Quality Control - part 3

**Crowdsourcing and Human Computation
Lecture 11**

**Instructor: Chris Callison-Burch
TA: Ellie Pavlick**

Website: crowdsourcing-class.org

Different Mechanisms for Quality Control

- Aggregation and redundancy

- Embedded gold standard data

- Reputation systems

- Economic incentives

- **Statistical models**

Expectation Maximization algorithm

- EM is an algorithm for finding the probabilities of unobserved varaibles

- We will use it to estimate how accurate Turkers' labels are, and infer how good each Turker is

- This is more sophisticated than voting

Dawid and Skene (1977)

- *Maximum Likelihood Estimation of Observer Error-rates using the EM Algorithm*

- Examined application to medical diagnosis

- Patients are sometimes treated by multiple physicians, who can give different diagnoses

- Why? Doctors may has different questions. Patient may describe history differently. Doctors may classify symptoms differently

Observer Error

- Given that different doctors have different opinions, they can't all be right.

- How often do individual physicians suffer from "observer error"? Are their errors systematic?
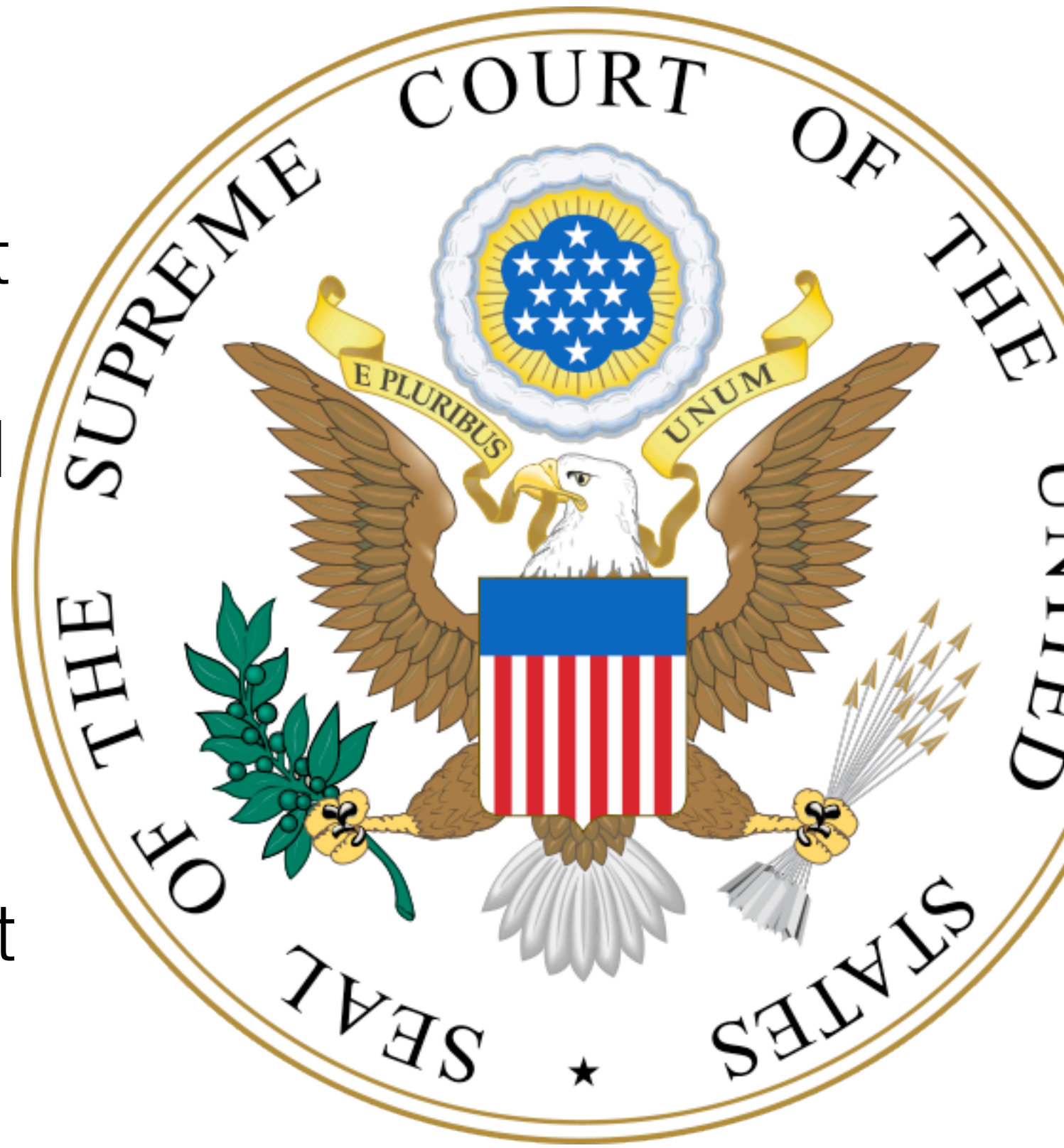
- Answers depend on the "true" diagnosis

