

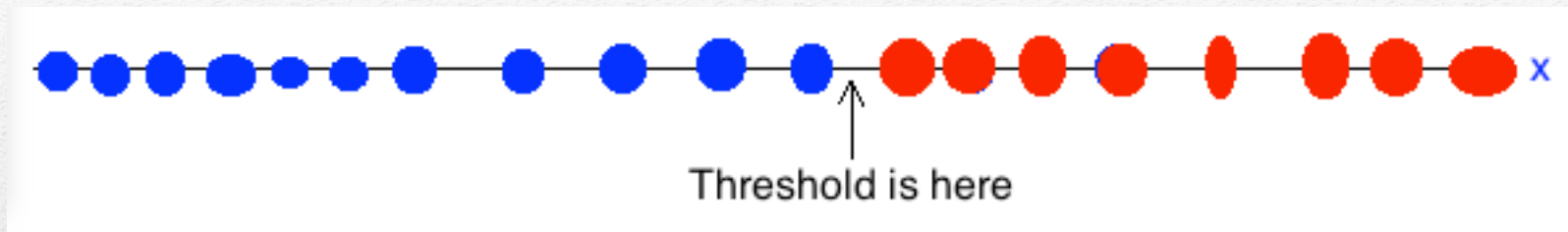
Active Learning with Label-Based Cost + Communication with Feedback

Gautam Ramasubramanian

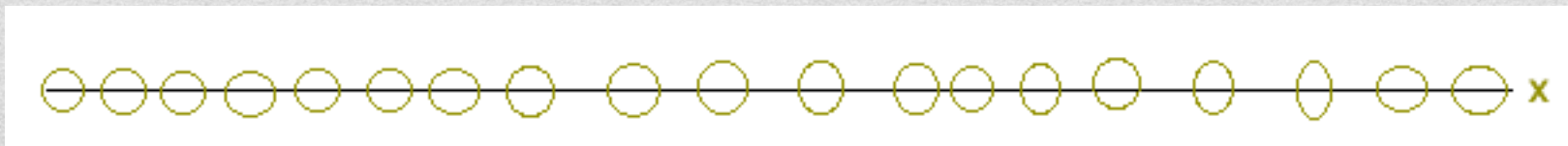


ACTIVE LEARNING

Supervised Learning vs. Active Learning

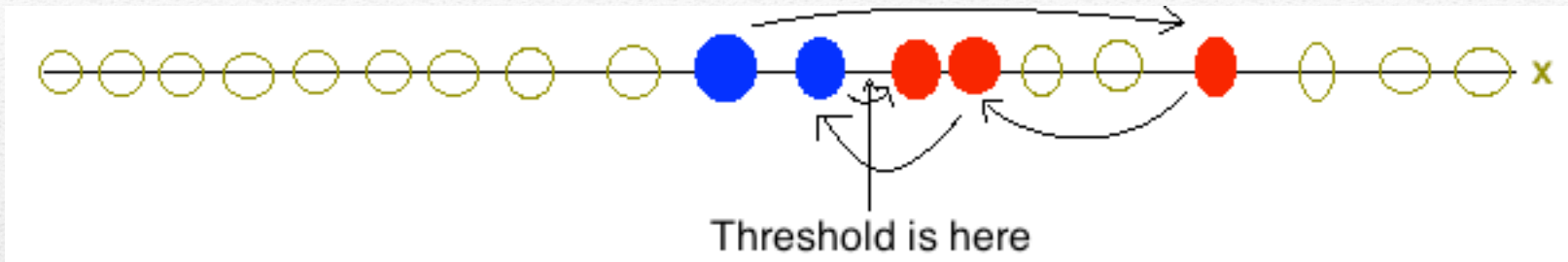


Supervised Learning Algorithm takes in training data and labels and produces a threshold, that can be used to predict labels for new data

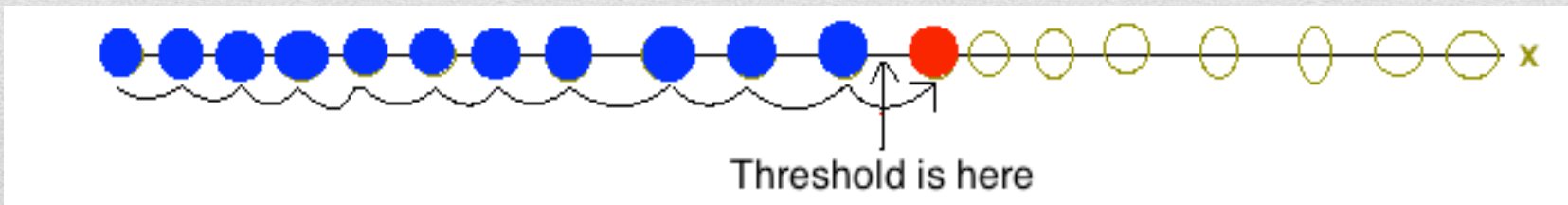


Active Learning finds ways to get that threshold when the training data is unlabeled and there is a cost associated to querying a label.

