CMPS 460 – Spring 2022

MACHINE LEARNING

Tamer Elsayed

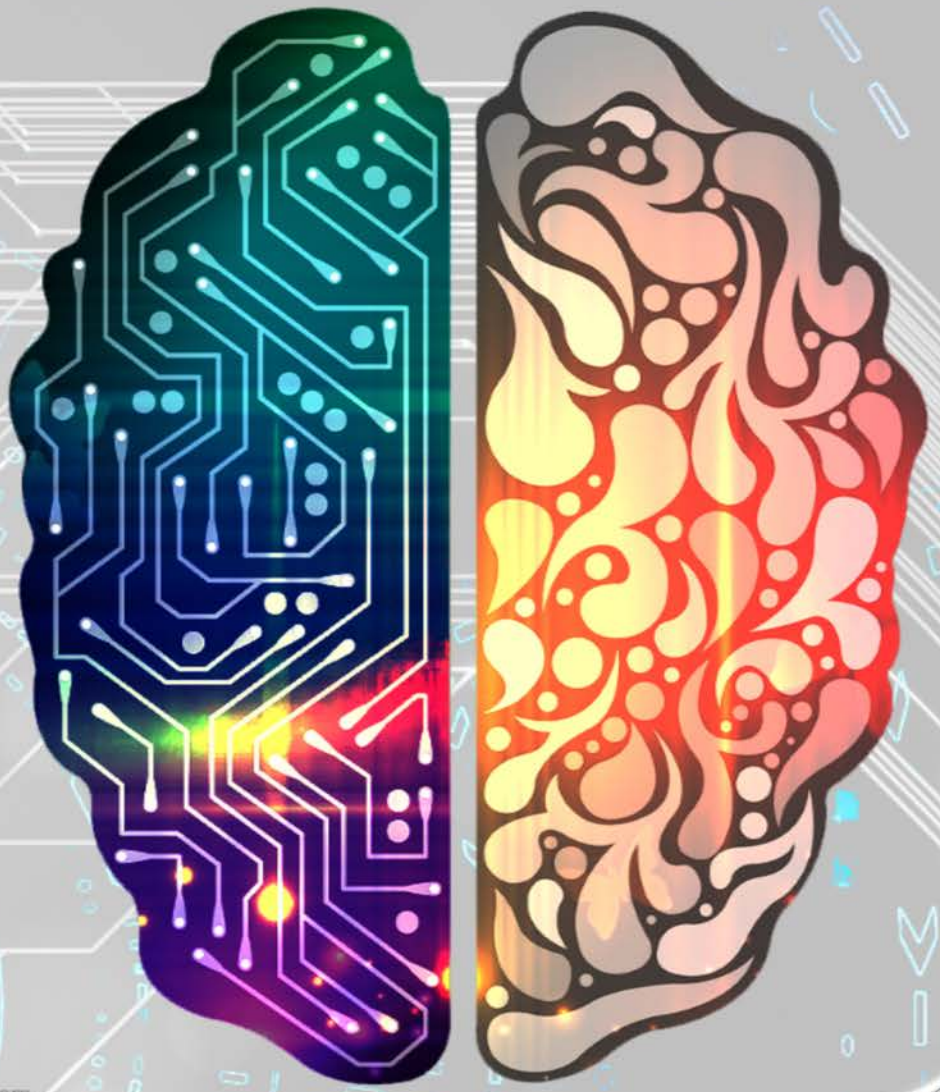


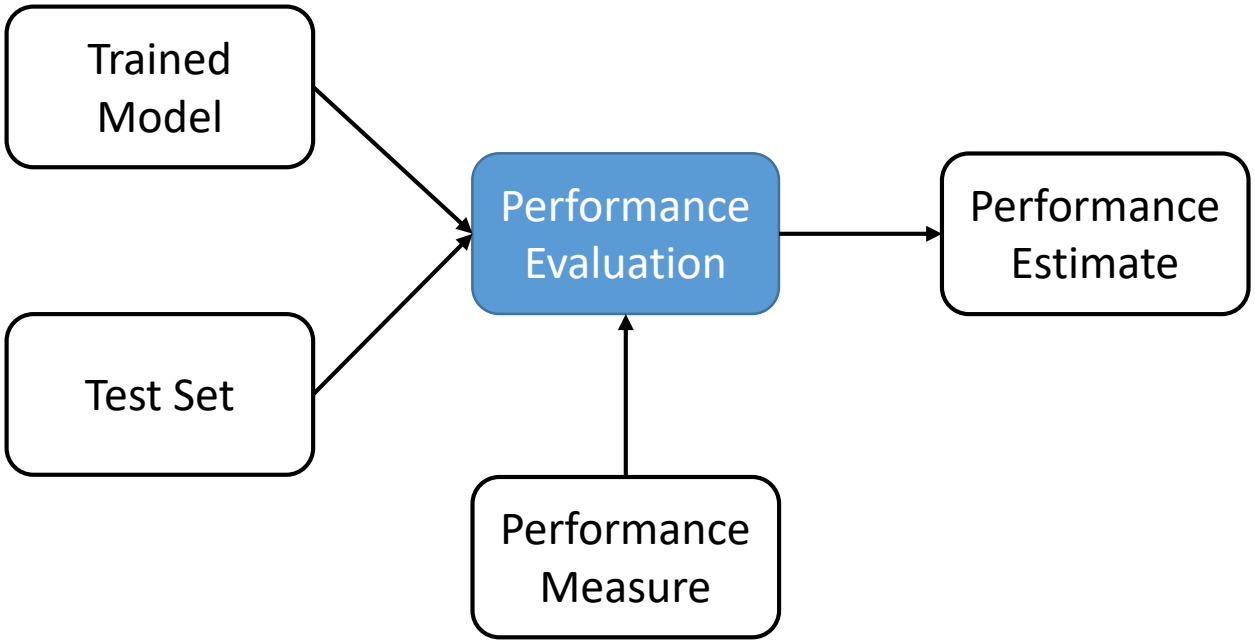
Image hosted by: WittySparks.com | Image source: Pixabay.com