Observer Error

- Observer error would be easy to calculate if we had ground truth

- Simply count the misdiagnoses and divide by the total number of diagnoses

- However, sometimes it is impossible to know what diagnosis is correct. Same set of symptoms can arise from multiple root causes.

"I know it when I see it"

I shall not today attempt further to define "hard-core pornography"; and perhaps I could never succeed in intelligibly doing so. But I know it when I see it.
—Justice Potter Stewart

url	worker1	worker2	worker3	worker4	worker5
google.com	porn	not porn	not porn	not porn	porn
panda-cam.gov	porn	porn	not porn	not porn	porn
sex-mission.com	porn	porn	porn	porn	not porn
sunnyfun.com	porn	not porn	not porn	not porn	porn
youporn.com	porn	porn	porn	porn	not porn

Solution?

- Can't have Justice Stewart rule on everything

- Instead, we will apply Dawid and Skene's EM algorithm, which iteratively

  1. Estimates the correct answers, using labels from multiple workers, and accounts for the quality of each worker

  2. Estimates the quality of the workers by comparing the submitted answers to the inferred correct answers

Inputs

- a set of **N** objects o$_1$ ... o$_N$
  google.com, panda-cam.gov, sex-mission.com, sunnyfun.com, youporn.com

- a set of **L** possible labels:
  {porn, not porn}

- Labels for each object by **K** workers
  worker1, worker2, worker3, worker4, worker5

Goal 1

- Recover the true class label $\mathbf{T}(o_n)$ for each object o_n when "gold" truth is unknown

- Since the true labels are not known / never directly observed, they are called *latent* variables

Goal 2

- For each worker who contributed labels, calculate their accuracy or reliability

- To calculate accuracy show how often they mistakenly choose one label when a different one is the actual truth

Chicken and egg problem

- If we knew what the **true class labels** were for each object for each object, then we could compute each Turker's accuracy

- If we had **accuracies for every Turker**, then we could infer what the true label for each object should be

Input: Labels $l[k][n]$ from worker (k) to object o_n,

Output: Confusion matrix $\pi_{ij}^{(k)}$ for each worker (k), Correct labels $T(o_n)$ for each object o_n, Class priors $Pr\{C\}$ for each class C

1 Initialize error rates $\pi_{ij}^{(k)}$ for each worker (k) (e.g., assume each worker is perfect);

2 Initialize correct label for each object $T(o_n)$ (e.g., using majority vote);

3 **while** *not converged* **do**

4   Estimate the correct label $T(o_n)$ for each object, using the labels $l[\cdot][n]$ assigned to o_n by workers, weighting the votes using the error rates $\pi_{ij}^{(k)}$;

5   Estimate the error rates $\pi_{ij}^{(k)}$, for each worker (k), using the correct labels $T(o_n)$ and the assigned labels $l[k][n]$;

6   Estimate the class priors $Pr\{C\}$, for each class C;

7 **end**

8 **return** *Estimated error rates* $\pi_{ij}^{(k)}$, *Estimated correct labels* $T(o_n)$, *Estimated class priors* $Pr\{C\}$

Algorithm 1: The EM algorithm for worker quality estimation.