Active Learning Strategies



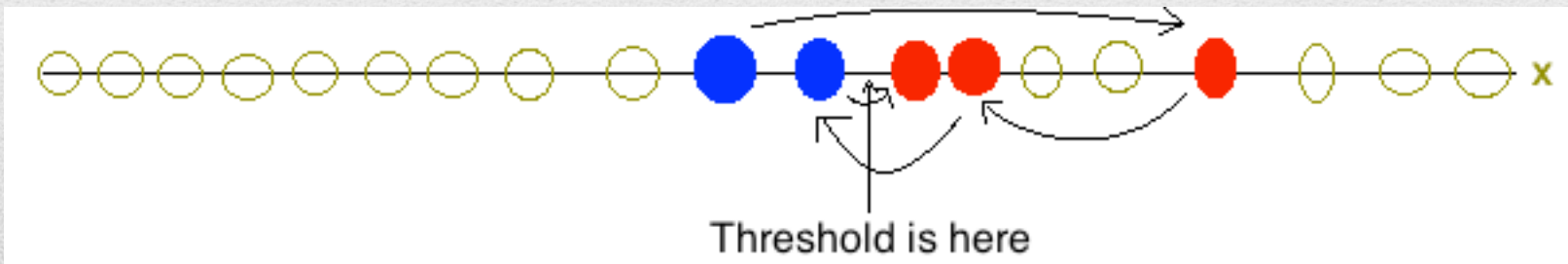
If the goal is to get the threshold in the minimum number of queries, then binary search is the best algorithm. The cost of querying is positive, and is the same regardless of the outcome (blue or red).



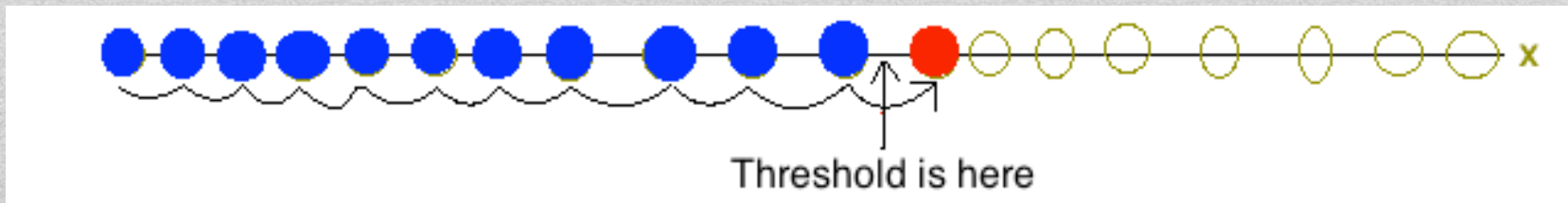
If the goal is to minimize the number of queries that come out red, we can do a linear search starting from the leftmost datum.

Label-Based Cost of Query

- Each Query has a certain cost, and that cost is dependent on the resulting label
- **Case #1** – $Cost(Blue) = Cost(Red) = C > 0$
 - *Best Strategy* – *Binary Search*



- **Case #2** – $Cost(Blue) = 0, Cost(Red) = C > 0$
 - *Best Strategy* – *Linear Search starting from Leftmost*



Label-Based Cost of Query

- **Case #3 (General)** – $Cost(Blue) = \alpha > 0$, $Cost(Red) = \beta > 0$
 - *Best Strategy* – ????