5.b

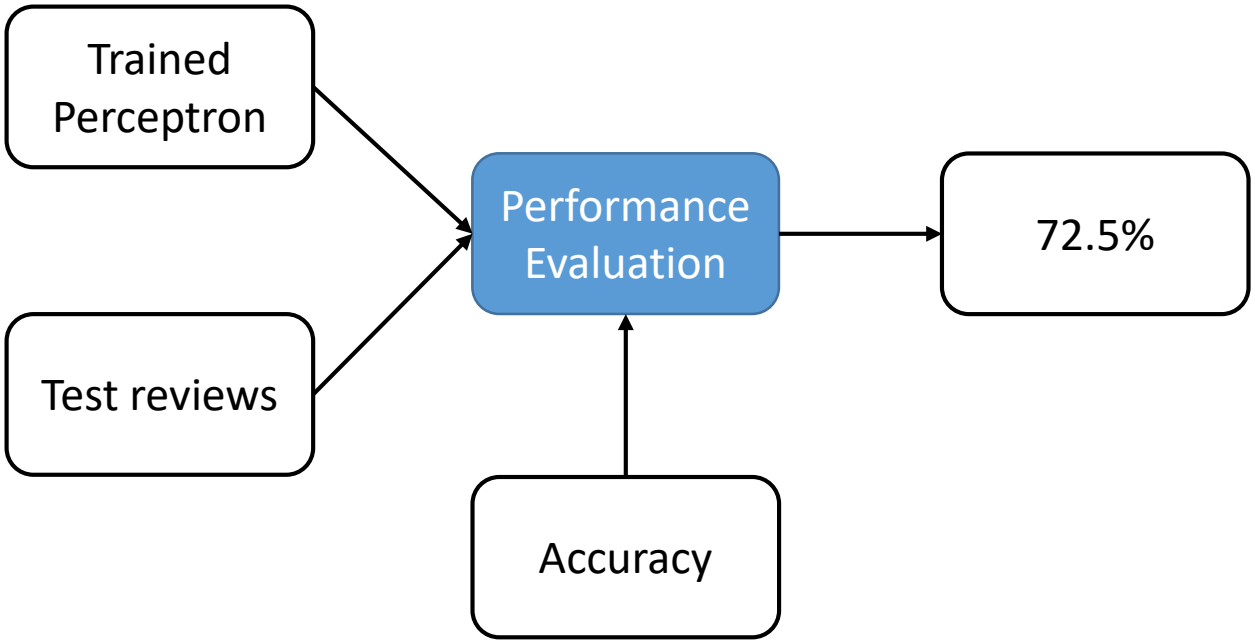
Practical Issues: Evaluation

Sec 5.5-5.6

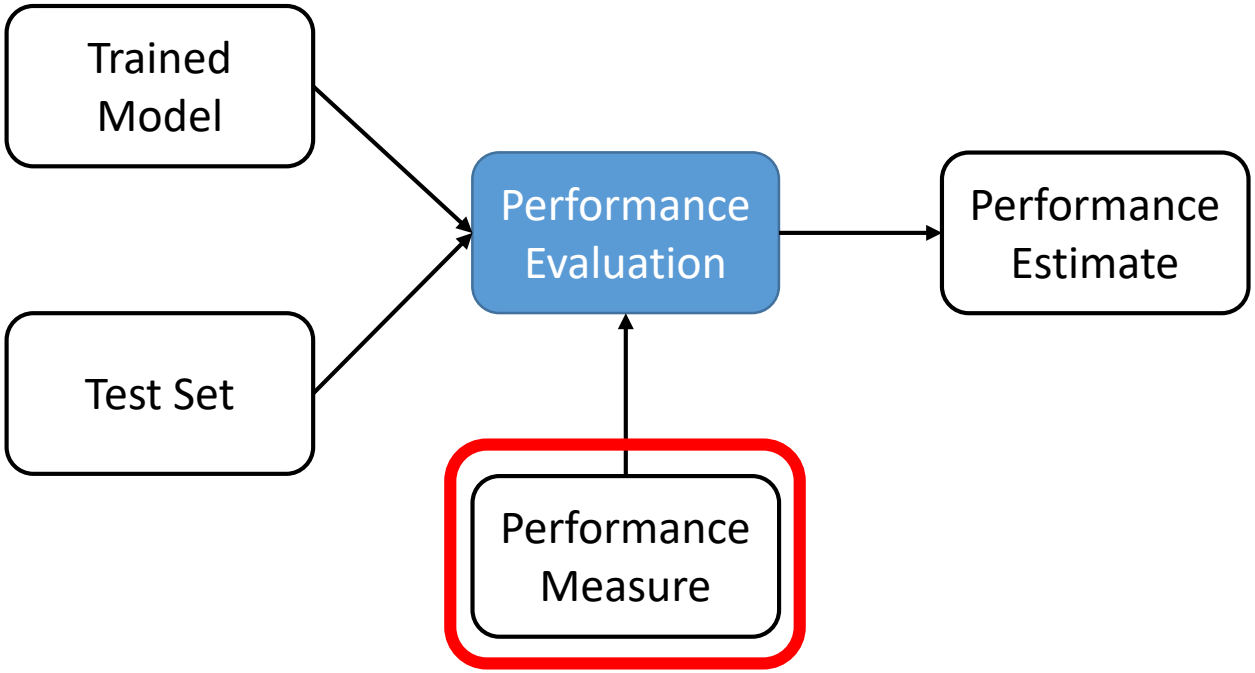
Performance Evaluation



Example


Trained Perceptron
Test reviews
Performance Evaluation
72.5%
Accuracy

Performance Evaluation

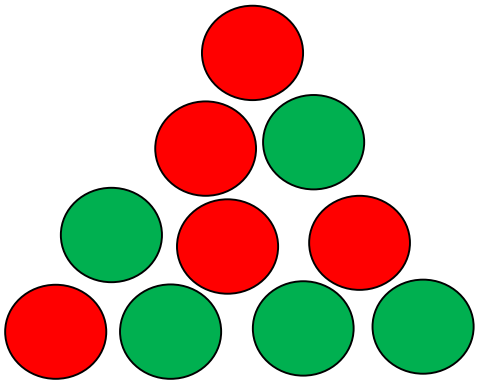


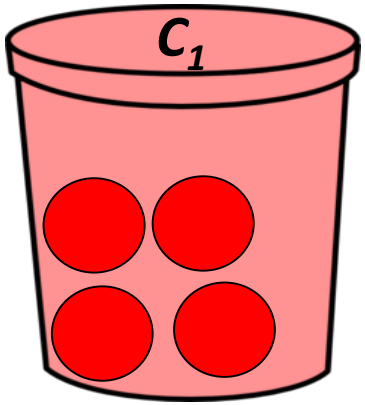
Accuracy

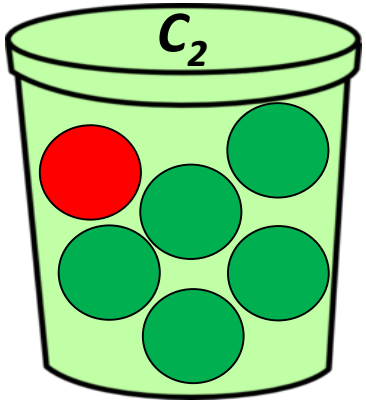
What fraction of the examples are classified correctly?