Input

url	worker1	worker2	worker3	worker4	worker5
google.com	porn	not porn	not porn	not porn	porn
panda-cam.gov	porn	porn	not porn	not porn	porn
sex-mission.com	porn	porn	porn	porn	not porn
sunnyfun.com	porn	not porn	not porn	not porn	porn
youporn.com	porn	porn	porn	porn	not porn

Output: "True" Labels

url	True Labels
google.com	not porn
panda-cam.gov	not porn
sex-mission.com	porn
sunnyfun.com	not porn
youporn.com	porn

Output: Worker Accuracies

Guess

w1	porn	not
porn	100%	0%
not	100%	0%

Truth

w2	porn	not
porn	100%	0%
not	33%	67%

w3	porn	not
porn	100%	0%
not	0%	100%

Guess

w4	porn	not
porn	100%	0%
not	0%	100%

Truth

w5	porn	not
porn	0%	100%
not	100%	0%

Initialize

Guess

w1	porn	not
porn	100%	0%
not	0%	100%

w2	porn	not
porn	100%	0%
not	0%	100%

w3	porn	not
porn	100%	0%
not	0%	100%

Guess

w4	porn	not
porn	100%	0%
not	0%	100%

w5	porn	not
porn	100%	0%
not	0%	100%

Truth

Initialize

url	worker 1	worker 2	worker 3	worker 4	worker 5
google.com	porn	not porn	not porn	not porn	porn
panda-cam.gov	porn	porn	not porn	not porn	porn
sex-mission.com	porn	porn	porn	porn	not porn
sunnyfun.com	porn	not porn	not porn	not porn	porn
youporn.com	porn	porn	porn	porn	not porn

p(porn)	p(not porn)
40%	60%
60%	40%
80%	20%
40%	60%
80%	20%

Re-Calculate Worker Scores

Guess

w1	porn	not
porn	40	0
not	60	0

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn	60%	40%
sex-mission.com	porn	porn	porn	porn	not porn	80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn	40%	60%
youporn.com	porn	porn	porn	porn	not porn	80%	20%

Re-Calculate Worker Scores

Guess

w1	porn	not
porn	100	0
not	100	0

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn	60%	40%
sex-mission.com	porn	porn	porn	porn	not porn	80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn	40%	60%
youporn.com	porn	porn	porn	porn	not porn	80%	20%

Re-Calculate Worker Scores

Guess

w1	porn	not
porn	180	0
not	120	0

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn	60%	40%
sex-mission.com	porn	porn	porn	porn	not porn	80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn	40%	60%
youporn.com	porn	porn	porn	porn	not porn	80%	20%

Re-Calculate Worker Scores

Guess

w1	porn	not
porn	220	0
not	180	0

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn	60%	40%
sex-mission.com	porn	porn	porn	porn	not porn	80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn	40%	60%
youporn.com	porn	porn	porn	porn	not porn	80%	20%

Re-Calculate Worker Scores

Guess

w1	porn	not
porn	300	0
not	200	0

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn	60%	40%
sex-mission.com	porn	porn	porn	porn	not porn	80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn	40%	60%
youporn.com	porn	porn	porn	porn	not porn	80%	20%

Re-Calculate Worker Scores

Guess		
w1	porn	not
	porn	not

Truth		
porn	60%	0
not	40%	0

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn	60%	40%
sex-mission.com	porn	porn	porn	porn	not porn	80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn	40%	60%
youporn.com	porn	porn	porn	porn	not porn	80%	20%

