

# Quality Control - part 4

**Crowdsourcing and Human Computation**

**Instructor: Chris Callison-Burch**

**Website: [crowdsourcing-class.org](http://crowdsourcing-class.org)**

# Different Mechanisms for Quality Control

- Aggregation and redundancy
- Embedded gold standard data
- Economic incentives
- Reputation systems
- **Statistical models**

# Expectation Maximization algorithm

- EM is an algorithm for finding the probabilities of unobserved variables
- We will use it to estimate how accurate workers' labels are, and infer how good each worker 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 have 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
sunnyfun.com	porn	not	not	not	porn
sex-mission.com	porn	porn	porn	porn	porn
google.com	not	porn	not	not	porn
youporn.com	porn	porn	porn	porn	not
yahoo.com	porn	not	not	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  $o_1 \dots o_N$   
sunnyfun.com, sex-mission.com,  
google.com, youporn.com, yahoo.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  $T(O_n)$  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
sunnyfun.com	porn	not	not	not	porn
sex-mission.com	porn	porn	porn	porn	porn
google.com	not	porn	not	not	porn
youporn.com	porn	porn	porn	porn	not
yahoo.com	porn	not	not	not	porn

# Output: “True” Labels

url	True Labels
sunnyfun.com	not
sex-mission.com	porn
google.com	not
youporn.com	porn
yahoo.com	not



# Output: Worker Accuracies

		Guess	
		w1	not
Truth	porn	100%	0%
	not	67%	33%

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

		Guess	
		w3	not
Truth	porn	100%	0%
	not	0%	100%

		Guess	
		w4	not
Truth	porn	100%	0%
	not	0%	100%

		Guess	
		w5	not
Truth	porn	50%	50%
	not	100%	0%

# Initialize

## Guess

<b>Truth</b>	<b>w1</b>	porn	not
	porn	100%	0%
	not	0%	100%

<b>Truth</b>	<b>w2</b>	porn	not
	porn	100%	0%
	not	0%	100%

<b>Truth</b>	<b>w3</b>	porn	not
	porn	100%	0%
	not	0%	100%

## Guess

<b>Truth</b>	<b>w4</b>	porn	not
	porn	100%	0%
	not	0%	100%

<b>Truth</b>	<b>w5</b>	porn	not
	porn	100%	0%
	not	0%	100%

# Initialize with majority vote

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

porn	not porn
40%	60%
100%	0%
40%	60%
80%	20%
40%	60%

# Pretend your labels are correct

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

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

Re-Calculate  
Worker  
Scores

Guess

Truth

<b>w1</b>	porn	not
porn	0	0
not	1	0

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

# Re-Calculate Worker Scores

## Guess

	<b>w1</b>	porn	not
<b>Truth</b>	porn	1	0
	not	1	0

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

# Re-Calculate Worker Scores

## Guess

Truth

<b>w1</b>	porn	not
porn	1	0
not	1	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex-mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

# Re-Calculate Worker Scores

## Guess

	<b>w1</b>	porn	not
<b>Truth</b>	porn	2	0
	not	1	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1



# Re-Calculate Worker Scores

## Guess

	<b>w1</b>	porn	not
<b>Truth</b>	porn	2	0
	not	2	1

url	worker 1	worker 2	worker 3	worker 4	worker 5	porn	not porn
sunnyfun.com	porn	not	not	not	porn	0	1
sex- mission.com	porn	porn	porn	porn	porn	1	0
google.com	not	porn	not	not	porn	0	1
youporn.com	porn	porn	porn	porn	not	1	0
yahoo.com	porn	not	not	not	porn	0	1

# Re-Calculate Worker Scores

**Guess**

	<b>w1</b>	porn	not
<b>Truth</b>	porn	100%	0%
	not	67%	33%

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

porn	not porn
0	1
1	0
0	1
1	0
0	1

# Re-Calculate Worker Scores

## Guess

	<b>w2</b>	porn	not
<b>Truth</b>	porn	0	0
	not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

# Re-Calculate Worker Scores

## Guess

	<b>w2</b>	porn	not
<b>Truth</b>	porn	1	0
	not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex-mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

# Re-Calculate Worker Scores

## Guess

	<b>w2</b>	porn	not
<b>Truth</b>	porn	1	0
	not	1	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex-mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

# Re-Calculate Worker Scores

## Guess

	<b>w2</b>	porn	not
<b>Truth</b>	porn	2	0
	not	1	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex-mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

