

Multi-Label and Multinomial Classification

Natalie Parde

UIC CS 421

What if we have more than two classes?

Many NLP classification tasks have more than two classes

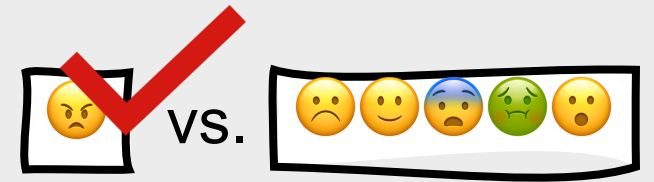
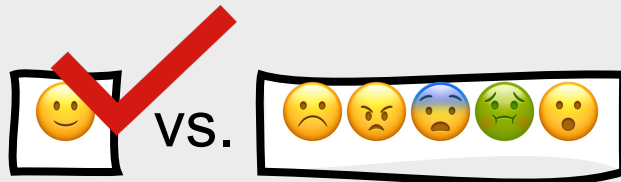
- Sentiment analysis (positive, negative, neutral)
- Part-of-speech tagging (noun, verb, adjective, etc.)
- Emotion detection (happy, sad, angry, surprised, afraid, disgusted)

Classification Paradigms

- Multi-label classification
- Multinomial classification

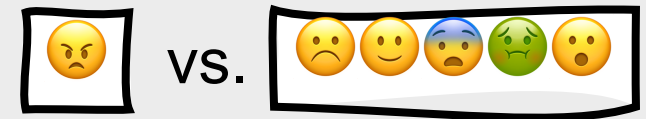
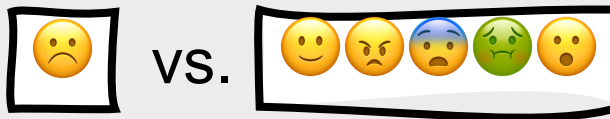
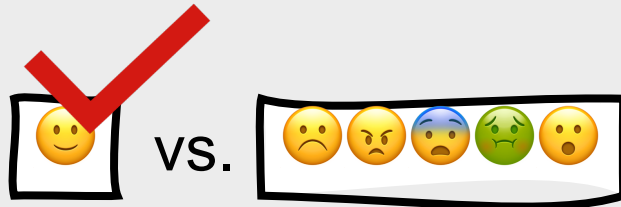
Multi-Label Classification

- Each document can be assigned more than one label
- How do we do this?
 - Build separate binary classifiers for each class
 - Positive class vs. every other class
 - Run each classifier on the test document
 - Each classifier makes its decision independently of the other classifiers, therefore allowing multiple labels to be assigned to the document



Multinomial Classification

- Each document can only be assigned one label
- How do we do this?
 - Same setup:
 - Build separate binary classifiers for each class
 - Run each classifier on the test document
 - Different outcome:
 - Choose the label from the classifier with the highest score



Multi-Class Contingency Matrix

		Actual		
		class 1	class 2	class 3
Predicted	class 1	a	b	c
	class 2	d	e	f
	class 3	g	h	i

Multi-Class Precision

		Actual		
		class 1	class 2	class 3
Predicted	class 1	a	b	c
	class 2	d	e	f
	class 3	g	h	i

$$\text{Precision} = \frac{a}{a+b+c}$$

Multi-Class Recall

		Actual		
		class 1	class 2	class 3
Predicted	class 1	a	b	c
	class 2	d	e	f
	class 3	g	h	i