Re-Calculate Worker Scores

Guess

w2	porn	not
porn		40
not		60

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5		p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn		40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn		60%	40%
sex-mission.com	porn	porn	porn	porn	not porn		80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn		40%	60%
youporn.com	porn	porn	porn	porn	not porn		80%	20%

Re-Calculate Worker Scores

Guess

w2	porn	not
Truth porn	60	40
not	40	60

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn	60%	40%
sex-mission.com	porn	porn	porn	porn	not porn	80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn	40%	60%
youporn.com	porn	porn	porn	porn	not porn	80%	20%

Re-Calculate Worker Scores

Guess

w2	porn	not
porn	140	40
not	60	60

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn	60%	40%
sex-mission.com	porn	porn	porn	porn	not porn	80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn	40%	60%
youporn.com	porn	porn	porn	porn	not porn	80%	20%

Re-Calculate Worker Scores

Guess

w2	porn	not
porn	140	100
not	60	100

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn	60%	40%
sex-mission.com	porn	porn	porn	porn	not porn	80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn	40%	60%
youporn.com	porn	porn	porn	porn	not porn	80%	20%

Re-Calculate Worker Scores

Guess

w2	porn	not
porn	220	100
not	80	100

(Truth on vertical axis)

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn	60%	40%
sex-mission.com	porn	porn	porn	porn	not porn	80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn	40%	60%
youporn.com	porn	porn	porn	porn	not porn	80%	20%

Re-Calculate Worker Scores

Guess

w2	porn	not
Truth porn	73%	50%
not	27%	50%

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn	60%	40%
sex-mission.com	porn	porn	porn	porn	not porn	80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn	40%	60%
youporn.com	porn	porn	porn	porn	not porn	80%	20%

Re-Calculate Worker Scores

Guess

w3	porn	not
porn		40
not		60

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5		p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn		40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn		60%	40%
sex-mission.com	porn	porn	porn	porn	not porn		80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn		40%	60%
youporn.com	porn	porn	porn	porn	not porn		80%	20%

Re-Calculate Worker Scores

w3 (Truth)	Guess porn	Guess not
porn		100
not		100

Guess — (column header above table)

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn	60%	40%
sex-mission.com	porn	porn	porn	porn	not porn	80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn	40%	60%
youporn.com	porn	porn	porn	porn	not porn	80%	20%

Re-Calculate Worker Scores

Guess

w3	porn	not
porn (Truth)	80	100
not (Truth)	20	100

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn	60%	40%
sex-mission.com	porn	porn	porn	porn	not porn	80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn	40%	60%
youporn.com	porn	porn	porn	porn	not porn	80%	20%

Re-Calculate Worker Scores

w3	Guess porn	Guess not
Truth porn	80	140
Truth not	20	160

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn	60%	40%
sex-mission.com	porn	porn	porn	porn	not porn	80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn	40%	60%
youporn.com	porn	porn	porn	porn	not porn	80%	20%

Re-Calculate Worker Scores

Guess		
w3	porn	not
porn	160	140
not	40	160

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn	60%	40%
sex-mission.com	porn	porn	porn	porn	not porn	80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn	40%	60%
youporn.com	porn	porn	porn	porn	not porn	80%	20%

Re-Calculate Worker Scores

Guess

w3	porn	not
porn	80%	47%
not	20%	53%

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn	60%	40%
sex-mission.com	porn	porn	porn	porn	not porn	80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn	40%	60%
youporn.com	porn	porn	porn	porn	not porn	80%	20%

Re-Calculate Worker Scores

Guess

w5	porn	not
porn (Truth)	40	
not	60	

url	worker 1	worker 2	worker 3	worker 4	worker 5		p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn		40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn		60%	40%
sex-mission.com	porn	porn	porn	porn	not porn		80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn		40%	60%
youporn.com	porn	porn	porn	porn	not porn		80%	20%

