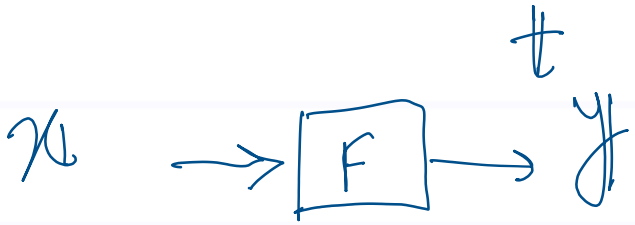


x, t → Sup

x → un sup

x, some t → Semi-sup

Data Annotation

EE698V - Machine Learning for Signal Processing

Vipul Arora

Data Collection

- Imbalance over
  - classes/regions
    - speech vs gun shots
    - images: tree vs lion
  - input variations, background conditions
    - speech/music in hostel mess vs that in library
    - lighting conditions; trees of India vs those of Britain

Data Annotation or Labelling

- Creating ground truth or targets
- Done manually

Audio Event Detection

- Strong labels
  - mark each event with time boundaries
  - slower
- Weak labels or tags
  - tag all events in the audio (no boundaries)
  - faster

Challenge: Inherent ambiguities in tasks

- Examples
  - Speech or music
  - Emotion recognition
  - People talking or TV
- Resolved the hard way

Challenge: Annotators' mistakes

- Because of
  - untrained annotators - TRAIN THEM
  - inattention - VERIFY
    - very boring job
- Crowdsourcing
  - multiple untrained annotators

Quality Control

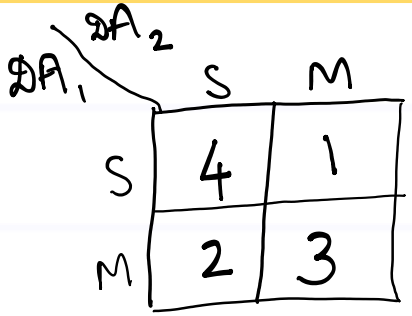
- Quantify the quality of annotation
  - metrics
    - Accuracy
    - Cohen’s Kappa
- Locate the mistakes

Quality Control

- Multiple annotators for the same sample
  - blind to others' annotations
- Compare mutually
- Inter Annotator Agreement (IAA) analysis

Annotation Quality Metrics

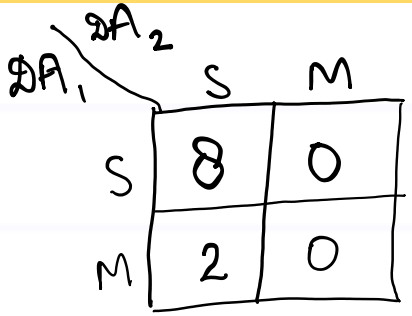
DA 1	DA 2
S	M
S	S
S	S
S	S
S	S
M	M
M	S
M	M
M	M
M	S



Accuracy = ? $\frac{7}{10}$

Annotation Quality Metrics

DA 1	DA 2
S	S
S	S
S	S
S	S
S	S
S	S
S	S
S	S
S	S
M	S
M	S



Accuracy = ?

$\frac{8}{10}$

Taking care of data imbalance

- You cannot use precision, recall, etc. here
- Use Cohen's Kappa measure
- Let us interpret accuracy in terms of probability theory

Accuracy

Random variables : y , y'

Accuracy = $P(y = y')$

Joint probability $P(y, y')$

then, Accuracy = $\sum_c P(y = c, y' = c)$

DA_1 \ DA_2	S	M
S	4	1
M	2	3

Accuracy = $\frac{4}{10} + \frac{3}{10} = 0.7$

DA_1 \ DA_2	S	M
S	8	0
M	2	0

Accuracy = $\frac{8}{10} + \frac{0}{10} = 0.8$ even if DA_2 is musically challenged

Cohen's Kappa

- The annotations are
  - good if $P(y = y')$ is high
  - bad if y and y' are independent
- So, measure $P(y = y') - P(y)P(y')$
- Perfect score should be 1, so normalize

$$\kappa = \frac{P(y = y') - P(y)P(y')}{1 - P(y)P(y')}$$

Example

$$k = \frac{\sum_c \left[P(y=c, y'=c) - P(y=c)\,P(y'=c) \right]}{1 - \sum_c P(y=c)\,P(y'=c)}$$

Random variables : y , y'

DA_1 \ DA_2	S	M
S	4	1
M	2	3

$$\text{Accuracy} = \frac{4}{10} + \frac{3}{10} = 0.7$$

$$k = \frac{\left(\frac{4}{10} - \frac{6}{10} \times \frac{5}{10}\right) + \left(\frac{3}{10} - \frac{4}{10} \times \frac{5}{10}\right)}{1 - \left(\frac{6}{10} \times \frac{5}{10} + \frac{4}{10} \times \frac{5}{10}\right)} = \frac{0.1 + 0.1}{1 - 0.5} = \frac{2}{5} = 0.4$$