# Re-Calculate Worker Scores

## Guess

	<b>w2</b>	porn	not
<b>Truth</b>	porn	2	0
	not	1	2

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex-mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

# Re-Calculate Worker Scores

## Guess

	<b>w2</b>	porn	not
<b>Truth</b>	porn	100%	0%
	not	33%	67%

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

porn	not porn
0	1
1	0
0	1
1	0
0	1



# Re-Calculate Worker Scores

## Guess

Truth

w3	porn	not
porn	0	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex-mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

# Re-Calculate Worker Scores

## Guess

Truth

w3	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex-mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

# Re-Calculate Worker Scores

## Guess

Truth

w3	porn	not
porn	1	0
not	0	2

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex-mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

# Re-Calculate Worker Scores

## Guess

Truth

w3	porn	not
porn	2	0
not	0	2

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

# Re-Calculate Worker Scores

## Guess

Truth

w3	porn	not
porn	2	0
not	0	3

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

# Re-Calculate Worker Scores

## Guess

Truth

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

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

porn	not porn
0	1
1	0
0	1
1	0
0	1

# Re-Calculate Worker Scores

## Guess

Truth

w4	porn	not
porn	0	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

# Re-Calculate Worker Scores

## Guess

Truth

w4	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex-mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1



# Re-Calculate Worker Scores

## Guess

Truth

w4	porn	not
porn	1	0
not	0	2

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex-mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

# Re-Calculate Worker Scores

## Guess

Truth

<b>w4</b>	porn	not
porn	2	0
not	0	2

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

# Re-Calculate Worker Scores

## Guess

Truth

<b>w4</b>	porn	not
porn	2	0
not	0	3

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

# Re-Calculate Worker Scores

## Guess

Truth

<b>w4</b>	porn	not
porn	100%	0%
not	0%	100%

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

porn	not porn
0	1
1	0
0	1
1	0
0	1

# Re-Calculate Worker Scores

## Guess

Truth

<b>w5</b>	porn	not
porn	0	0
not	1	0

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		0	1
sex- mission.com	porn	porn	porn	porn	porn		1	0
google.com	not	porn	not	not	porn		0	1
youporn.com	porn	porn	porn	porn	not		1	0
yahoo.com	porn	not	not	not	porn		0	1

# Re-Calculate Worker Scores

## Guess

Truth

<b>w5</b>	porn	not
porn	1	0
not	1	0

url	worker 1	worker 2	worker 3	worker 4	worker 5	guess	truth
sunnyfun.com	porn	not	not	not	porn	0	1
sex-mission.com	porn	porn	porn	porn	porn	1	0
google.com	not	porn	not	not	porn	0	1
youporn.com	porn	porn	porn	porn	not	1	0
yahoo.com	porn	not	not	not	porn	0	1

# Re-Calculate Worker Scores

## Guess

Truth

<b>w5</b>	porn	not
porn	1	0
not	2	0

url	worker 1	worker 2	worker 3	worker 4	worker 5	guess	truth
sunnyfun.com	porn	not	not	not	porn	0	1
sex-mission.com	porn	porn	porn	porn	porn	1	0
google.com	not	porn	not	not	porn	0	1
youporn.com	porn	porn	porn	porn	not	1	0
yahoo.com	porn	not	not	not	porn	0	1

# Re-Calculate Worker Scores

## Guess

	<b>w5</b>	porn	not
<b>Truth</b>	porn	1	1
	not	2	0

url	worker 1	worker 2	worker 3	worker 4	worker 5	porn	not porn
sunnyfun.com	porn	not	not	not	porn	0	1
sex- mission.com	porn	porn	porn	porn	porn	1	0
google.com	not	porn	not	not	porn	0	1
youporn.com	porn	porn	porn	porn	not	1	0
yahoo.com	porn	not	not	not	porn	0	1



# Re-Calculate Worker Scores

## Guess

	<b>w5</b>	porn	not
<b>Truth</b>	porn	1	1
	not	3	0

url	worker 1	worker 2	worker 3	worker 4	worker 5	porn	not porn
sunnyfun.com	porn	not	not	not	porn	0	1
sex- mission.com	porn	porn	porn	porn	porn	1	0
google.com	not	porn	not	not	porn	0	1
youporn.com	porn	porn	porn	porn	not	1	0
yahoo.com	porn	not	not	not	porn	0	1

