Proofs of Learning

John Langford

What is learning?

Learning = ability to predict

proof of Learning?

1. = Evidence of learning

2. Impossible in general

- Proof by assumed limited problem space
- Proof by assumption of independence ($\leftarrow$ us)

The Model

- (unknown) distribution D

- Draw all samples (x, y) independently from D

Proofs of learning

1. Holdout: Used in practice

2. Training: Not used, often loose

3. Combined: Best of both

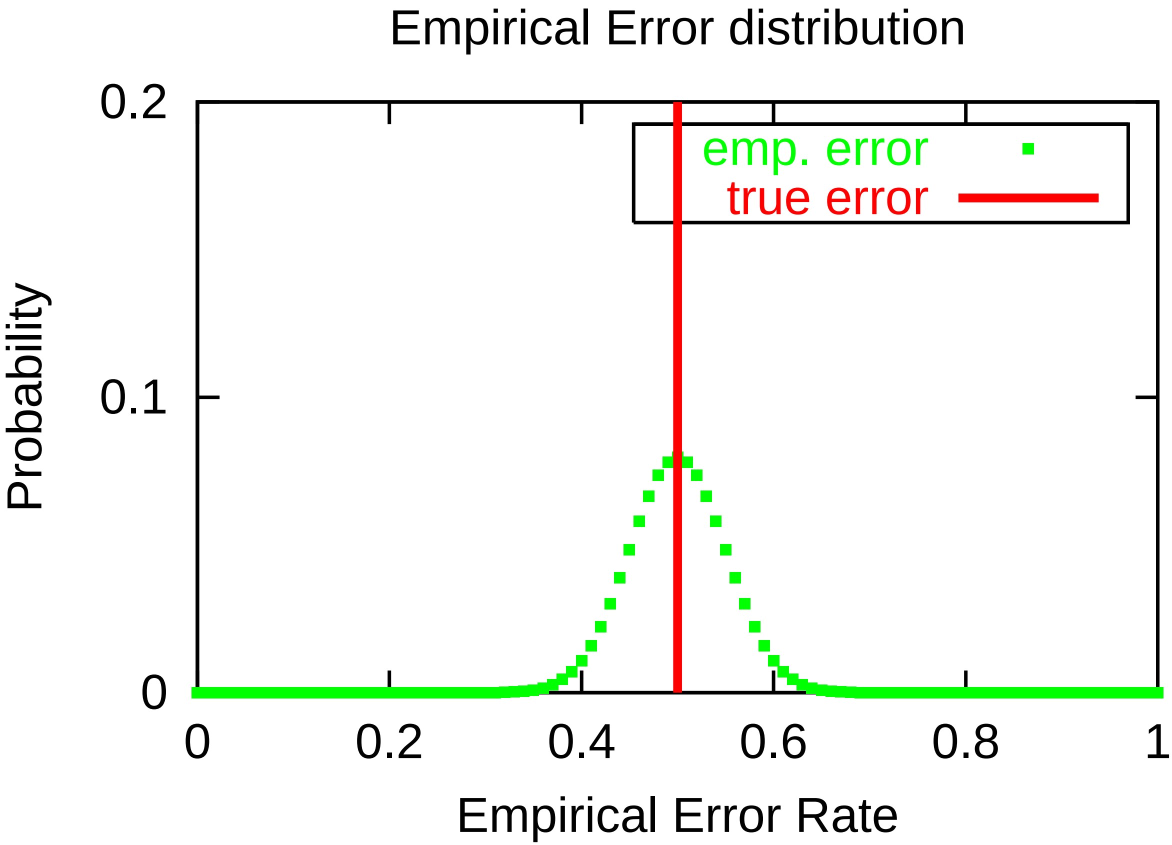
Terminology

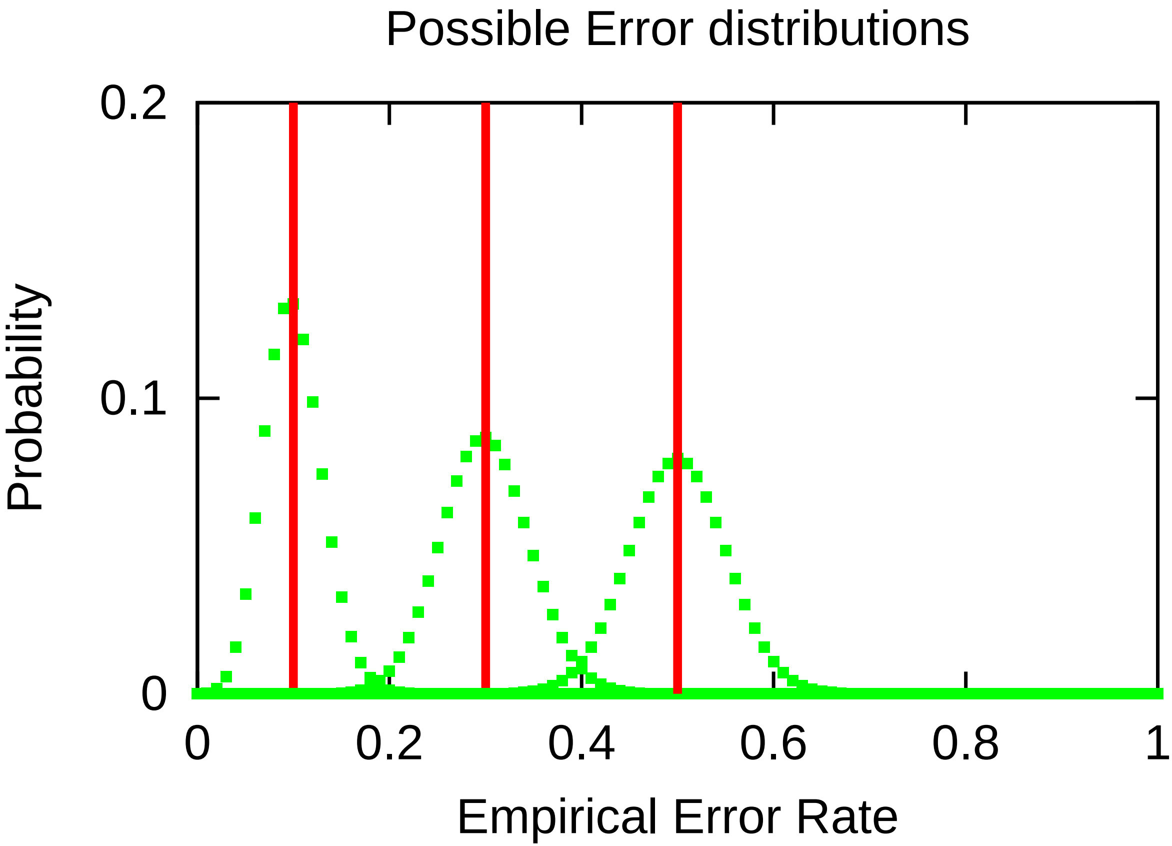
- Samples $= S$

- Learning algorithm $A: \; S \to h$

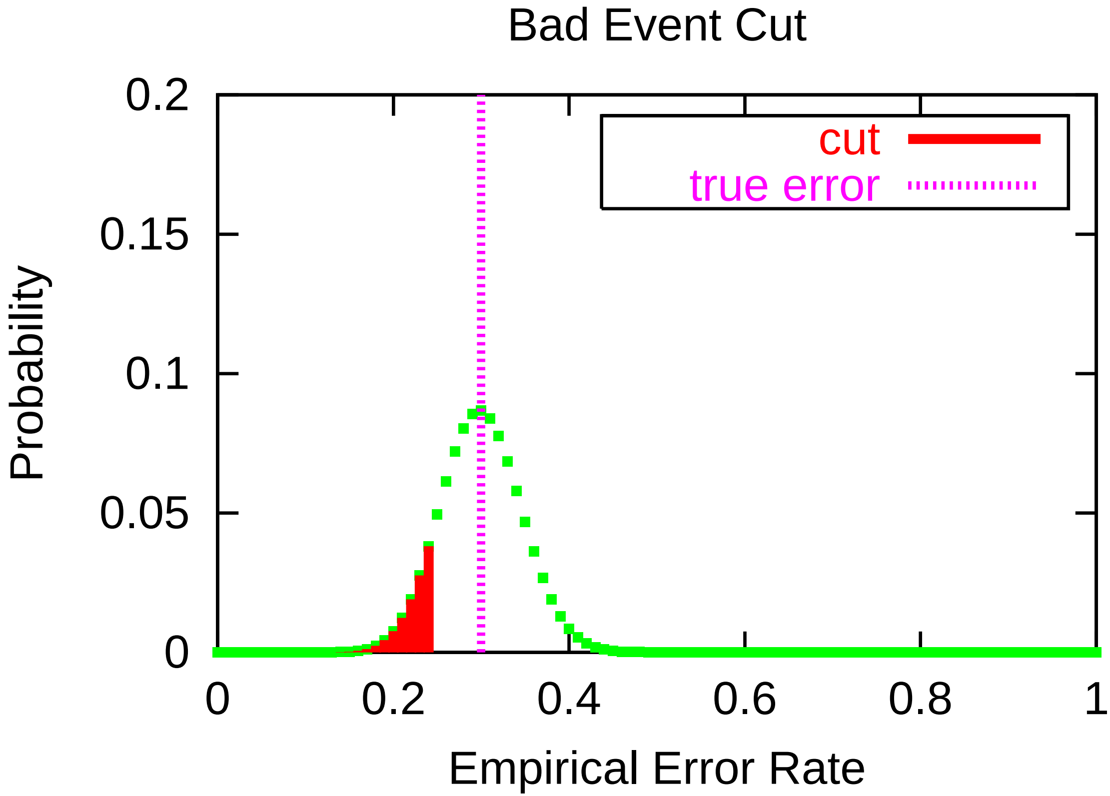
- True Error rate $e(h) = \mathsf{Pr}_D(h(x) \neq y)$

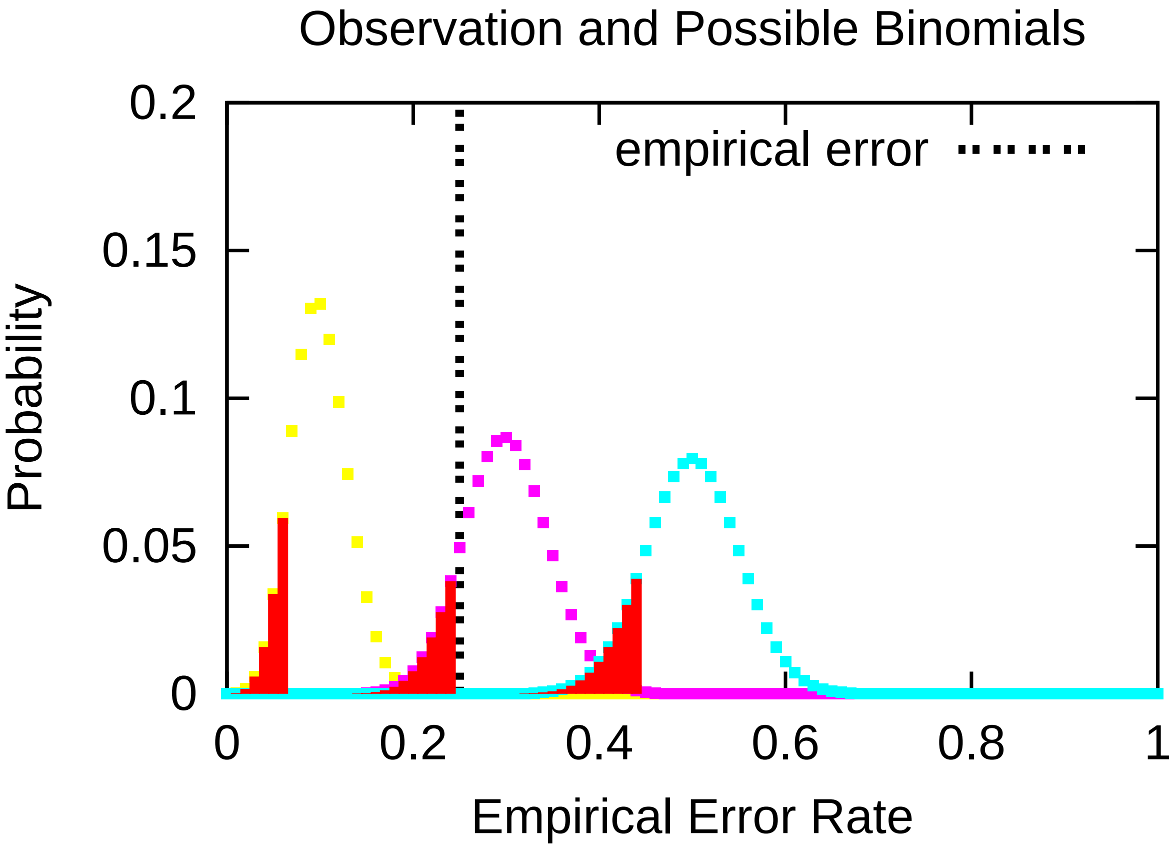
- Empirical Error rate $\widehat{e}(h) = \mathsf{Pr}_S(h(x) \neq y)$