Re-Calculate Worker Scores

w5	porn	not
porn	100	
not	100	

Guess

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5		p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn		40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn		60%	40%
sex-mission.com	porn	porn	porn	porn	not porn		80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn		40%	60%
youporn.com	porn	porn	porn	porn	not porn		80%	20%

Re-Calculate Worker Scores

Guess

w5	porn	not
porn	100	80
not	100	20

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5		p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn		40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn		60%	40%
sex-mission.com	porn	porn	porn	porn	not porn		80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn		40%	60%
youporn.com	porn	porn	porn	porn	not porn		80%	20%

Re-Calculate Worker Scores

Guess

w5	porn	not
porn	140	80
not	160	20

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn	60%	40%
sex-mission.com	porn	porn	porn	porn	not porn	80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn	40%	60%
youporn.com	porn	porn	porn	porn	not porn	80%	20%

Re-Calculate Worker Scores

Guess

w5	porn	not
porn	140	160
not	160	40

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5		p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn		40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn		60%	40%
sex-mission.com	porn	porn	porn	porn	not porn		80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn		40%	60%
youporn.com	porn	porn	porn	porn	not porn		80%	20%

Re-Calculate Worker Scores

Guess

w5	porn	not
porn	47%	80%
not	53%	20%

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	40%	60%
panda-cam.gov	porn	porn	not porn	not porn	porn	60%	40%
sex-mission.com	porn	porn	porn	porn	not porn	80%	20%
sunnyfun.com	porn	not porn	not porn	not porn	porn	40%	60%
youporn.com	porn	porn	porn	porn	not porn	80%	20%

Update Estimates for True Labels

Guess

w1	porn	not
porn	60%	0
not	40%	0

(Truth labels rows)

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	60	40
panda-cam.gov	porn	porn	not porn	not porn	porn	60	40
sex-mission.com	porn	porn	porn	porn	not porn	60	40
sunnyfun.com	porn	not porn	not porn	not porn	porn	60	40
youporn.com	porn	porn	porn	porn	not porn	60	40

Update Estimates for True Labels

Guess

w2	porn	not
porn (Truth)	73%	50%
not (Truth)	27%	50%

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	110	90
panda-cam.gov	porn	porn	not porn	not porn	porn	60	40
sex-mission.com	porn	porn	porn	porn	not porn	60	40
sunnyfun.com	porn	not porn	not porn	not porn	porn	60	40
youporn.com	porn	porn	porn	porn	not porn	60	40

Update Estimates for True Labels

Guess

w2	porn	not
porn	73%	50%
not	27%	50%

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	110	90
panda-cam.gov	porn	porn	not porn	not porn	porn	133	67
sex-mission.com	porn	porn	porn	porn	not porn	60	40
sunnyfun.com	porn	not porn	not porn	not porn	porn	60	40
youporn.com	porn	porn	porn	porn	not porn	60	40

Update Estimates for True Labels

Guess

w2	porn	not
porn	73%	50%
not	27%	50%

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	110	90
panda-cam.gov	porn	porn	not porn	not porn	porn	133	67
sex-mission.com	porn	porn	porn	porn	not porn	133	67
sunnyfun.com	porn	not porn	not porn	not porn	porn	60	40
youporn.com	porn	porn	porn	porn	not porn	60	40

Update Estimates for True Labels

Guess		
w2	porn	not

Truth	porn	not
porn	73%	50%
not	27%	50%

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	110	90
panda-cam.gov	porn	porn	not porn	not porn	porn	133	67
sex-mission.com	porn	porn	porn	porn	not porn	133	67
sunnyfun.com	porn	not porn	not porn	not porn	porn	110	90
youporn.com	porn	porn	porn	porn	not porn	60	40