# Re-Calculate Worker Scores

## Guess

	<b>w5</b>	porn	not
<b>Truth</b>	porn	1	1
	not	3	0

url	worker 1	worker 2	worker 3	worker 4	worker 5	porn	not porn
sunnyfun.com	porn	not	not	not	porn	0	1
sex- mission.com	porn	porn	porn	porn	porn	1	0
google.com	not	porn	not	not	porn	0	1
youporn.com	porn	porn	porn	porn	not	1	0
yahoo.com	porn	not	not	not	porn	0	1

# Re-Calculate Worker Scores

## Guess

	<b>w5</b>	porn	not
<b>Truth</b>	porn	50%	50%
	not	100%	0%

url	worker 1	worker 2	worker 3	worker 4	worker 5	porn	not porn
sunnyfun.com	porn	not	not	not	porn	0	1
sex-mission.com	porn	porn	porn	porn	porn	1	0
google.com	not	porn	not	not	porn	0	1
youporn.com	porn	porn	porn	porn	not	1	0
yahoo.com	porn	not	not	not	porn	0	1

# Update Estimates for True Labels

		Guess	
		w1	w2
Truth	porn	1.00	0.00
	not	0.67	0.33

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		1.00	0.67
sex- mission.com	porn	porn	porn	porn	porn			
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

		Guess	
		w2	porn
Truth	porn	1.00	0.00
	not	0.33	0.67

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		1.00	1.34
sex- mission.com	porn	porn	porn	porn	porn			
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

## Guess

Truth

w3	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		1.00	2.34
sex- mission.com	porn	porn	porn	porn	porn			
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

## Guess

Truth

<b>w4</b>	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		1.00	3.34
sex- mission.com	porn	porn	porn	porn	porn			
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

		Guess		
		w5	porn	not
Truth	porn	0.5	0.5	
	not	1	0	

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		1.50	4.34
sex- mission.com	porn	porn	porn	porn	porn			
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			



# Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		1.50	4.34
sex- mission.com	porn	porn	porn	porn	porn			
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn			
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

		Guess	
		w1	porn
Truth	porn	1.00	0.00
	not	0.67	0.33

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex-mission.com	porn	porn	porn	porn	porn		1.00	0.67
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

		Guess	
		w2	porn
Truth	porn	1.00	0.00
	not	0.33	0.67

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		2.00	1.00
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

## Guess

Truth

w3	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex-mission.com	porn	porn	porn	porn	porn		3.00	1.00
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

## Guess

	<b>w4</b>	porn	not
<b>Truth</b>	porn	1	0
	not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		4.00	1.00
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

		Guess		
		w5	porn	not
Truth	porn	0.5	0.5	
	not	1	0	

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex-mission.com	porn	porn	porn	porn	porn		4.50	2.00
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		4.50	2.00
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			



# Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn			
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

		Guess	
		w1	porn
Truth	porn	1.00	0.00
	not	0.67	0.33

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex-mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		0.00	0.33
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

		Guess	
		w2	porn
Truth	porn	1.00	0.00
	not	0.33	0.67

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex-mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		1.00	0.67
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

## Guess

Truth

w3	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex-mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		1.00	1.67
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

## Guess

Truth

<b>w4</b>	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		1.00	2.67
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

		Guess	
		w5	porn
Truth	porn	0.5	0.5
	not	1	0

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex-mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		1.50	3.76
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		1.50	3.76
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not			
yahoo.com	porn	not	not	not	porn			



# Update Estimates for True Labels

		Guess	
		w1	porn
Truth	porn	1.00	0.00
	not	0.67	0.33

url	worker 1	worker 2	worker 3	worker 4	worker 5	porn	not porn
sunnyfun.com	porn	not	not	not	porn	26%	74%
sex-mission.com	porn	porn	porn	porn	porn	69%	31%
google.com	not	porn	not	not	porn	29%	71%
youporn.com	porn	porn	porn	porn	not	1.00	0.67
yahoo.com	porn	not	not	not	porn		

# Update Estimates for True Labels

		Guess	
		w2	porn
Truth	porn	1.00	0.00
	not	0.33	0.67

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex-mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		2.00	1.00
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