Empirical Error distribution
emp. error
true error
Probability
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
Empirical Error Rate

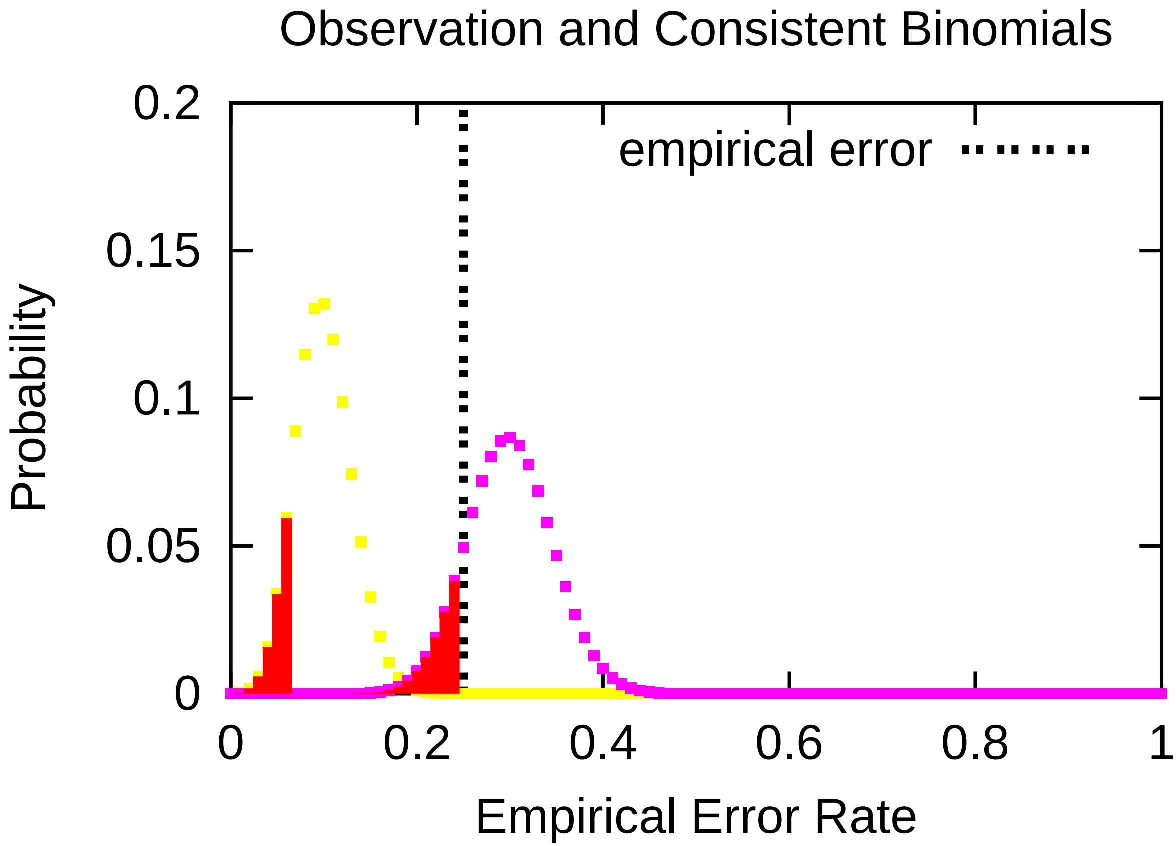

Possible Error distributions
Probability
Empirical Error Rate
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1

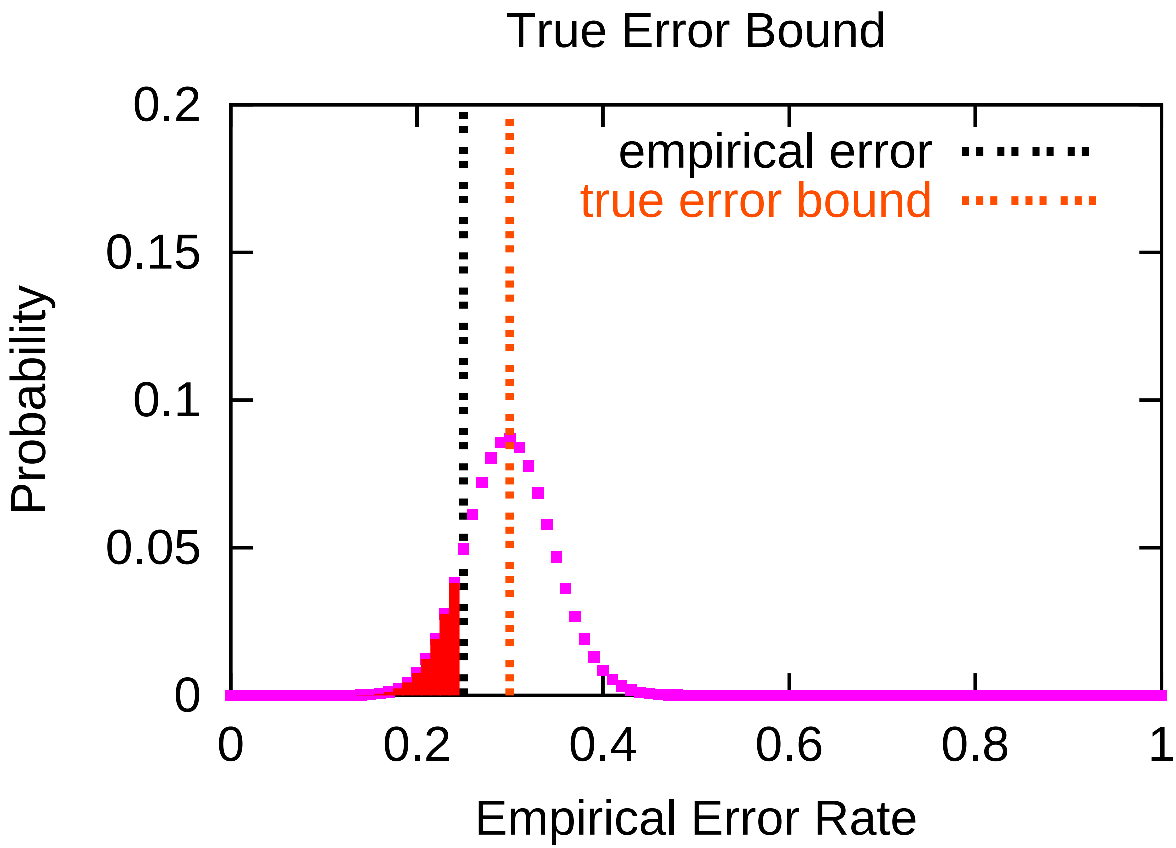
How can you prove learning?