Update Estimates for True Labels

Guess

w2	porn	not
Truth porn	73%	50%
not	27%	50%

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	110	90
panda-cam.gov	porn	porn	not porn	not porn	porn	133	67
sex-mission.com	porn	porn	porn	porn	not porn	133	67
sunnyfun.com	porn	not porn	not porn	not porn	porn	110	90
youporn.com	porn	porn	porn	porn	not porn	133	67

Update Estimates for True Labels

w3	porn	not
porn	80%	47%
not	20%	53%

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
oogle.com	porn	not porn	not porn	not porn	porn	157	143
panda-cam.gov	porn	porn	not porn	not porn	porn	133	67
sex-mission.com	porn	porn	porn	porn	not porn	133	67
sunnyfun.com	porn	not porn	not porn	not porn	porn	110	90
youporn.com	porn	porn	porn	porn	not porn	133	67

Update Estimates for True Labels

Guess

w3	porn	not
porn	80%	47%
not	20%	53%

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5		p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn		157	143
panda-cam.gov	porn	porn	not porn	not porn	porn		180	120
sex-mission.com	porn	porn	porn	porn	not porn		133	67
sunnyfun.com	porn	not porn	not porn	not porn	porn		110	90
youporn.com	porn	porn	porn	porn	not porn		133	67

Update Estimates for True Labels

Guess

w3	porn	not
porn	80%	47%
not	20%	53%

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	157	143
panda-cam.gov	porn	porn	not porn	not porn	porn	180	120
sex-mission.com	porn	porn	porn	porn	not porn	213	87
sunnyfun.com	porn	not porn	not porn	not porn	porn	110	90
youporn.com	porn	porn	porn	porn	not porn	133	67

Update Estimates for True Labels

Guess

w3	porn	not
porn	80%	47%
not	20%	53%

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	157	143
panda-cam.gov	porn	porn	not porn	not porn	porn	180	120
sex-mission.com	porn	porn	porn	porn	not porn	213	87
sunnyfun.com	porn	not porn	not porn	not porn	porn	157	143
youporn.com	porn	porn	porn	porn	not porn	133	67

Update Estimates for True Labels

Truth	w3	porn	not
	porn	80%	47%
	not	20%	53%

url	worker 1	worker 2	worker 3	worker 4	worker 5		p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn		157	143
panda-cam.gov	porn	porn	not porn	not porn	porn		180	120
sex-mission.com	porn	porn	porn	porn	not porn		213	87
sunnyfun.com	porn	not porn	not porn	not porn	porn		157	143
youporn.com	porn	porn	porn	porn	not porn		213	87

Update Estimates for True Labels

w4	porn	not
porn	80%	47%
not	20%	53%

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5		p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn		204	196
panda-cam.gov	porn	porn	not porn	not porn	porn		180	120
sex-mission.com	porn	porn	porn	porn	not porn		213	87
sunnyfun.com	porn	not porn	not porn	not porn	porn		157	143
youporn.com	porn	porn	porn	porn	not porn		213	87

Update Estimates for True Labels

Guess

w4	porn	not
Truth porn	80%	47%
not	20%	53%

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	204	196
panda-cam.gov	porn	porn	not porn	not porn	porn	227	173
sex-mission.com	porn	porn	porn	porn	not porn	213	87
sunnyfun.com	porn	not porn	not porn	not porn	porn	157	143
youporn.com	porn	porn	porn	porn	not porn	213	87

Update Estimates for True Labels

Guess

w4	porn	not
porn	80%	47%
not	20%	53%

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	204	196
panda-cam.gov	porn	porn	not porn	not porn	porn	227	173
sex-mission.com	porn	porn	porn	porn	not porn	293	107
sunnyfun.com	porn	not porn	not porn	not porn	porn	157	143
youporn.com	porn	porn	porn	porn	not porn	213	87