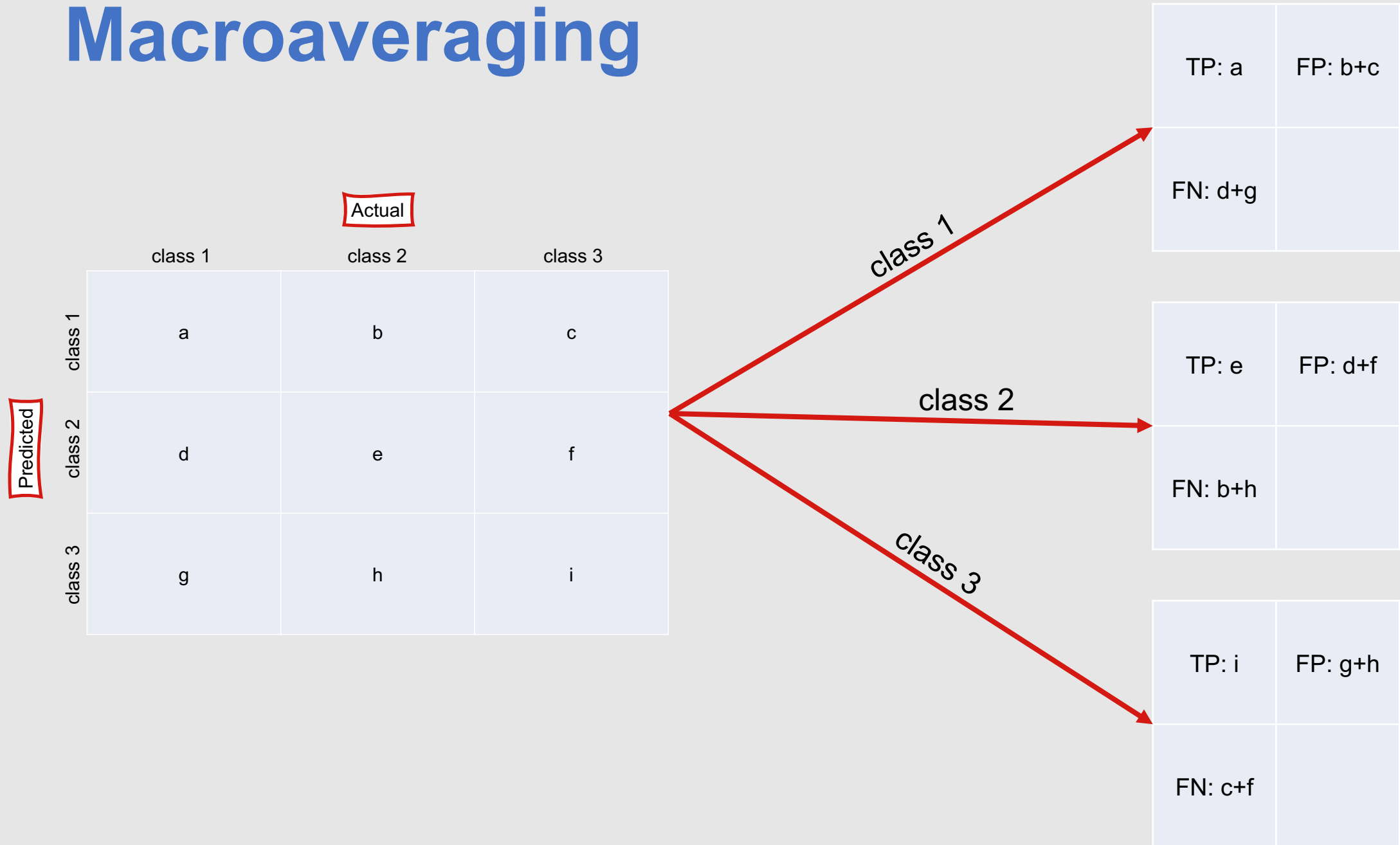
$$\text{Precision} = \frac{a}{a+b+c}$$

$$\text{Recall} = \frac{a}{a+d+g}$$

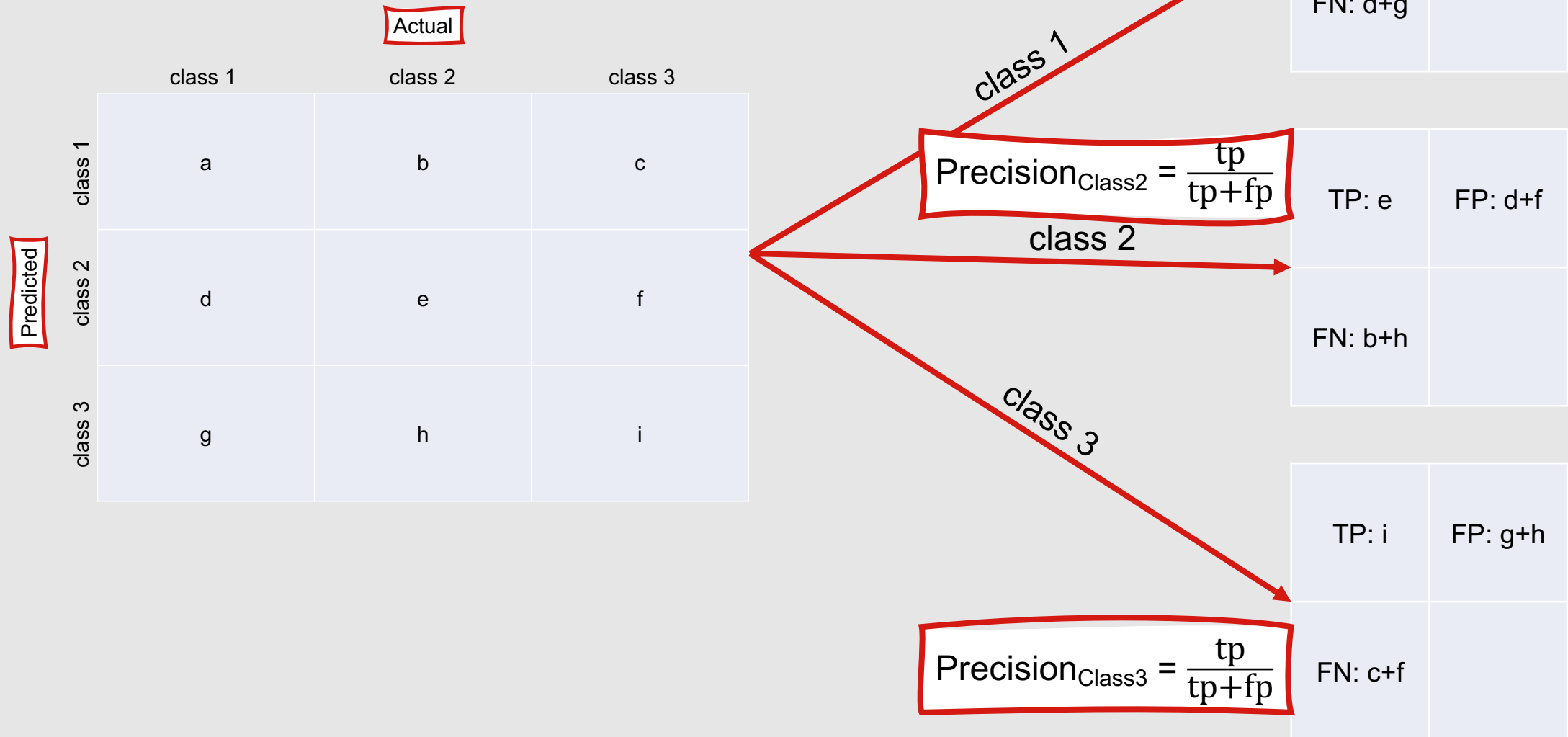
Macroaveraging and Microaveraging

- We can check the system's **overall performance** in multi-class classification settings by combining all of the precision values (or all of the recall values) in two ways:
 - **Macroaveraging**
 - **Microaveraging**
- **Macroaveraging:** Compute the performance for each class, and then average over all classes
- **Microaveraging:** Collect decisions for all classes into a single contingency table, and compute precision and recall from that table