## Guess

Truth

w3	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5	porn	not porn
sunnyfun.com	porn	not	not	not	porn	26%	74%
sex-mission.com	porn	porn	porn	porn	porn	69%	31%
google.com	not	porn	not	not	porn	29%	71%
youporn.com	porn	porn	porn	porn	not	3.00	1.00
yahoo.com	porn	not	not	not	porn		

# Update Estimates for True Labels

## Guess

	<b>w4</b>	porn	not
<b>Truth</b>	porn	1	0
	not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		4.00	1.00
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

		Guess		
		w5	porn	not
Truth	porn	0.5	0.5	
	not	1	0	

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex-mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		4.50	1.00
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		4.50	1.00
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		82%	18%
yahoo.com	porn	not	not	not	porn			

# Update Estimates for True Labels

		Guess	
		w1	w2
Truth	porn	1.00	0.00
	not	0.67	0.33

url	worker	worker	worker	worker	worker	porn	not porn
	1	2	3	4	5		
sunnyfun.com	porn	not	not	not	porn	26%	74%
sex-mission.com	porn	porn	porn	porn	porn	69%	31%
google.com	not	porn	not	not	porn	29%	71%
youporn.com	porn	porn	porn	porn	not	82%	18%
yahoo.com	porn	not	not	not	porn	1.00	0.67



# Update Estimates for True Labels

		Guess		
		w2	porn	not
Truth	porn	1.00	0.00	
	not	0.33	0.67	

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex-mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		82%	18%
yahoo.com	porn	not	not	not	porn		1.00	1.34

# Update Estimates for True Labels

## Guess

Truth

w3	porn	not
porn	1	0
not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5	porn	not porn
sunnyfun.com	porn	not	not	not	porn	26%	74%
sex- mission.com	porn	porn	porn	porn	porn	69%	31%
google.com	not	porn	not	not	porn	29%	71%
youporn.com	porn	porn	porn	porn	not	82%	18%
yahoo.com	porn	not	not	not	porn	1.00	2.34

# Update Estimates for True Labels

## Guess

	<b>w4</b>	porn	not
<b>Truth</b>	porn	1	0
	not	0	1

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		82%	18%
yahoo.com	porn	not	not	not	porn		1.00	3.34

# Update Estimates for True Labels

		Guess		
		w5	porn	not
Truth	porn	0.5	0.5	
	not	1	0	

url	worker 1	worker 2	worker 3	worker 4	worker 5	porn	not porn
sunnyfun.com	porn	not	not	not	porn	26%	74%
sex-mission.com	porn	porn	porn	porn	porn	69%	31%
google.com	not	porn	not	not	porn	29%	71%
youporn.com	porn	porn	porn	porn	not	82%	18%
yahoo.com	porn	not	not	not	porn	1.50	4.34

# Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		82%	18%
yahoo.com	porn	not	not	not	porn		1.50	4.34

# Update Estimates for True Labels

url	worker 1	worker 2	worker 3	worker 4	worker 5		porn	not porn
sunnyfun.com	porn	not	not	not	porn		26%	74%
sex- mission.com	porn	porn	porn	porn	porn		69%	31%
google.com	not	porn	not	not	porn		29%	71%
youporn.com	porn	porn	porn	porn	not		82%	18%
yahoo.com	porn	not	not	not	porn		26%	74%

# Iteration 0

url	porn	not porn
sunnyfun.com	40%	60%
sex-mission.com	100%	0%
google.com	40%	60%
youporn.com	80%	20%
yahoo.com	40%	60%

## Guess

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

  

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

  

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

  

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

# Iteration 1

url	porn	not porn
sunnyfun.com	26%	74%
sex-mission.com	69%	31%
google.com	29%	71%
youporn.com	82%	18%
yahoo.com	26%	74%

		Guess	
		w1	
Truth	porn	100%	0%
	not	67%	33%
	w2		
		w2	
porn	100%	0%	
not	33%	67%	
w3&4			
		w3&4	
porn	100%	0%	
not	0%	100%	
w5			
		w5	
porn	50%	50%	
not	100%	0%	



# Repeat until convergence

- You can continue to iterate until your values converge
- For this example, we converge after the first iteration

# 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

<b>w1</b>	porn	not
porn	100%	0%
not	100%	0%

100

<b>w2</b>	porn	not
porn	100%	0%
not	33%	67%

33

<b>w3</b>	porn	not
porn	100%	0%
not	0%	100%

0

<b>w4</b>	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 and still allow 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 porn, 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