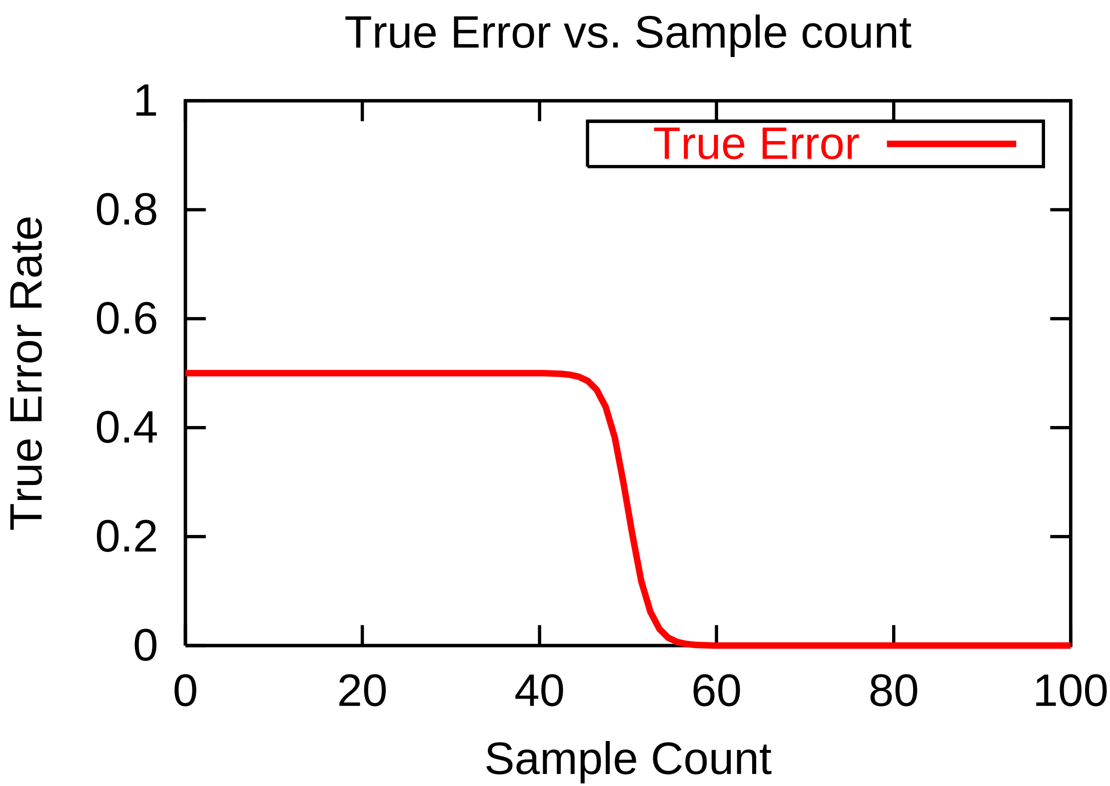
Observation and Possible Binomials
empirical error
Probability
0.2
0.15
0.1
0.05
0
0
0.2
0.4
0.6
0.8
1
Empirical Error Rate


Observation and Consistent Binomials
empirical error
Probability
Empirical Error Rate
0
0.05
0.1
0.15
0.2
0
0.2
0.4
0.6
0.8
1


True Error Bound
empirical error
true error bound
Probability
Empirical Error Rate
0.2
0.15
0.1
0.05
0
0
0.2
0.4
0.6
0.8
1