Update Estimates for True Labels

Guess

w4	porn	not
porn	80%	47%
not	20%	53%

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	204	196
panda-cam.gov	porn	porn	not porn	not porn	porn	227	173
sex-mission.com	porn	porn	porn	porn	not porn	293	107
sunnyfun.com	porn	not porn	not porn	not porn	porn	204	196
youporn.com	porn	porn	porn	porn	not porn	213	87

Update Estimates for True Labels

w4	porn	not
porn	80%	47%
not	20%	53%

Truth

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	204	196
panda-cam.gov	porn	porn	not porn	not porn	porn	227	173
sex-mission.com	porn	porn	porn	porn	not porn	293	107
sunnyfun.com	porn	not porn	not porn	not porn	porn	204	196
youporn.com	porn	porn	porn	porn	not porn	293	107

Update Estimates for True Labels

w5	porn	not
porn	47%	80%
not	53%	20%

(Truth)

url	worker 1	worker 2	worker 3	worker 4	worker 5		p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn		251	249
panda-cam.gov	porn	porn	not porn	not porn	porn		227	173
sex-mission.com	porn	porn	porn	porn	not porn		293	107
sunnyfun.com	porn	not porn	not porn	not porn	porn		204	196
youporn.com	porn	porn	porn	porn	not porn		293	107

Update Estimates for True Labels

w5	Guess porn	Guess not
Truth porn	47%	80%
Truth not	53%	20%

url	worker 1	worker 2	worker 3	worker 4	worker 5		p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn		251	249
panda-cam.gov	porn	porn	not porn	not porn	porn		274	226
sex-mission.com	porn	porn	porn	porn	not porn		293	107
sunnyfun.com	porn	not porn	not porn	not porn	porn		204	196
youporn.com	porn	porn	porn	porn	not porn		293	107

Update Estimates for True Labels

Truth \ Guess (w5)	porn	not
porn	47%	80%
not	53%	20%

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	251	249
panda-cam.gov	porn	porn	not porn	not porn	porn	274	226
sex-mission.com	porn	porn	porn	porn	not porn	373	127
sunnyfun.com	porn	not porn	not porn	not porn	porn	204	196
youporn.com	porn	porn	porn	porn	not porn	293	107

Update Estimates for True Labels

w5	porn	not
Guess		
porn	47%	80%
not	53%	20%

Truth (row label for porn/not)

url	worker 1	worker 2	worker 3	worker 4	worker 5	p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn	251	249
panda-cam.gov	porn	porn	not porn	not porn	porn	274	226
sex-mission.com	porn	porn	porn	porn	not porn	373	127
sunnyfun.com	porn	not porn	not porn	not porn	porn	251	249
youporn.com	porn	porn	porn	porn	not porn	293	107

Update Estimates for True Labels

w5	porn	not
porn	47%	80%
not	53%	20%

Guess (column header), Truth (row header)

url	worker 1	worker 2	worker 3	worker 4	worker 5		p(porn)	p(not porn)
google.com	porn	not porn	not porn	not porn	porn		251	249
panda-cam.gov	porn	porn	not porn	not porn	porn		274	226
sex-mission.com	porn	porn	porn	porn	not porn		373	127
sunnyfun.com	porn	not porn	not porn	not porn	porn		251	249
youporn.com	porn	porn	porn	porn	not porn		373	127

Normalize

url	worker 1	worker 2	worker 3	worker 4	worker 5
google.com	porn	not porn	not porn	not porn	porn
panda-cam.gov	porn	porn	not porn	not porn	porn
sex-mission.com	porn	porn	porn	porn	not porn
sunnyfun.com	porn	not porn	not porn	not porn	porn
youporn.com	porn	porn	porn	porn	not porn

p(porn)	p(not porn)
50.2%	49.8%
54.8%	45.2%
74.6%	25.4%
50.2%	49.8%
74.6%	25.4%

Iteration 0

url	p(porn)	p(not porn)
google.com	40%	**60%**
panda-cam.gov	**60%**	40%
sex-mission.com	**80%**	20%
sunnyfun.com	40%	**60%**
youporn.com	**80%**	20%