Acc = ?

= 9/10

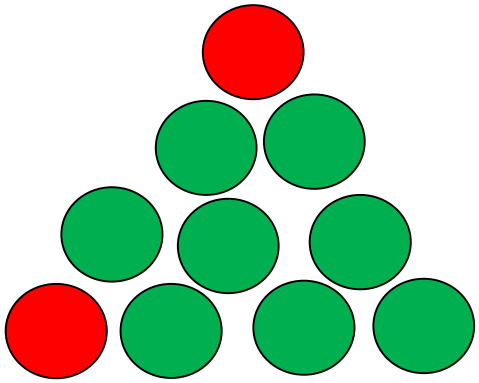


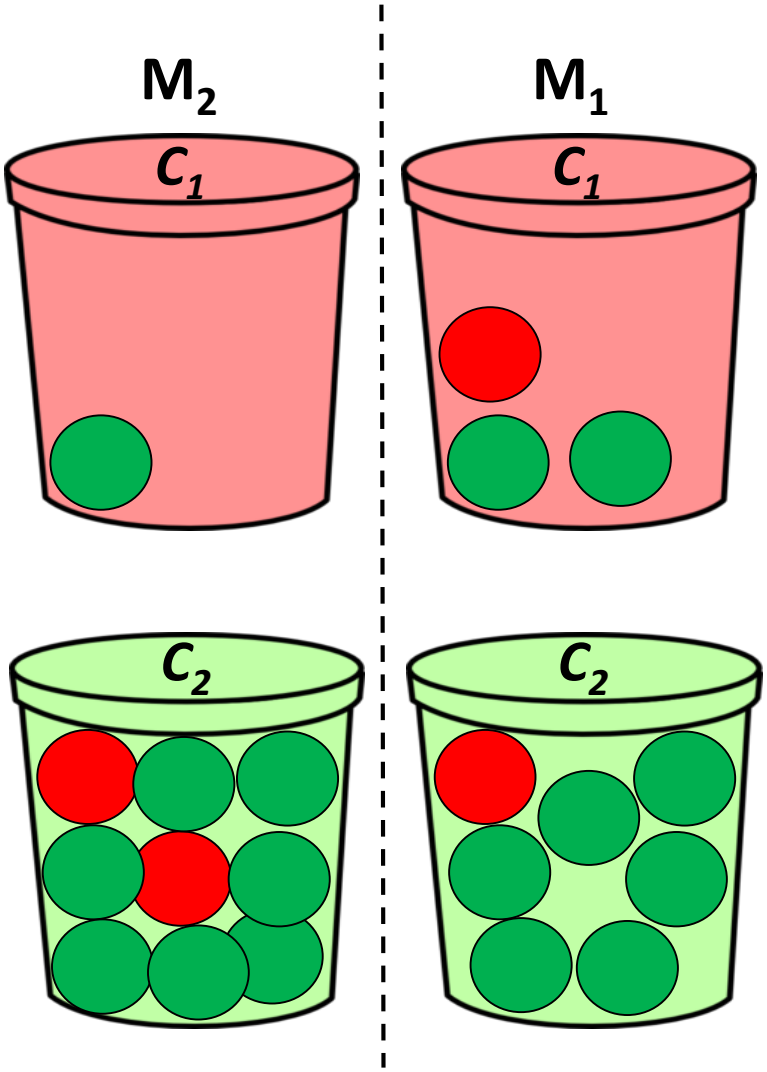




Accuracy

- Acc(M_1) = ?
- Acc(M_2) = ?





What's the problem?

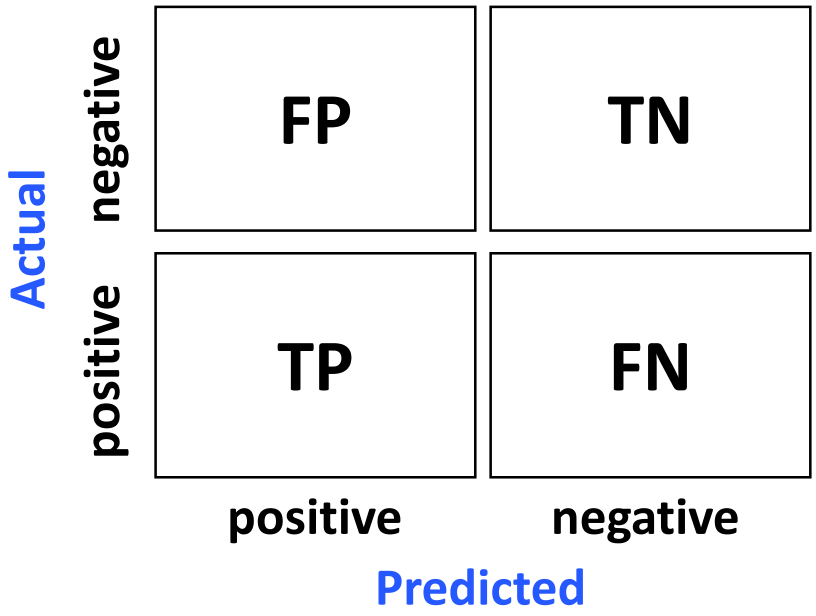
Problem with Accuracy?

- Imbalanced data (distribution of classes)!
- Some errors matter more than others …
  - Given medical record, predict patient has COVID or not
  - Given an email, detect spam

- When classes are highly unbalanced, we focus on one ***target class*** (usually the rare class), denoted as the "positive" class.