Holdout bound = wasted examples

$\Rightarrow$ No learning

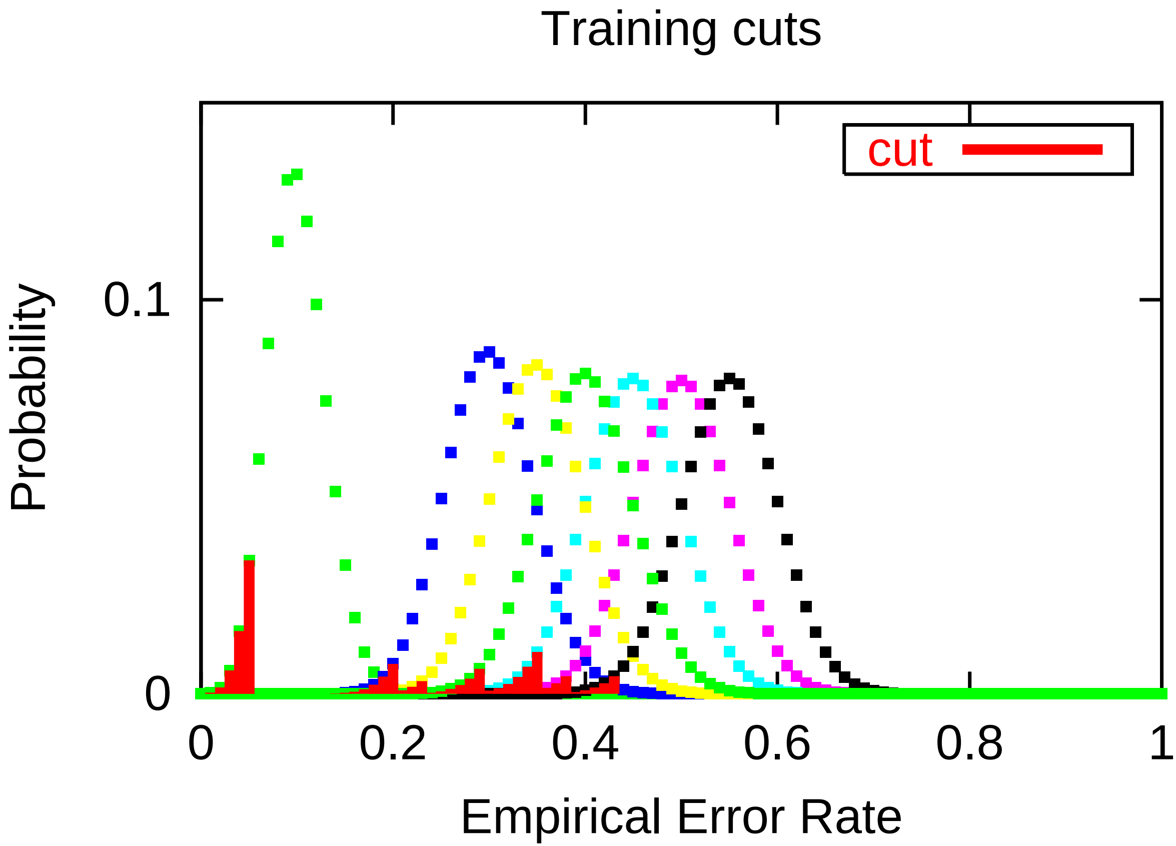


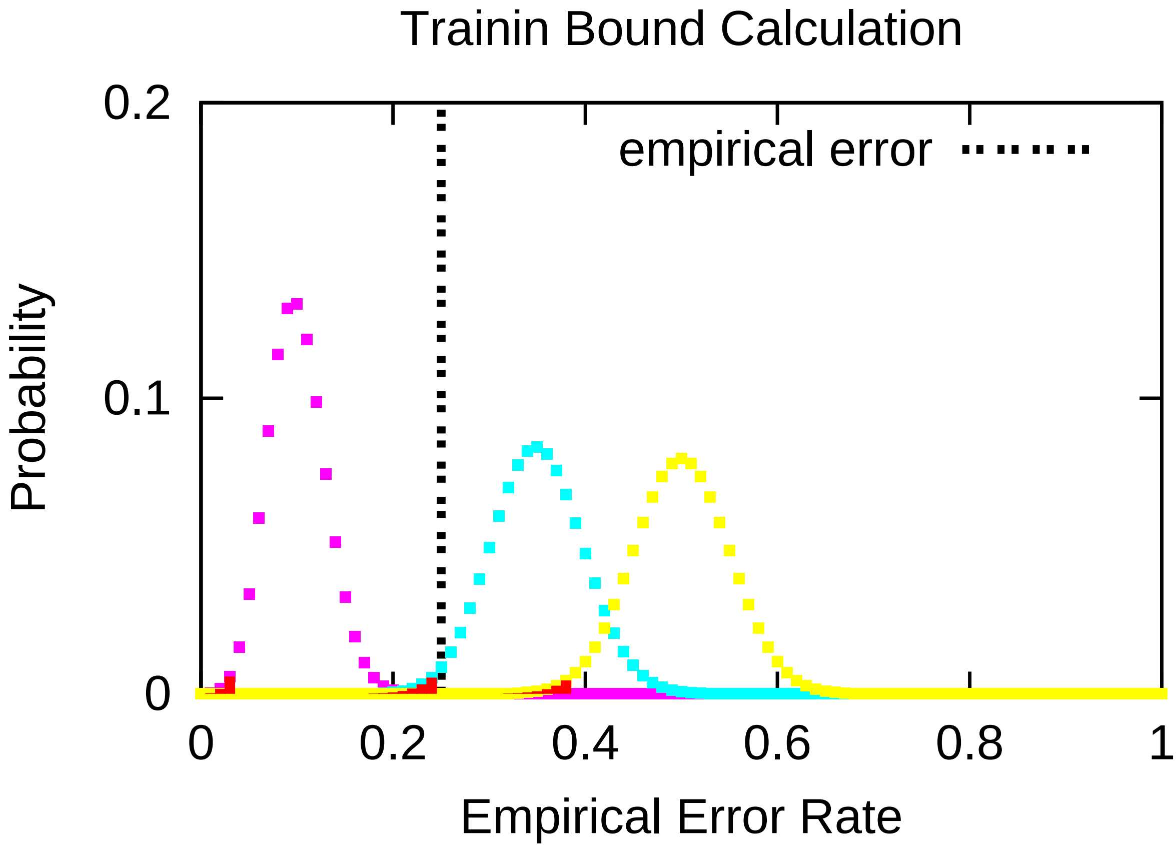
Proofs of learning

1. Holdout: Used in practice

2. Training: Not used, often loose

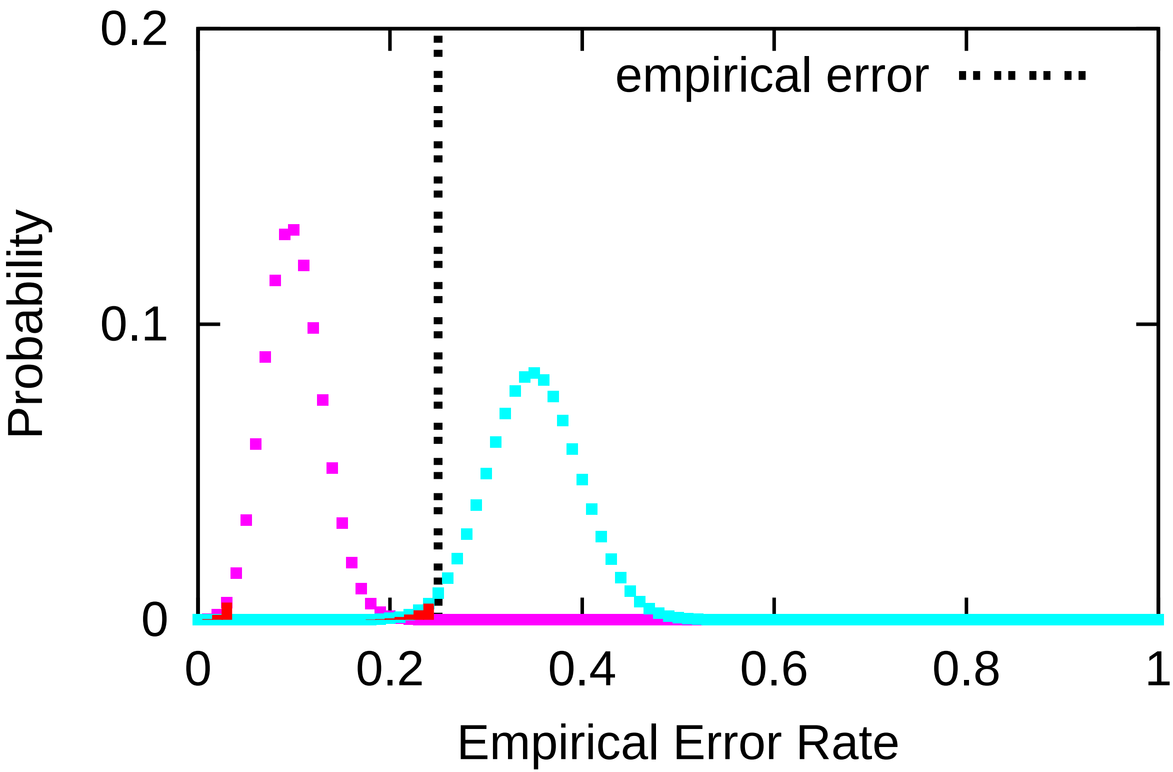
3. Combined: Best of both

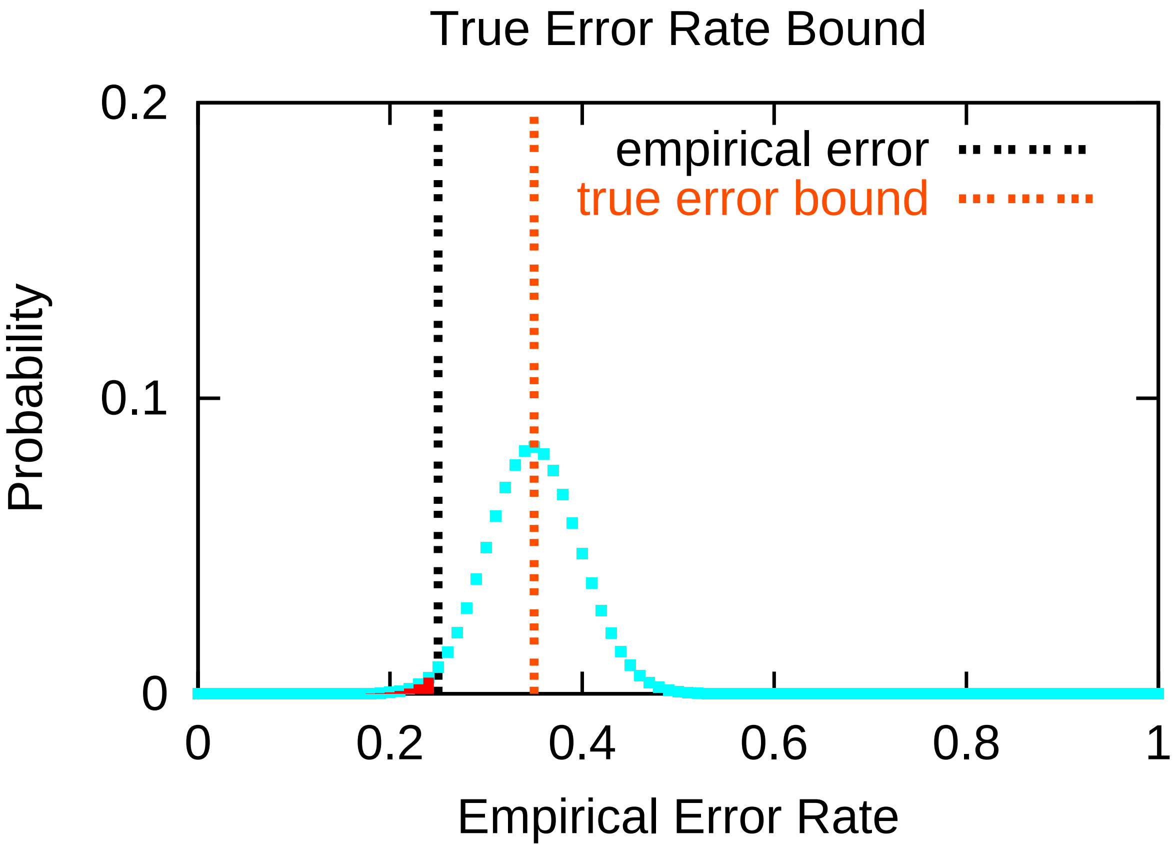

Training cuts
cut
Probability
Empirical Error Rate
0
0.1
0
0.2
0.4
0.6
0.8
1



Trainin Bound Calculation
empirical error
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1
Probability
Empirical Error Rate

Consistent Error Rates





True Error Rate Bound
empirical error
true error bound
Probability
Empirical Error Rate
0.2
0.1
0
0
0.2
0.4
0.6
0.8
1

Training A&D

- + Much theoretical work
- - Often too loose
- + Use all examples in training

Can we combine?