Macroaveraging



Macroaveraging



Macroaveraging

		Actual		
		class 1	class 2	class 3
Predicted	class 1	a	b	c
	class 2	d	e	f
	class 3	g	h	i

$$\text{Precision}_{\text{Class1}} = \frac{tp}{tp+fp}$$

TP: a	FP: b+c
FN: d+g	

$$\text{Precision}_{\text{Class2}} = \frac{tp}{tp+fp}$$

TP: e	FP: d+f
FN: b+h	

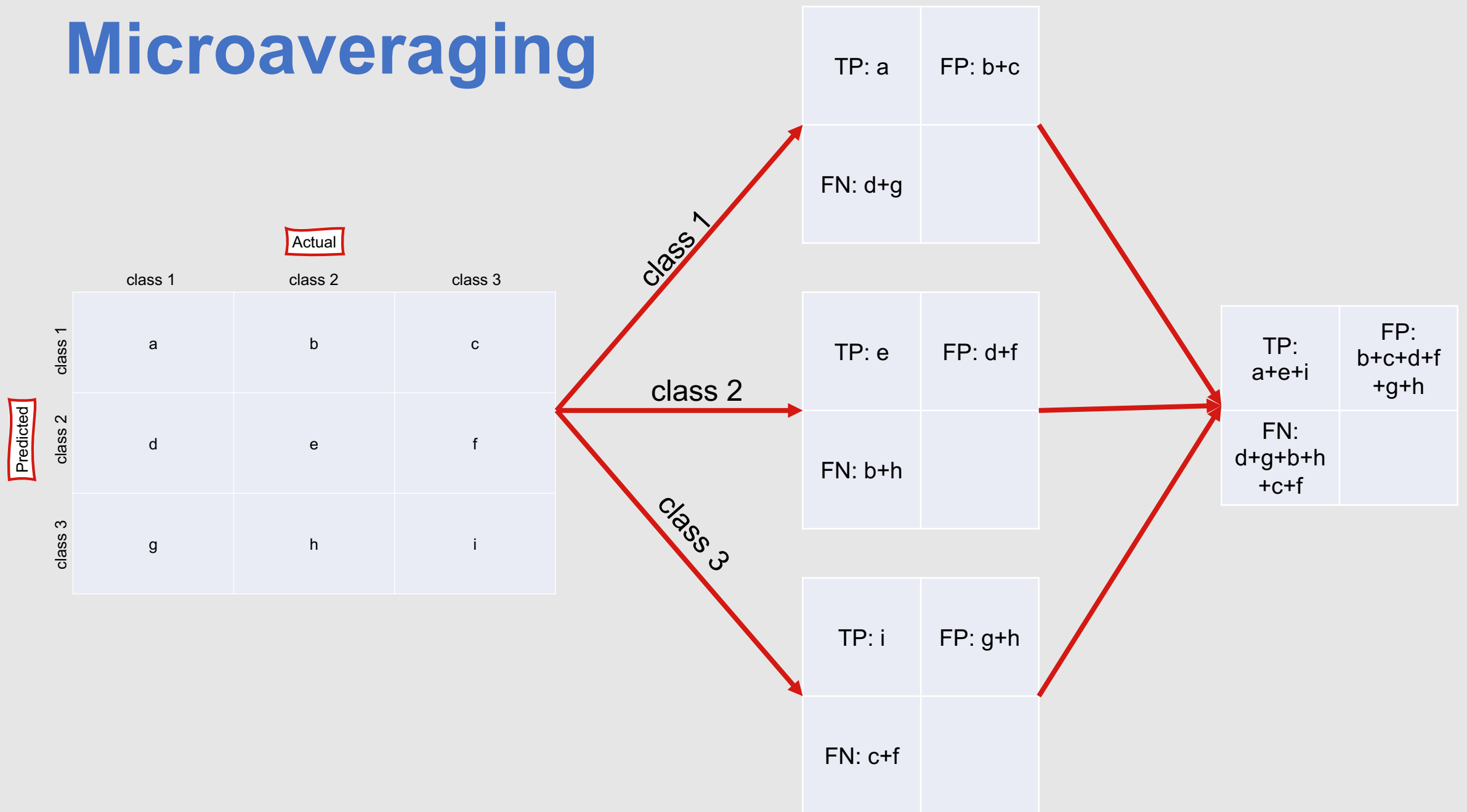
$$\text{Precision}_{\text{Class3}} = \frac{tp}{tp+fp}$$