Precision/Recall/F1
for the target class (positive)

Confusion Matrix



Confusion Matrix

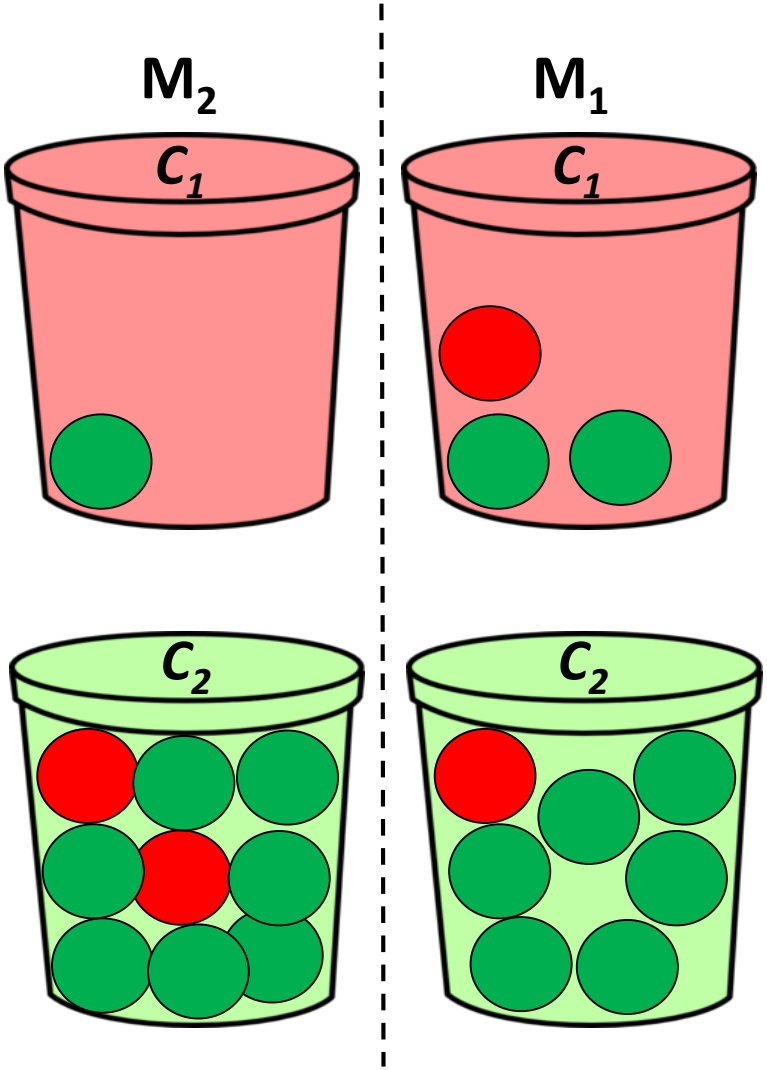
positive

M_1

Actual \ Predicted	positive	negative
negative	?	?
positive	?	?

M_2

Actual \ Predicted	positive	negative
negative	?	?
positive	?	?



Confusion Matrix

positive

M_1

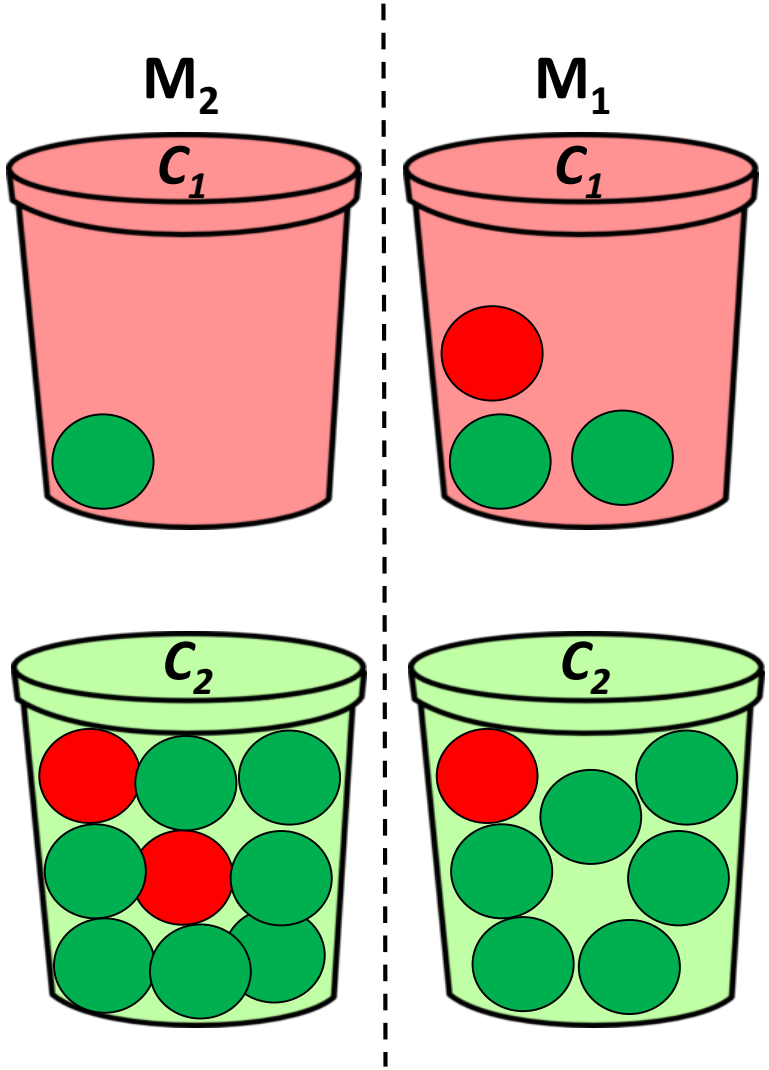
Actual	positive	negative
negative	2	6
positive	1	1

Predicted

M_2

Actual	positive	negative
negative	1	7
positive	0	2

Predicted



Precision

M_1

Actual	positive	negative
negative	2	6
positive	1	1

Predicted

M_2

Actual	positive	negative
negative	1	7
positive	0	2

Predicted

What fraction of those predicted as positive are actually positive?

$$P = \frac{TP}{TP + FP}$$

$$P(M_1) =$$

$$P(M_2) =$$

Precision: % of positive predictions that are correct

Recall

M_1

Actual \ Predicted	positive	negative
negative	2	6
positive	1	1

What fraction of the actual positive examples are predicted as positive?