Proofs of learning

1. Holdout: Used in practice
2. Training: Not used, often loose
3. Combined: Best of both

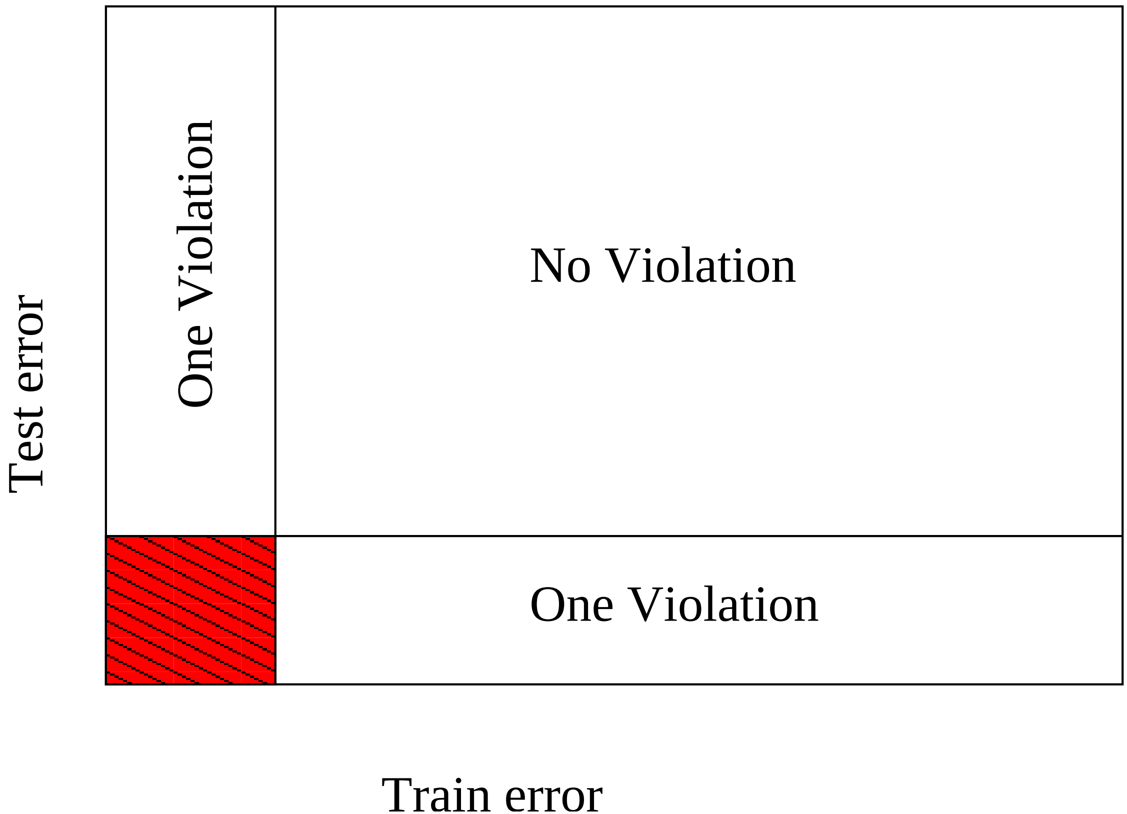
Combination techniques

1. Conjunction