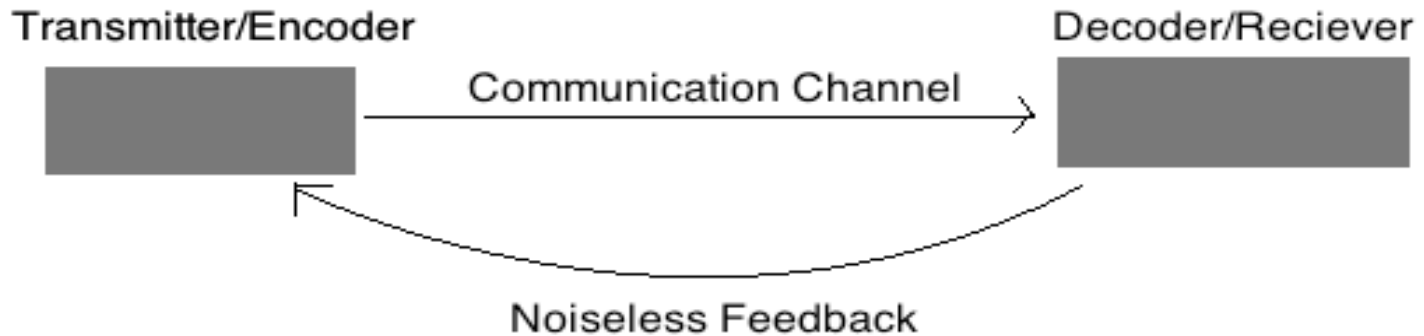


There are good guesses as to what the strategy could be, but we have to provide this game with an appropriate mathematical model and then use that model to definitively prove the best strategy.