$$R = \frac{TP}{TP + FN}$$

M_2

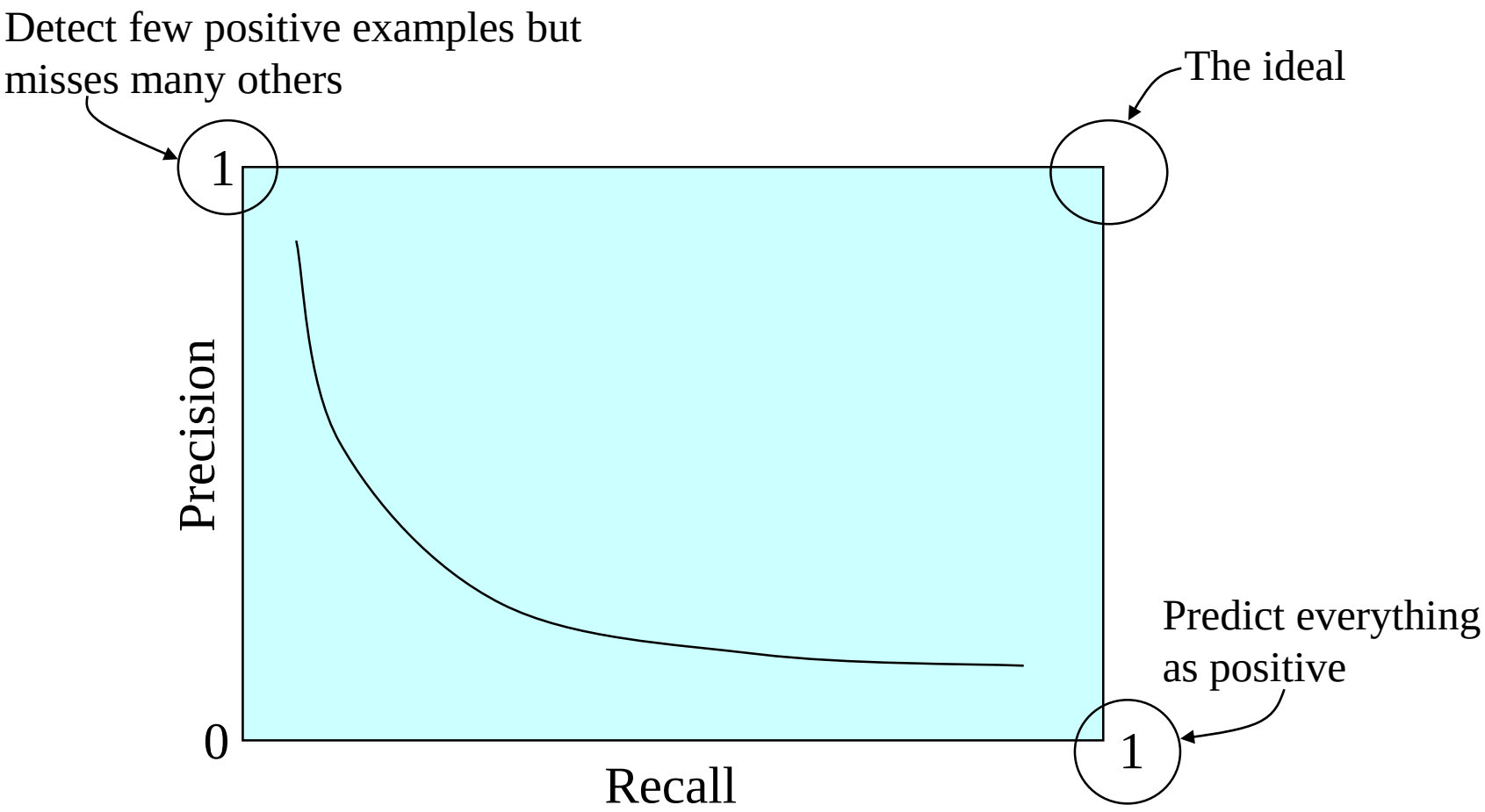
Actual \ Predicted	positive	negative
negative	1	7
positive	0	2

$$R(M_1) =$$

$$R(M_2) =$$

Recall: % of gold positive examples that are found

Trade-off between P & R



Example

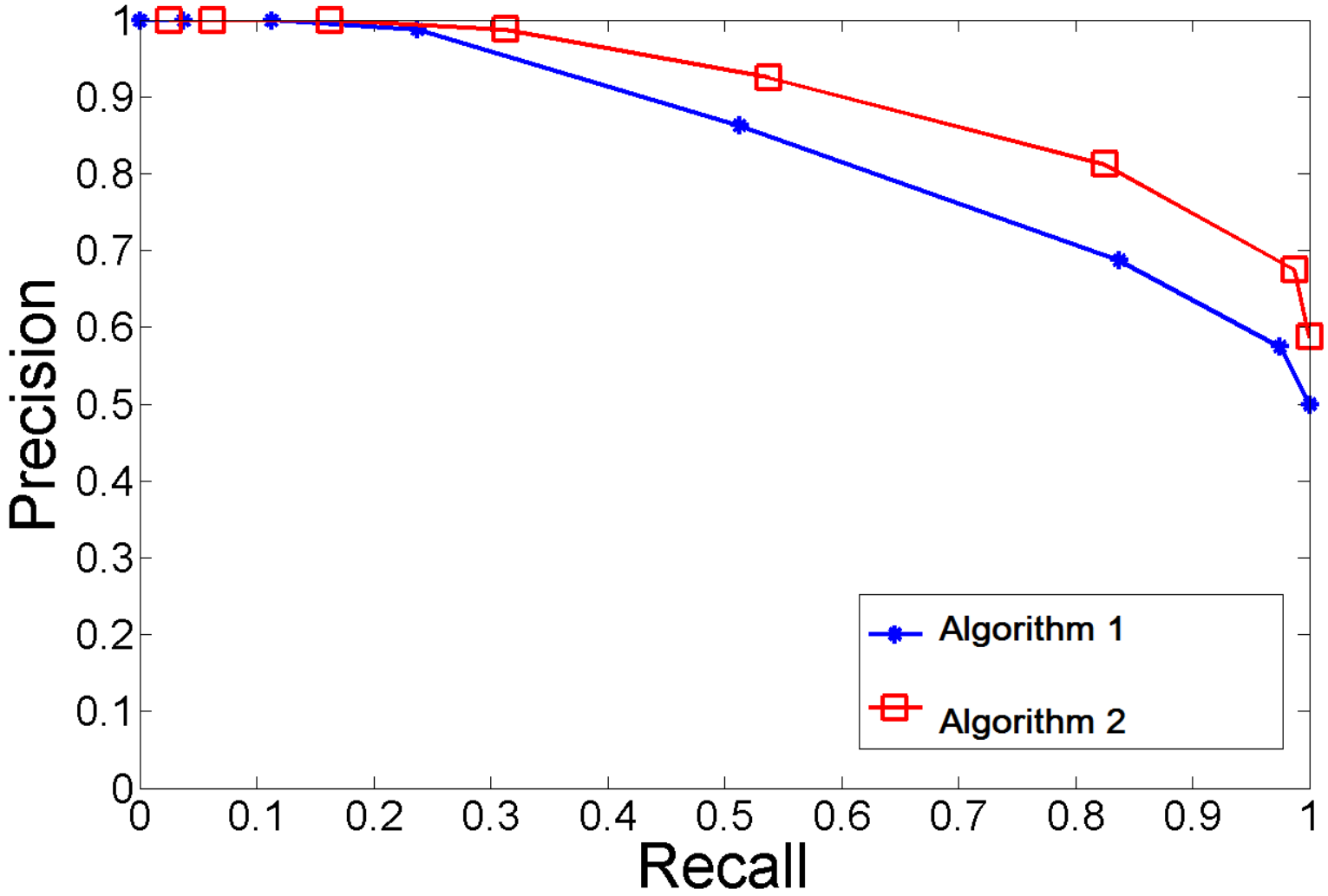
Test example	Spam score
email #25	0.94
email #37	0.89
email #2	0.76
email #15	0.73
email #116	0.61
email #64	0.54
email #7	0.42
.....	.....
.....	.....
.....	.....
email #38	0.24
email #10	0.16
email #25	0.13
email #168	0.02