Guess

Truth

w1	porn	not
porn	100%	0%
not	0%	100%

w2	porn	not
porn	100%	0%
not	0%	100%

w3&4	porn	not
porn	100%	0%
not	0%	100%

w5	porn	not
porn	100%	0%
not	0%	100%

Iteration 1

url	p(porn)	p(not porn)
google.com	**50.2%**	49.8%
panda-cam.gov	**54.8%**	45.2%
sex-mission.com	**74.6%**	25.4%
sunnyfun.com	**50.2%**	49.8%
youporn.com	**74.6%**	25.4%

Guess

Truth

w1	porn	not
porn	**60%**	0
not	40%	0

w2	porn	not
porn	**73%**	50%
not	27%	50%

w3&4	porn	not
porn	**80%**	47%
not	20%	**53%**

w5	porn	not
porn	47%	**80%**
not	**53%**	20%

Iterate until you converge

url	p(porn)	p(not porn)
google.com	8.25%	**91.75%**
panda-cam.gov	25%	**75%**
sex-mission.com	**100%**	0%
sunnyfun.com	8.25%	**91.75%**
youporn.com	**100%**	0%

Guess

w1	porn	not
porn	100%	0%
not	100%	0%

w2	porn	not
porn	100%	0%
not	33%	67%

w3&4	porn	not
porn	100%	0%
not	0%	100%

w5	porn	not
porn	0%	100%
not	100%	0%

Truth

Question

- How would you use gold standard data in the EM process?

EM Algorithm

- Re-Calculate Worker Scores over two steps:

  1. Estimate the probability that each answer is correct, using labels from multiple workers weighted by the probability that they are correct

  2. Estimate the quality of the workers by comparing their submitted answers to the inferred correct answers

Confusion Matrix gives us worker error

- From the confusion matrix we can measure the overall error rate for each worker

- Sum of the non-diagonal elements of the confusion matrix (weighted by the priors)

- This results in a single, scalar value as the quality score for each worker

Worker Error

w1	porn	not
porn	100%	0%
not	100%	0%

100

w2	porn	not
porn	100%	0%
not	33%	67%

33

w3&4	porn	not
porn	100%	0%
not	0%	100%

0

w5	porn	not
porn	0%	100%
not	100%	0%

200

Is worker5 the worst?

url	worker1	worker2	worker3	worker4	worker5
google.com	porn	not porn	not porn	not porn	porn
panda-cam.gov	porn	porn	not porn	not porn	porn
sex-mission.com	porn	porn	porn	porn	not porn
sunnyfun.com	porn	not porn	not porn	not porn	porn
youporn.com	porn	porn	porn	porn	not porn

Advanced Topics

- Bias versus error

- How noisy can the workers be for us to still converge to a correct solution?

Bias versus Error

- Error rate alone is not sufficient to measure the inherent value of a worker.

- For example, workers may be careful but biased

- In a non-binary case, this is more apparent

- What if instead of asking our workers to label sites porn or not pron, we asked them to label the G, PG, R, X?

Bias versus Error

- Parents with young children tend to be more conservative

- They tend to classify PG-rated sites as R-rated sites, and R-rated sites as X-rated.

- Such workers give consistently and predictably incorrect answers

- It is possible to automatically correct for bias

Implications

- Unlike with spammers, with biased workers it is possible to "reverse" the errors

- We can recover a label assignment of much higher quality

- In the presence of systematic bias, the naive measurement of error rate results in underestimates of the true quality of the worker

- This potentially leads to incorrect rejections and blocks of legitimate workers

For more details

- Check out two papers by Panos Ipeirotis and his collaborators

- *Managing Crowdsourcing Workers* discusses separating error and bias

- *Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers* discusses how noisy judgements can be, with us still getting good quality results

Discussion time: Anyone need help coming up with ideas for their final project?