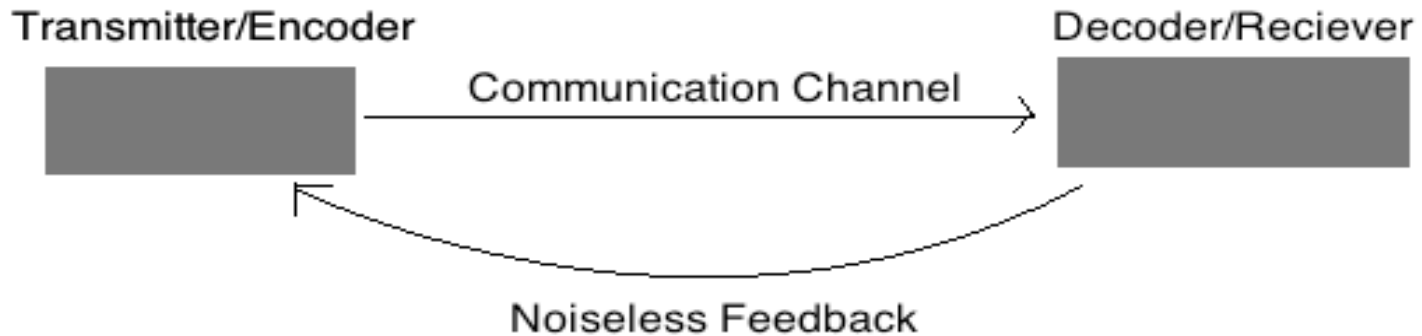


COMMUNICATION WITH FEEDBACK

Communication with Feedback



Communication with Feedback

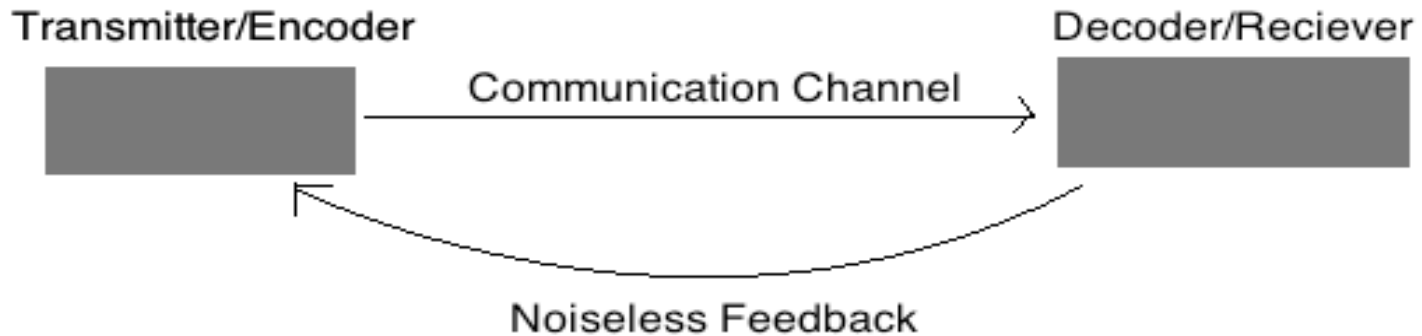


10

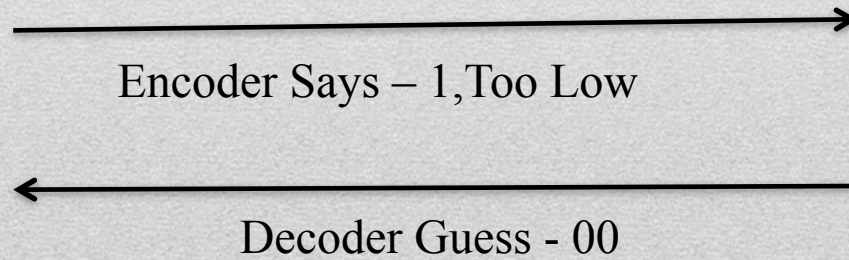


00
01
10
11

Communication with Feedback



10



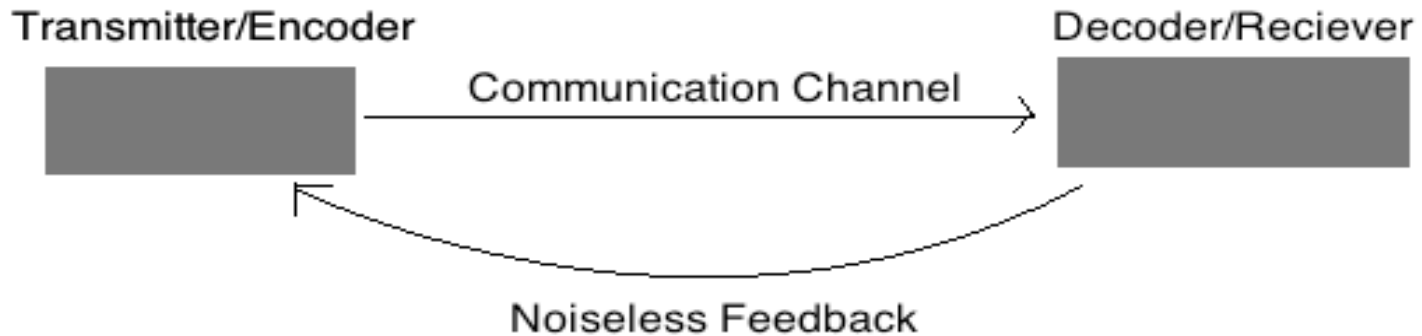
00

01

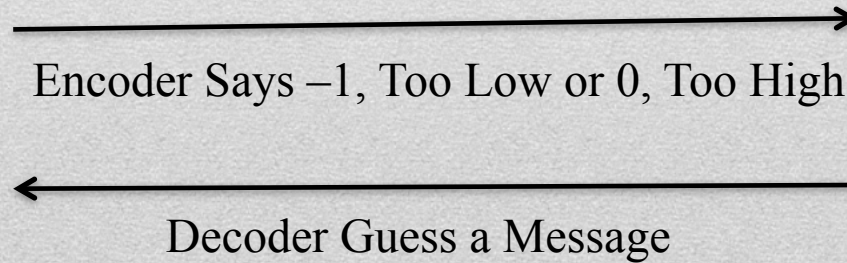
10

11

Communication with Feedback



10



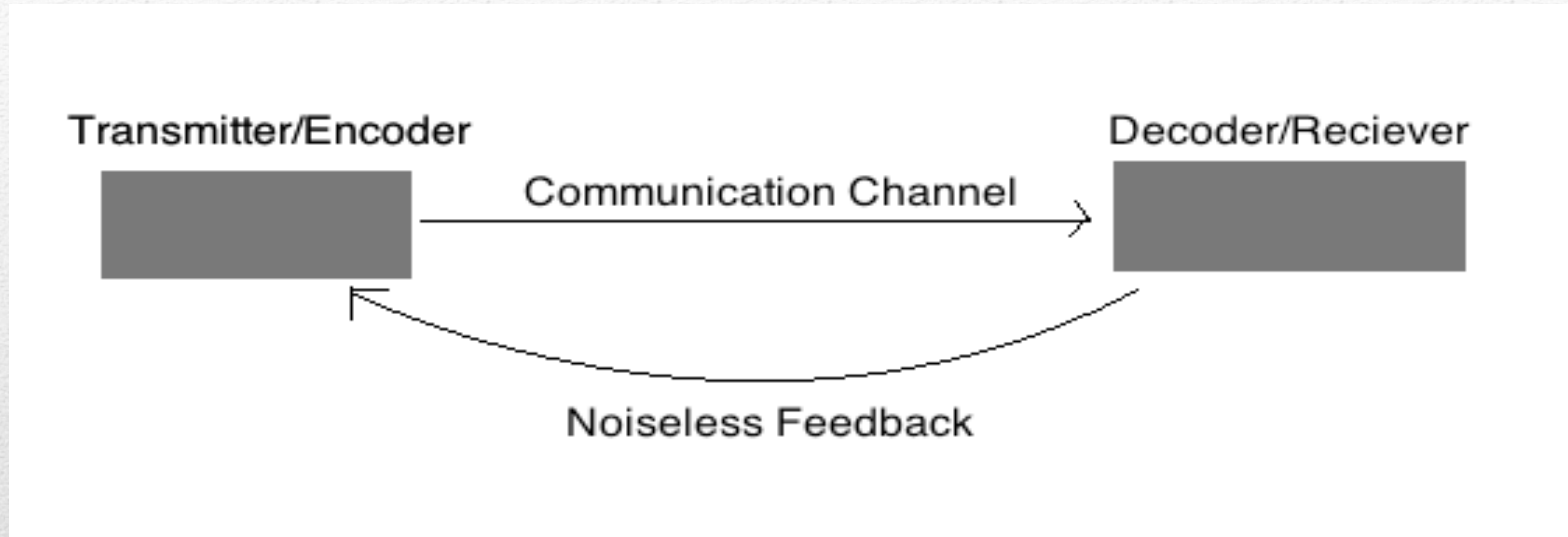
00

01

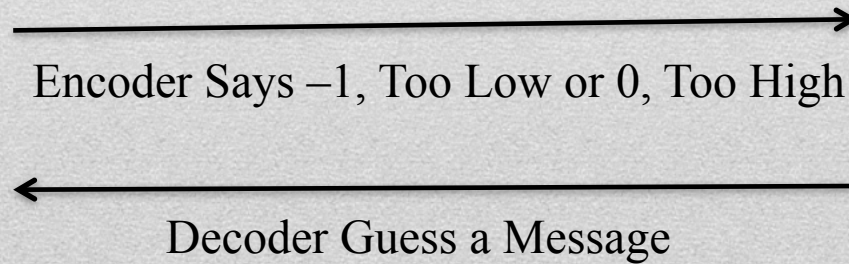
10

11

Communication with Feedback



10

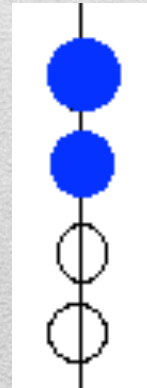


00

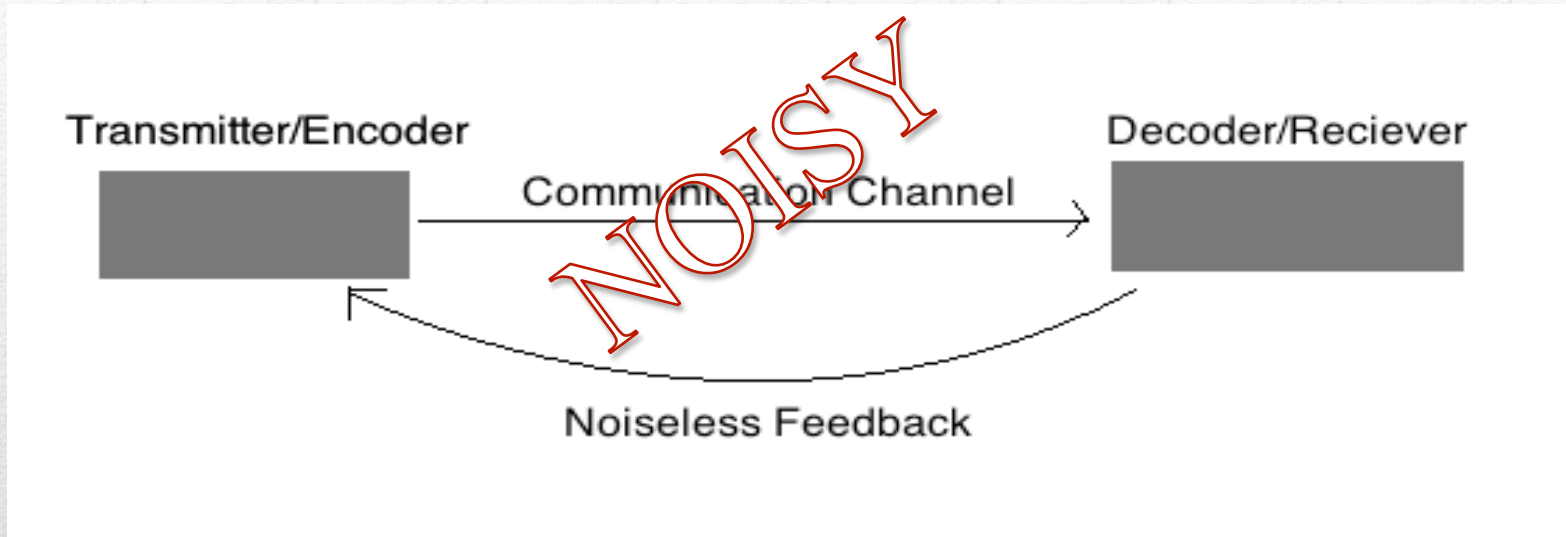