TP: i	FP: g+h
FN: c+f	

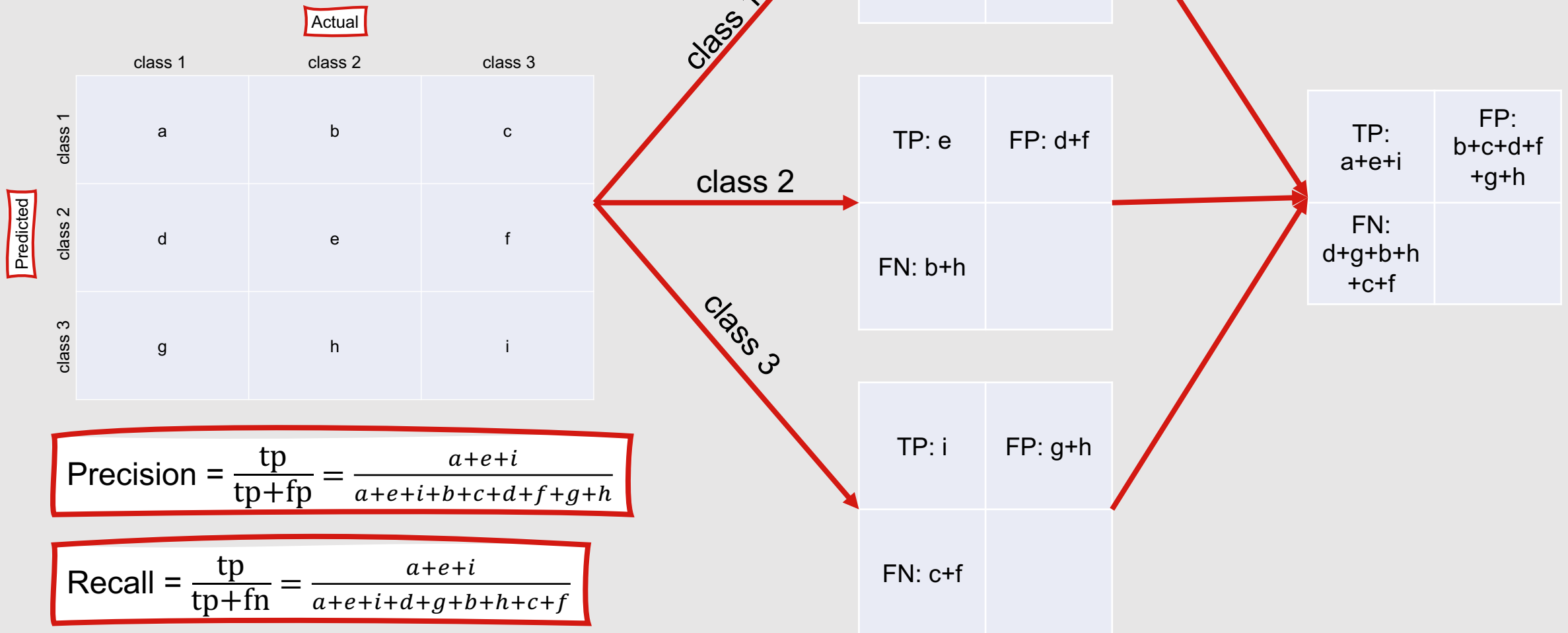
Macroaveraged Precision

$$= \frac{\text{Precision}_{\text{Class1}} + \text{Precision}_{\text{Class2}} + \text{Precision}_{\text{Class3}}}{3}$$

Microaveraging



Microaveraging



What's better: Microaveraging or macroaveraging?

- Depends on the scenario!
- Microaverages tend to be dominated by more frequent classes, since the counts are all pooled together
- Macroaverages tend to be more evenly distributed across classes
- Thus, if performance on all classes is equally important, macroaveraging is probably better; if performance on the most frequent class is more important, microaveraging is probably better