P=0.67, R=0.2

P=0.5, R=0.3

P=0.34, R=0.9

Precision/Recall Curve



A Combined Measure: F-measure

- F_1 measure

$$F_1 = \frac{2 * P * R}{P + R}$$

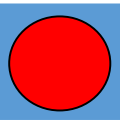
  - Harmonic mean of P and R

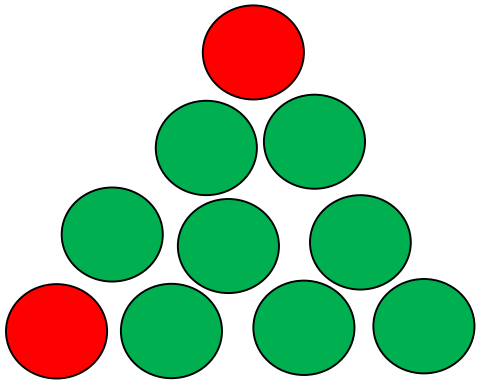
Why?

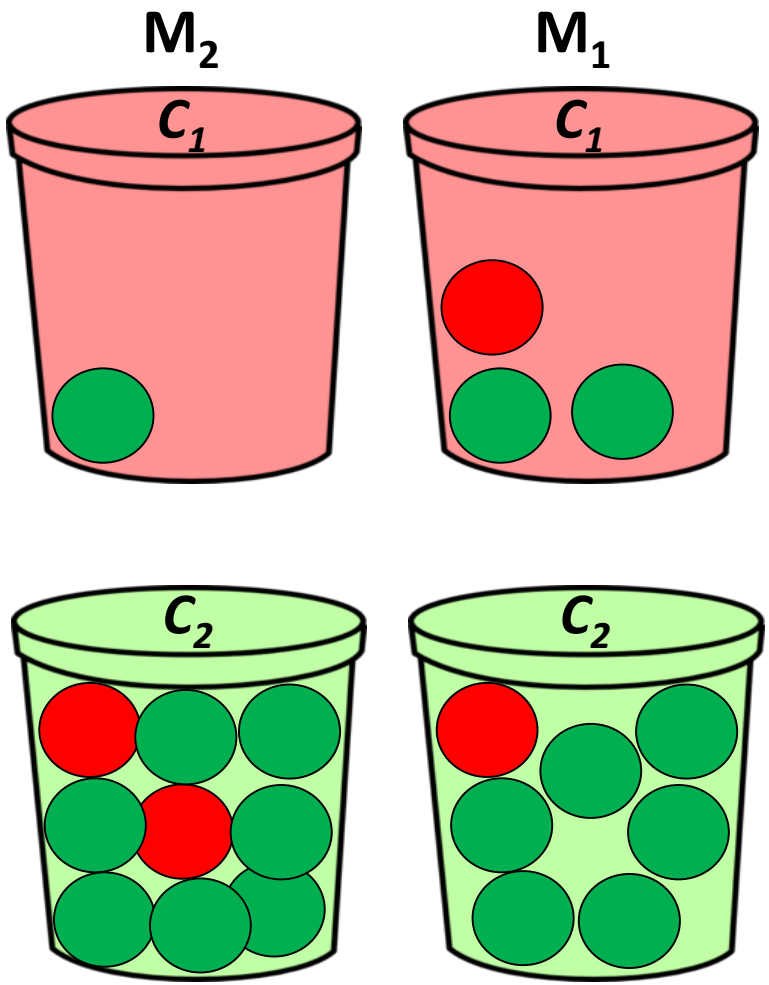
- Weighted F measure

$$F_\beta = \frac{(\beta^2+1) * P * R}{\beta^2 * P + R}$$

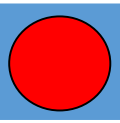
Binary Classification

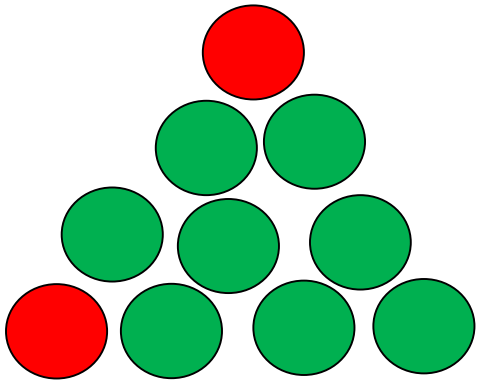
	M_1	M_2
Precision	?	?
Recall	?	?
F1	?	?