01

10

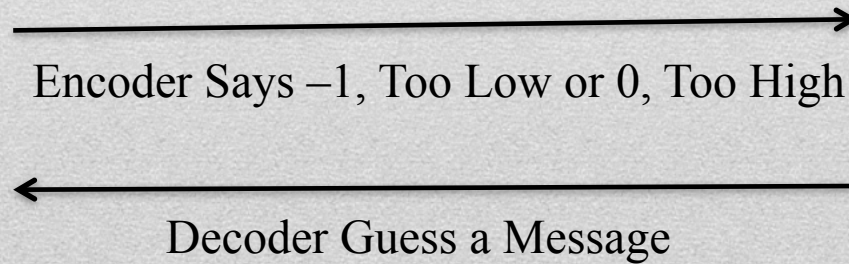
11



Communication with Feedback



10

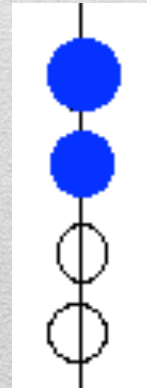


00

01

10

11



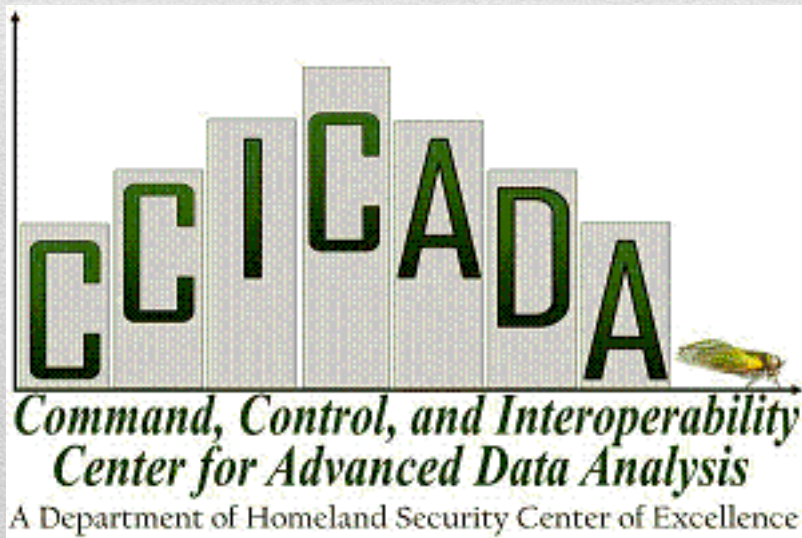
To Summarize

- We Are Analyzing the Number Guessing Game

```
Please enter a low number: 1
Please enter a high number: 10
Guess a number between 1 and 10
7
Too High
4 guesses left.
Guess a number between 1 and 7
4
Too High
3 guesses left.
Guess a number between 1 and 4
2
Too Low
2 guesses left.
Guess a number between 2 and 4
3
Correct
Try again? y/n:
```


Acknowledgements

- My Mentor
 - Professor Anand Sarwate
- CCICADA
- NSF and DIMACS
 - Professor Lazaros Gallos, Erik Amorim



DIMACS

References (for Pictures)

[1] *Shannon and Weaver Model of Communication*. [Online]. Available: <http://communicationtheory.org/shannon-and-weaver-model-of-communication/>

[2] “SVM-HUSTLE—an iterative semi-supervised machine learning approach for pairwise protein remote homology detection” â€ *Bioinformatics*. [Online]. Available: <http://bioinformatics.oxfordjournals.org/content/24/6/783/F1.expansion.html>

[3] *Computer Science*. [Online]. Available: <http://tech.sasaustin.org/compsci/Students/Projects/GuessingGame/GuessingGame.html>
