Performance Measures: Precision, Recall, and F1

Dirk Hovy

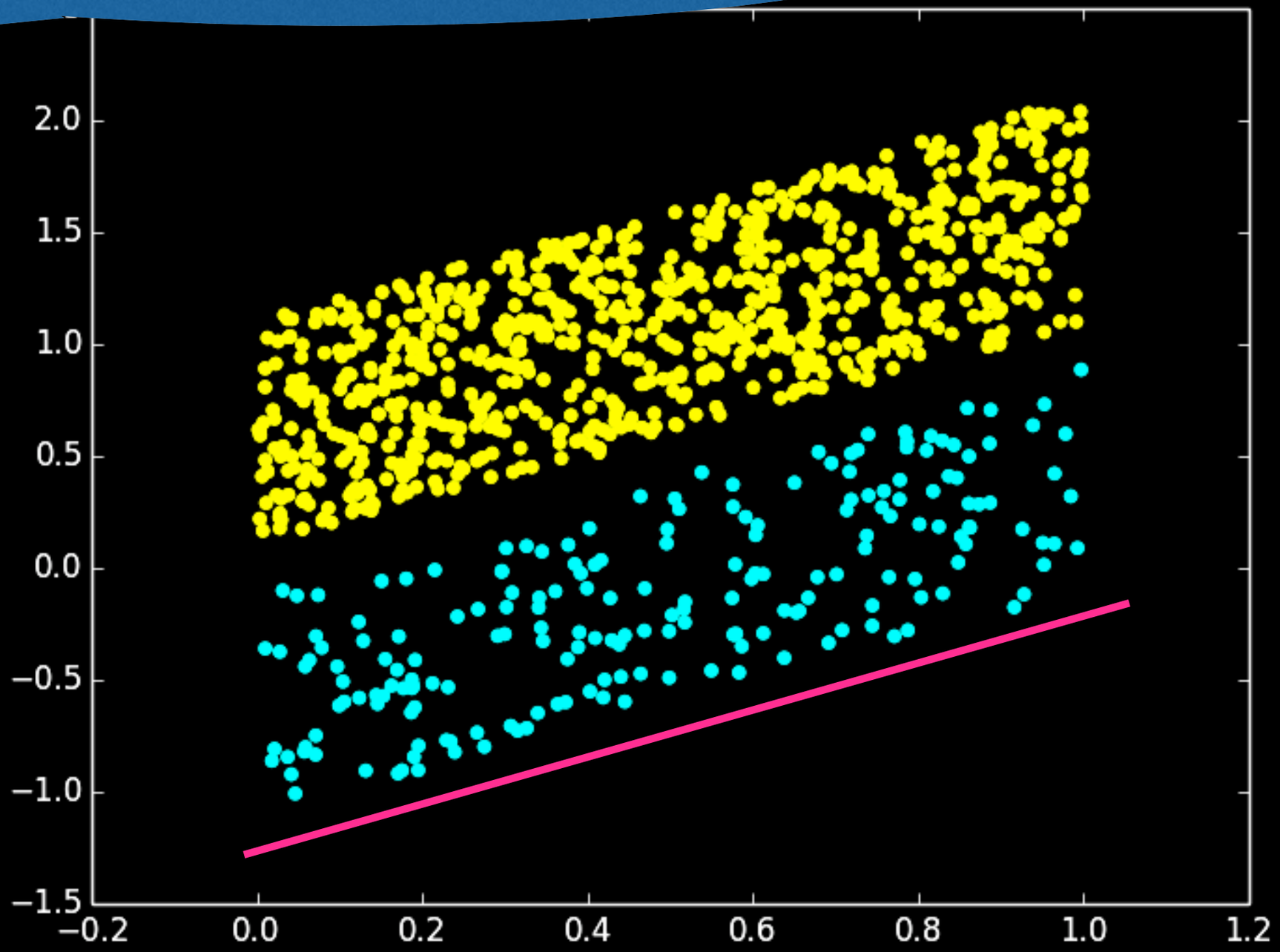
Center for Language Technology
University of Copenhagen

KØBENHAVNS UNIVERSITET

Performance Problems

I have a classifier that's 70% accurate!