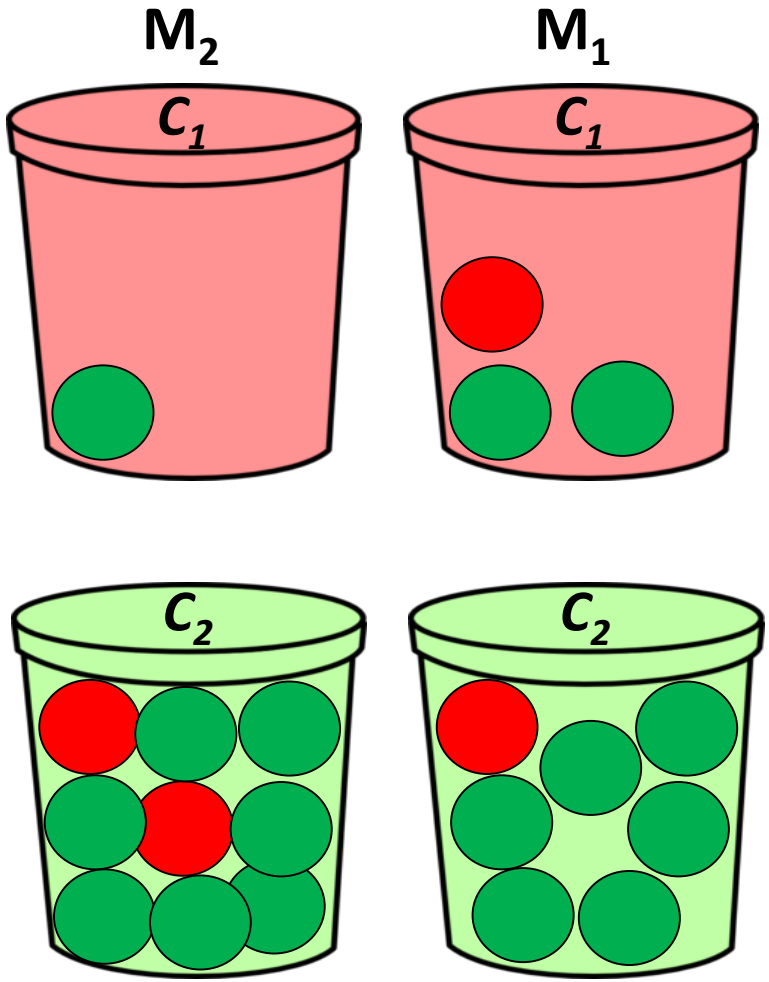




Binary Classification

	M_1	M_2
Precision	1/3 = 0.33	0/1 = 0
Recall	1/2 = 0.5	0/2 = 0
F1	0.4	0





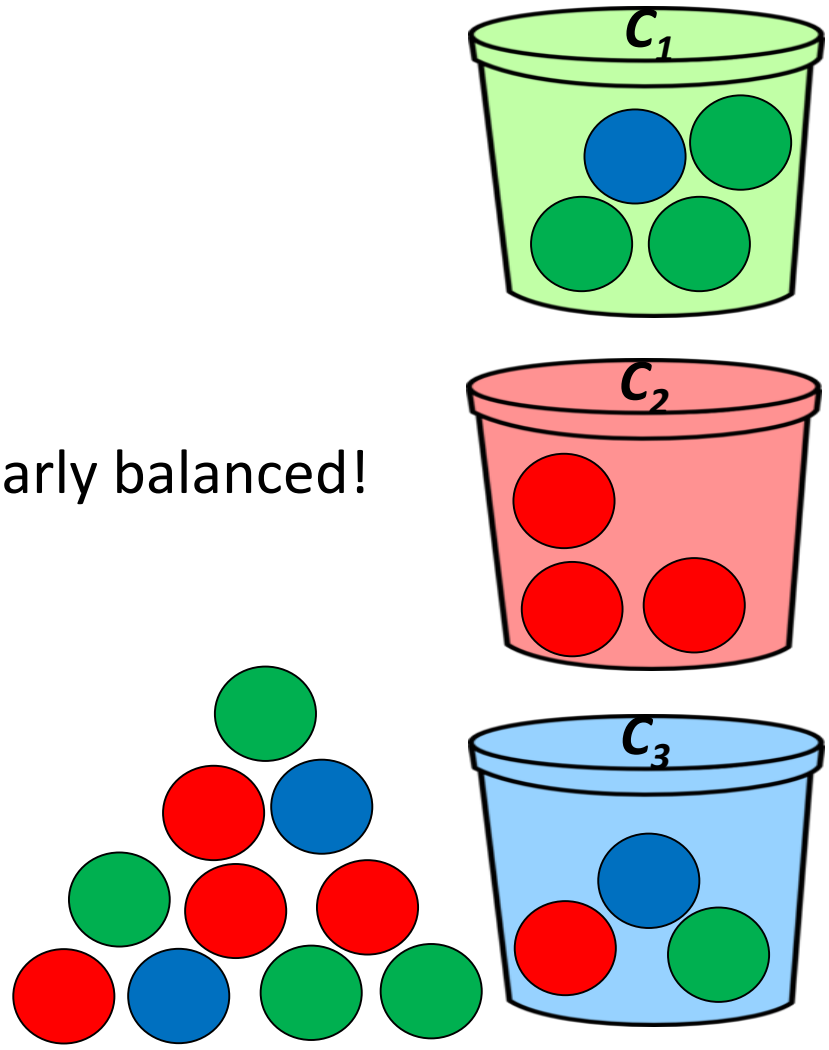
Multi-Class Classification

- Accuracy = ?

  = (3+3+1)/10 = 0.7

What's the problem?

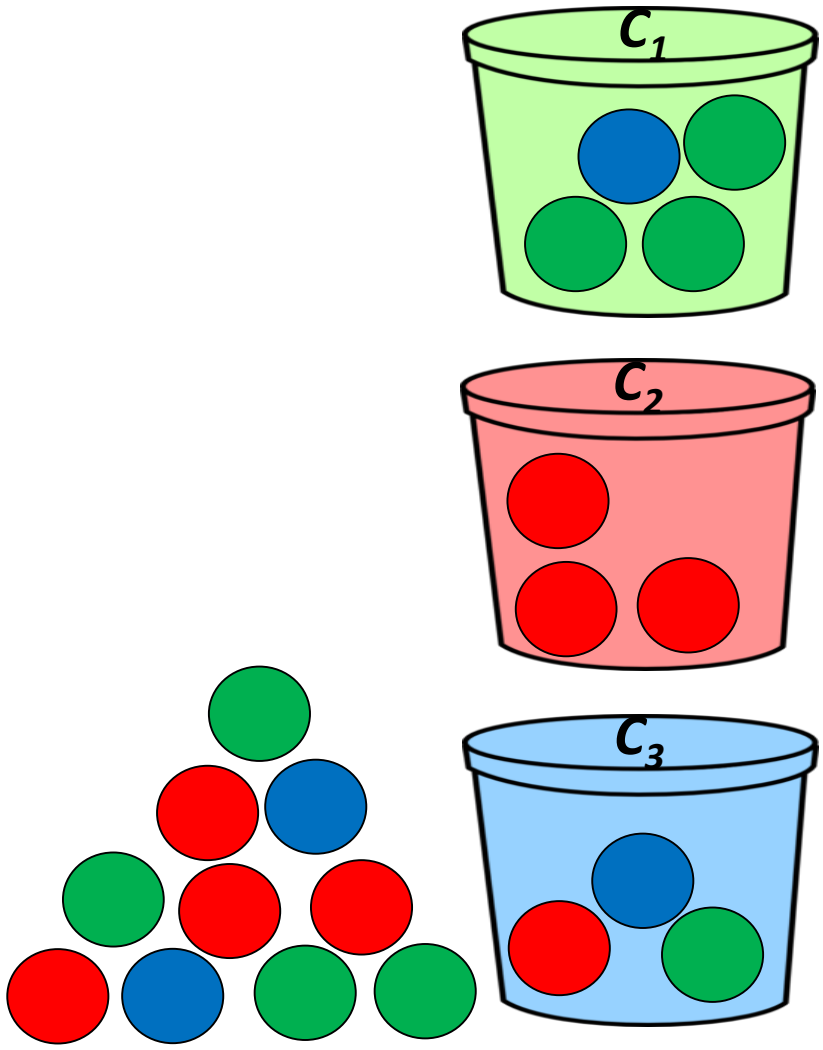
- Good measure when classes are nearly balanced!



