 - tp$$
$$\therefore fp = tp + 10$$
$$\therefore fn = 20 - tp$$

$$precision = \frac{tp}{tp + fp} = \frac{tp}{2tp + 10} = \frac{1}{2 + 10/tp}$$

when tp decrease, **precision will decrease**

$$recall = \frac{tp}{tp + fn} = \frac{tp}{20}$$

when tp decrease, **recall will decrease**

$$F1 = \frac{2}{1/recall + 1/precision}$$

when *recall* and *precision* decrease, **F1 will decrease**

1.1.5 OT1.

Consider the equations for accuracy and F1. When will accuracy be equal, greater, or less than F1?

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn} = \frac{1}{1 + \frac{fp+fn}{tp+\mathbf{tn}}}$$

$$F1 = \frac{2tp}{2tp + fp + fn} = \frac{1}{1 + \frac{fp+fn}{tp+\mathbf{tp}}}$$

$$\therefore Accuracy = F1 \quad ; \; tn = tp$$

$$\therefore Accuracy > F1 \quad ; \; tn > tp$$

$$\therefore Accuracy < F1 \quad ; \; tn < tp$$