x	y	ŷ
frog	1	1
deer	1	1
wolf	1	1
dog	1	1
bear	1	1
fish	1	1
bird	1	0
cat	1	0
stone	0	1
tree	0	0



		predicted	
		1	0
gold	1	TP	FN
	0	FP	TN

True and False

TARGET = LABEL 1

x	y	ŷ	
frog	1	1	
deer	1	1	
wolf	1	1	true positive
dog	1	1	
bear	1	1	
fish	1	1	
bird	1	0	false negative
cat	1	0	
stone	0	1	false positive
tree	0	0	true negative

accuracy = (TP+TN) / (P + N)

precision = TP / (TP + FP)

recall = TP / (TP + FN)

F1 = 2 (prec x rec) / (prec + rec)

accuracy = 7/10 = 0.7

precision = 6/7 = 0.86

recall = 6/8 = 0.75

F1 = 0.81

Changing Target

		predicted	
		1	0
gold	1	TP	FN
	0	FP	TN

TARGET = LABEL 0

x	y	ŷ	
frog	0	0	
deer	0	0	
wolf	0	0	
dog	0	0	true negative
bear	0	0	
fish	0	0	
bird	0	1	false positive
cat	0	1	
stone	1	0	false negative
tree	1	1	true positive

accuracy = (TP+TN) / (P + N)

precision = TP / (TP + FP)

recall = TP / (TP + FN)

F1 = 2 (prec x rec) / (prec + rec)

accuracy = 7/10 = 0.7

precision = 1/3 = 0.33

recall = 1/2 = 0.5

F1 = 0.4

KØBENHAVNS UNIVERSITET

		predicted	
		1	0
gold	1	TP	FN
	0	FP	TN

MICRO Averaging

WEIGH BY CLASS SIZE

accuracy = (TP+TN) / (P + N)

precision = TP / (TP + FP)

recall = TP / (TP + FN)

F1 = 2 (prec x rec) / (prec + rec)

TARGET=1 TARGET=0

x	**y**	**ŷ**	**x**	**y**	**ŷ**
frog	1	1	frog	0	0
deer	1	1	deer	0	0
wolf	1	1	wolf	0	0
dog	1	1	dog	0	0
bear	1	1	bear	0	0
fish	1	1	fish	0	0
bird	1	1	bird	0	0
cat	1	0	cat	0	1
stone	0	1	stone	1	0
tree	0	0	tree	1	1

acc = 7/10 + 7/10 = 14/20 = 0.7

prec = 6/7 + 1/3 = 7/10 = 0.7

rec = 6/8 + 1/2 = 7/10 = 0.7

F1 = 0.7

KØBENHAVNS UNIVERSITET

		predicted	
		1	0
gold	1	TP	FN
	0	FP	TN

MACRO Averaging

WEIGH ALL CLASSES EQUALLY

accuracy = (TP+TN) / (P + N)

precision = TP / (TP + FP)

recall = TP / (TP + FN)

F1 = 2 (prec x rec) / (prec + rec)

TARGET=1

x	**y**	**ŷ**
frog	1	1
deer	1	1
wolf	1	1
dog	1	1
bear	1	1
fish	1	1
bird	1	1
cat	1	0
stone	0	1
tree	0	0

TARGET=0

x	**y**	**ŷ**
frog	0	0
deer	0	0
wolf	0	0
dog	0	0
bear	0	0
fish	0	0
bird	0	0
cat	0	1
stone	1	0
tree	1	1

acc = (0.7 + 0.7) / 2 = 0.7

prec = (0.86 + 0.33) / 2 = 0.6

rec = (0.5 + 0.75) / 2 = 0.63

F1 = 0.61

		predicted	
		“1”	“0”
gold	1	TP	FN
	0	FP	TN

PREDICT MAJORITY CLASS FOR ALL

“Cheating”: Total Recall

TARGET = LABEL 1

x	y	ŷ	
frog	1	1	
deer	1	1	
wolf	1	1	
dog	1	1	true positive
bear	1	1	
fish	1	1	
bird	1	1	
cat	1	1	
stone	0	1	false positive
tree	0	1	

accuracy = (TP+TN) / (P + N)

precision = TP / (TP + FP)

recall = TP / (TP + FN)

F1 = 2 (prec x rec) / (prec + rec)

accuracy = 8/10 = 0.8

precision = 8/10 = 0.8

recall = 8/8 = 1.0

F1 = 0.9

Take-home points

- **accuracy** can be too general
- **precision** and **recall** are per-class measures
- **precision** = how many of instances labeled as target class are actually *in* target class?
- **recall** = how many of *all* target class instances in data identified correctly?
- **F1** = symmetric mean of precision and recall

KØBENHAVNS UNIVERSITET

Thanks!

KØBENHAVNS UNIVERSITET

Questions?

KØBENHAVNS UNIVERSITET