Confusion Matrix

Predicted

Actual



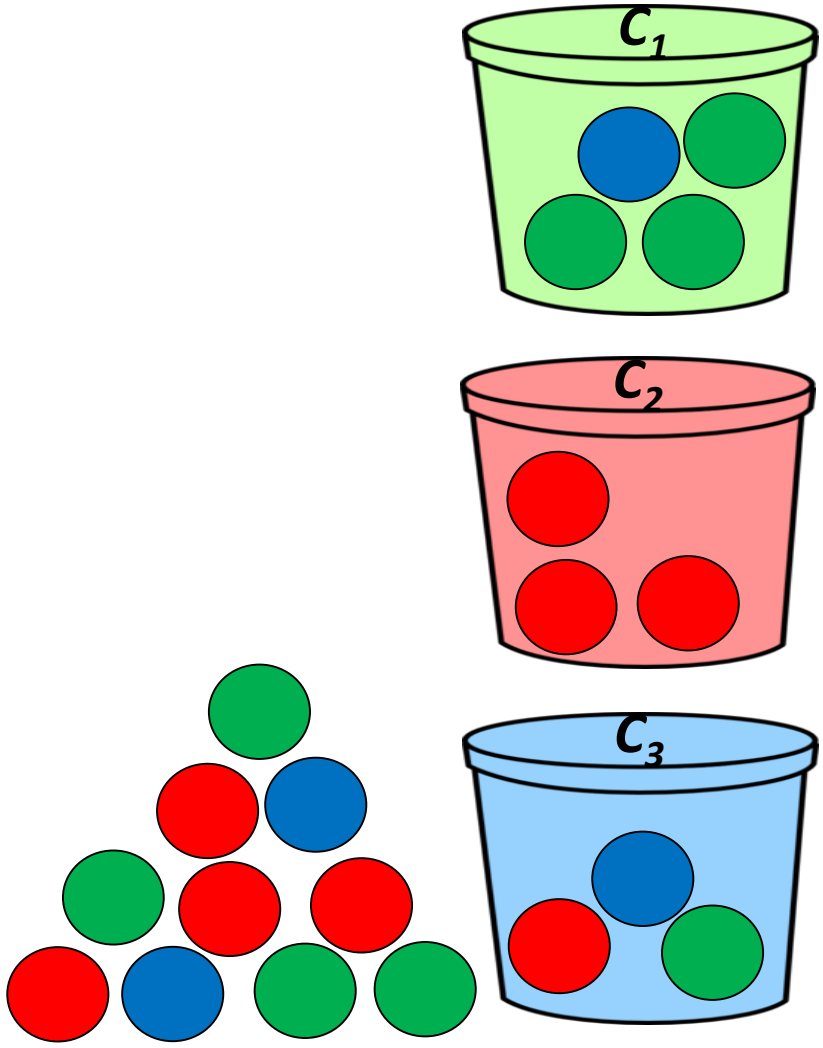
Confusion Matrix

Predicted

Actual

	3	0	1
	0	3	1
	1	0	1

P			
R			
F1			



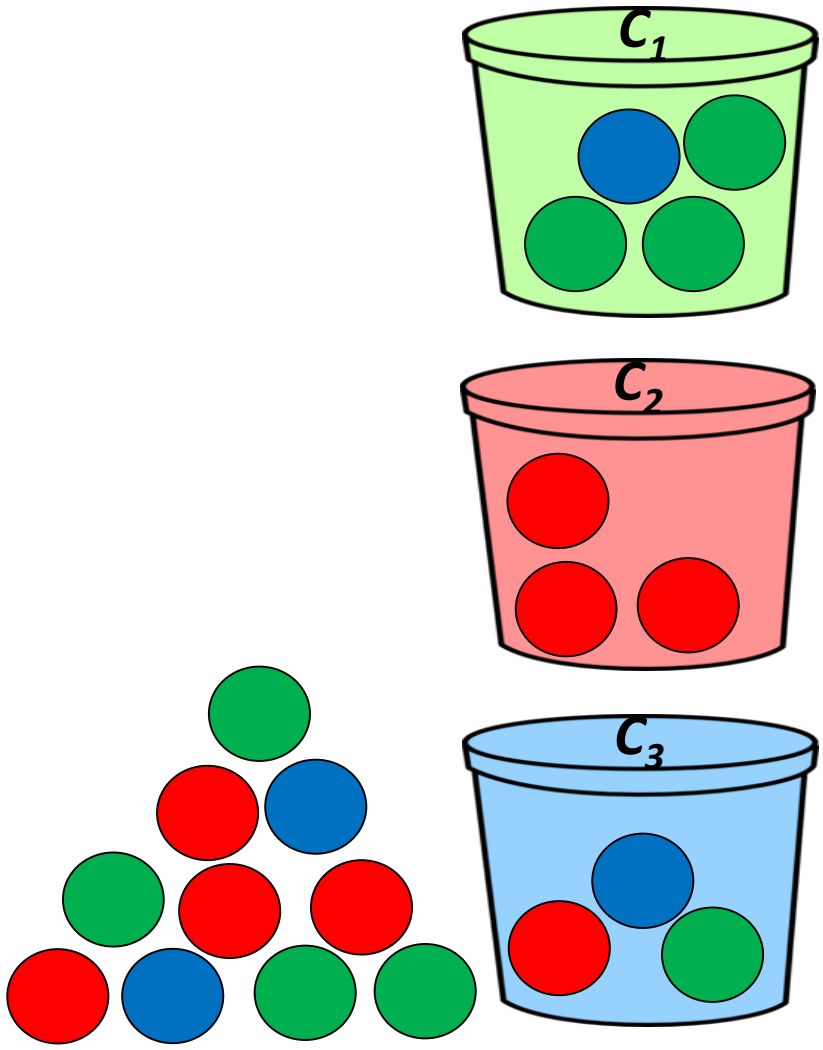
Confusion Matrix

Predicted

Actual

	3	0	1
	0	3	1
	1	0	1

P	0.75	1	0.333
R	0.75	0.75	0.5
F1	0.75	0.86	0.4



Macro-F1 = (0.75+0.86+0.4)/3 = **0.67**