DA_1 \ DA_2	S	M
S	8	0
M	2	0

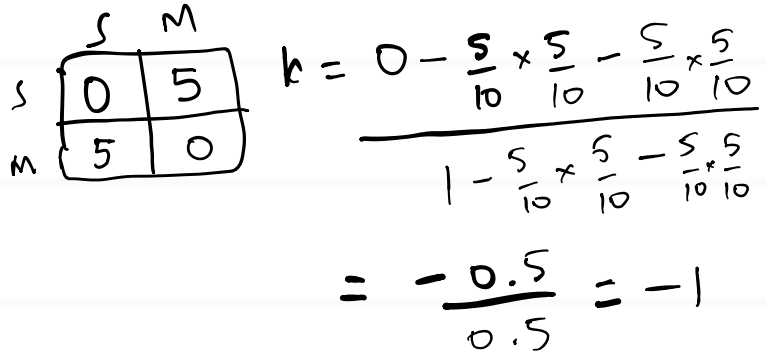
$$\text{Accuracy} = \frac{8}{10} + \frac{0}{10} = 0.8$$

even if DA_2 is musically challenged

$$k = \frac{\left(\frac{8}{10} - \frac{10}{10} \times \frac{8}{10}\right) + \left(0 - \frac{2}{10} \times \frac{0}{10}\right)}{1 - \left(\frac{8}{10} \times \frac{10}{10} + \frac{2}{10} \times \frac{0}{10}\right)} = 0$$

Interpreting Cohen's Kappa scores

- If $\kappa = 1$
  - perfect agreement
- If $\kappa = 0$
  - independence; random annotations
- If $\kappa < 0$
  - more disagreements than agreements
- Generally, $\kappa > 0.4$ is considered good



Verification

- Locate the samples where annotators disagree
- Pass them on for annotation to an expert (verifier)

DA 1	DA 2	
S	M	X
S	S	
S	S	
S	S	
S	S	
M	M	
M	S	X
M	M	
M	M	
M	S	X

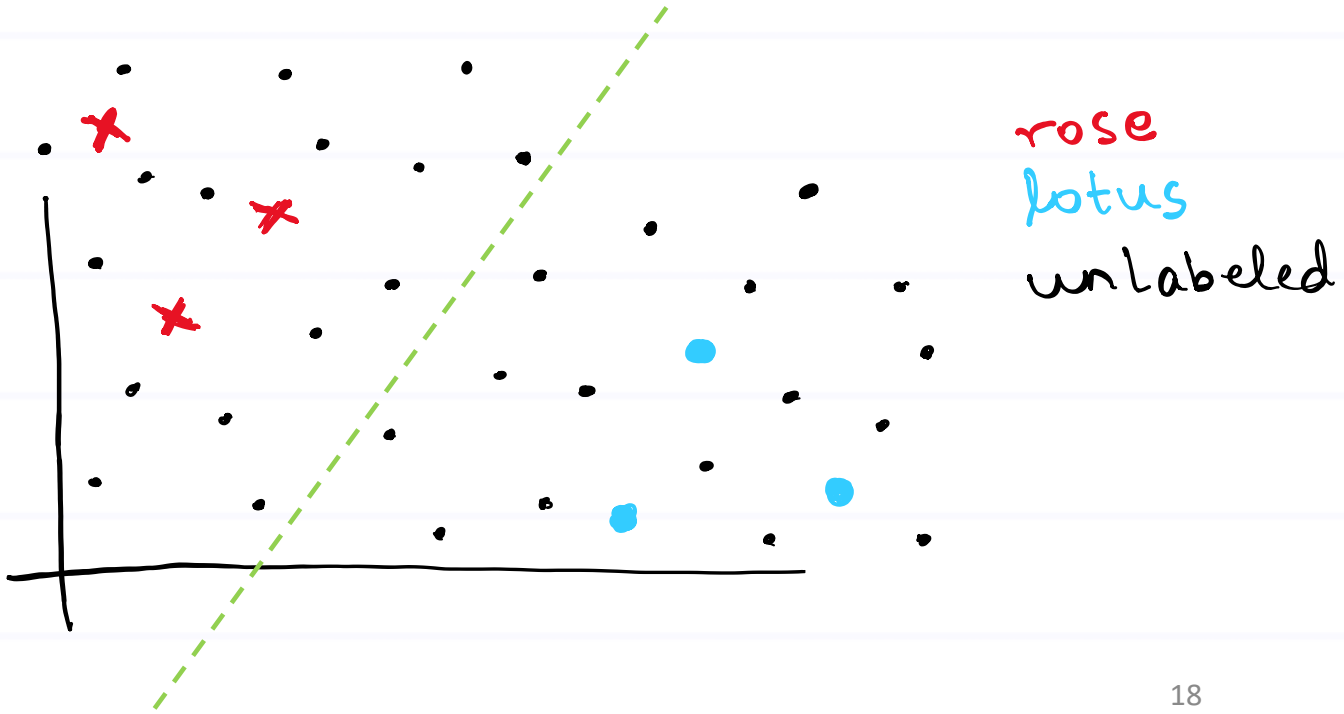
Annotator Quality

- Can you use the metric to find which DA needs more training?

κ	DA 1	DA 2	DA 3	DA 4
DA 1	1.0			
DA 2	0.1	1.0		
DA 3	0.6	0.2	1.0	
DA 4	0.7	-0.1	0.5	1.0

Speeding up

- Identify which samples are most informative
  - Annotate only those
- ACTIVE LEARNING
  - Which points to label?



References

- Wikipedia
- https://www.youtube.com/watch?v=CRz6qkoSqvk
- Active Learning:
  - B. Settles, “Active learning literature survey,” University of Wisconsin, Madison, Computer Sciences Technical Report 1648, 2010.
  - V. Arora, A. Lahiri, and H. Reetz, ‘Phonological Feature Based Mispronunciation Detection and Diagnosis Using Multi-Task DNNs and Active Learning’, in Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2017, https://doi.org/10.21437/Interspeech.2017