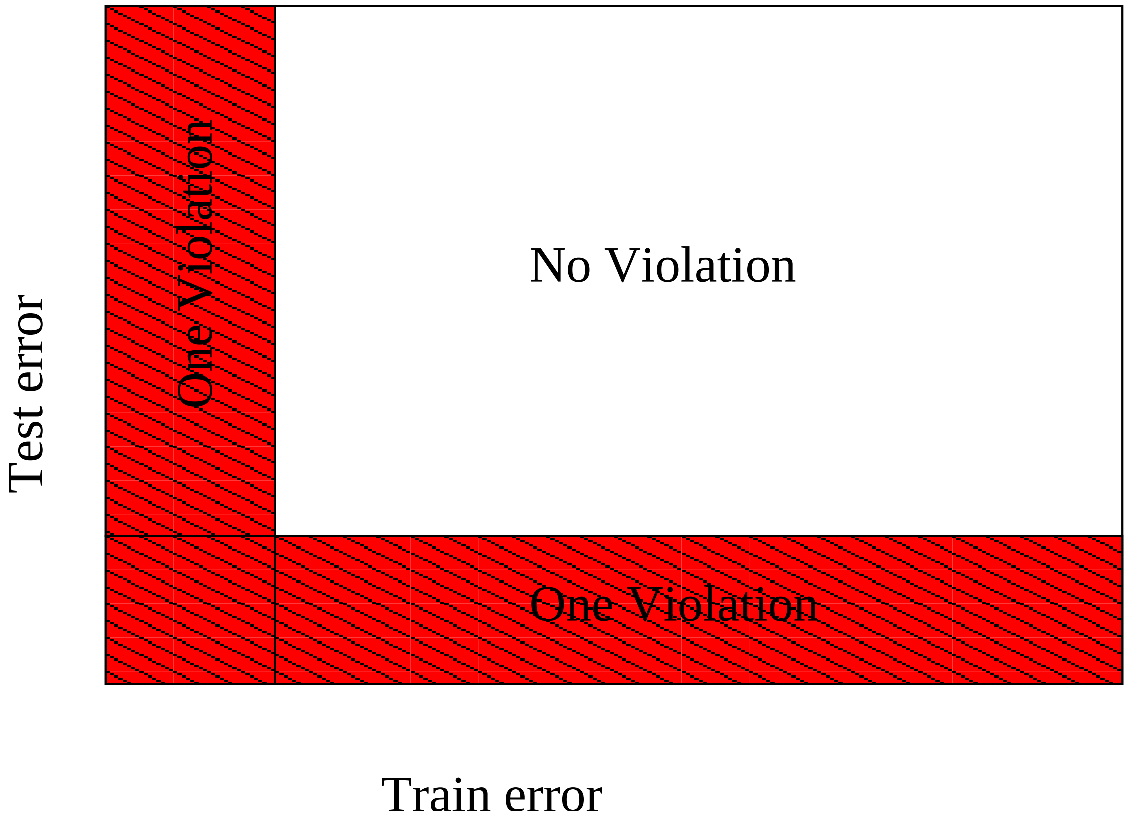
2. Disjunction

3. More general?

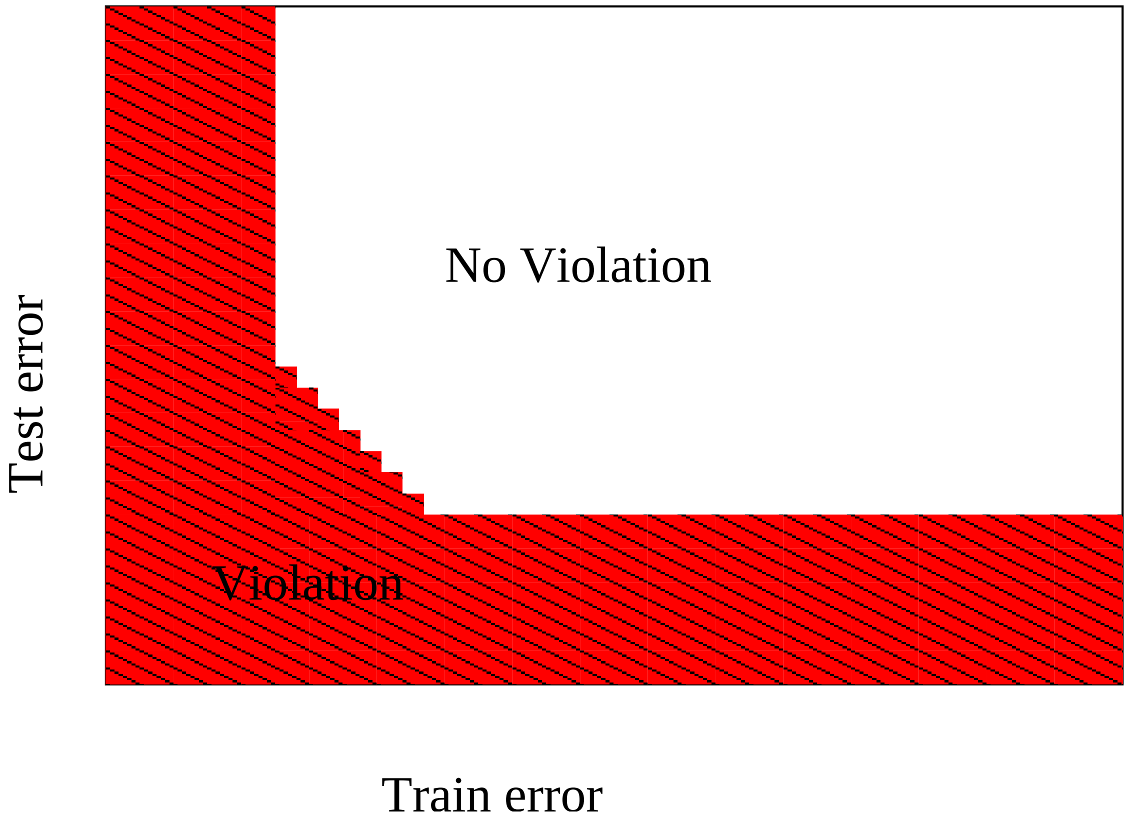
Conjunction



Disjunction



General

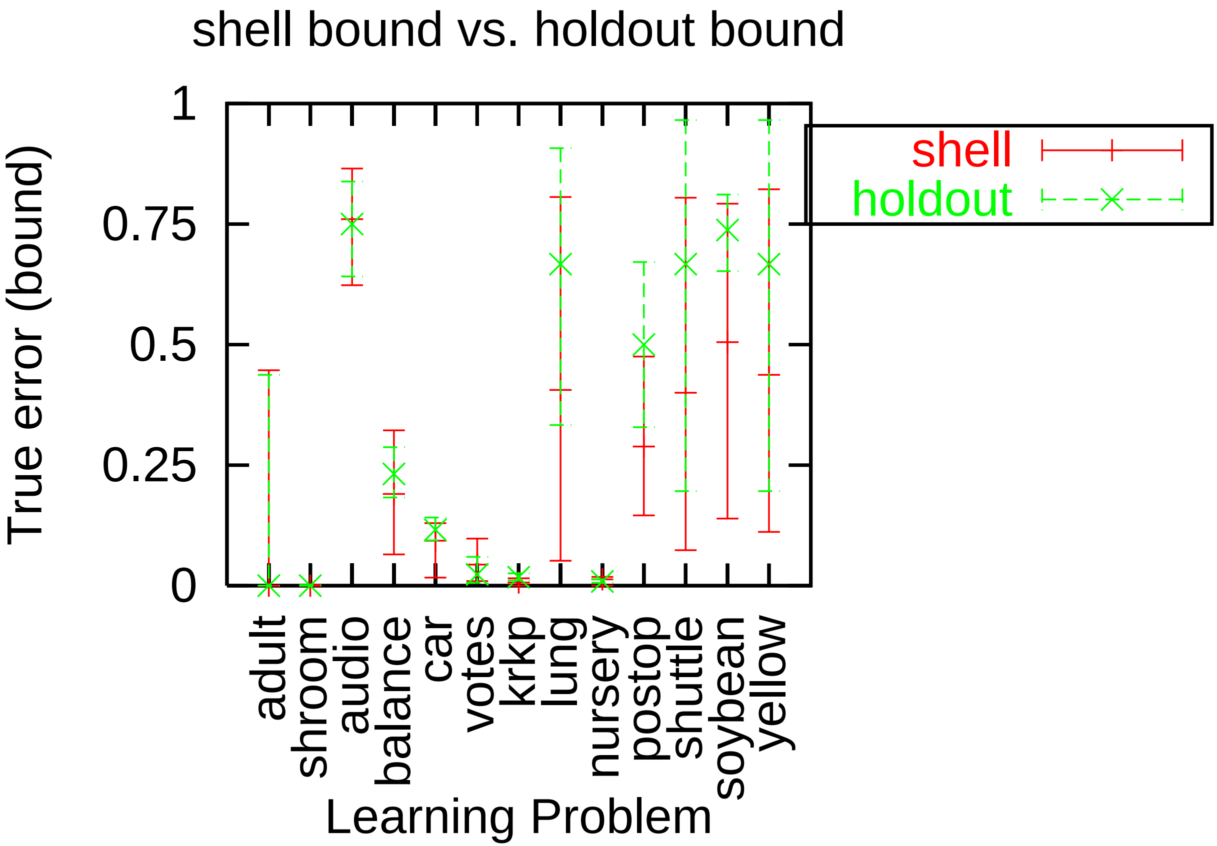


Combined bound implications

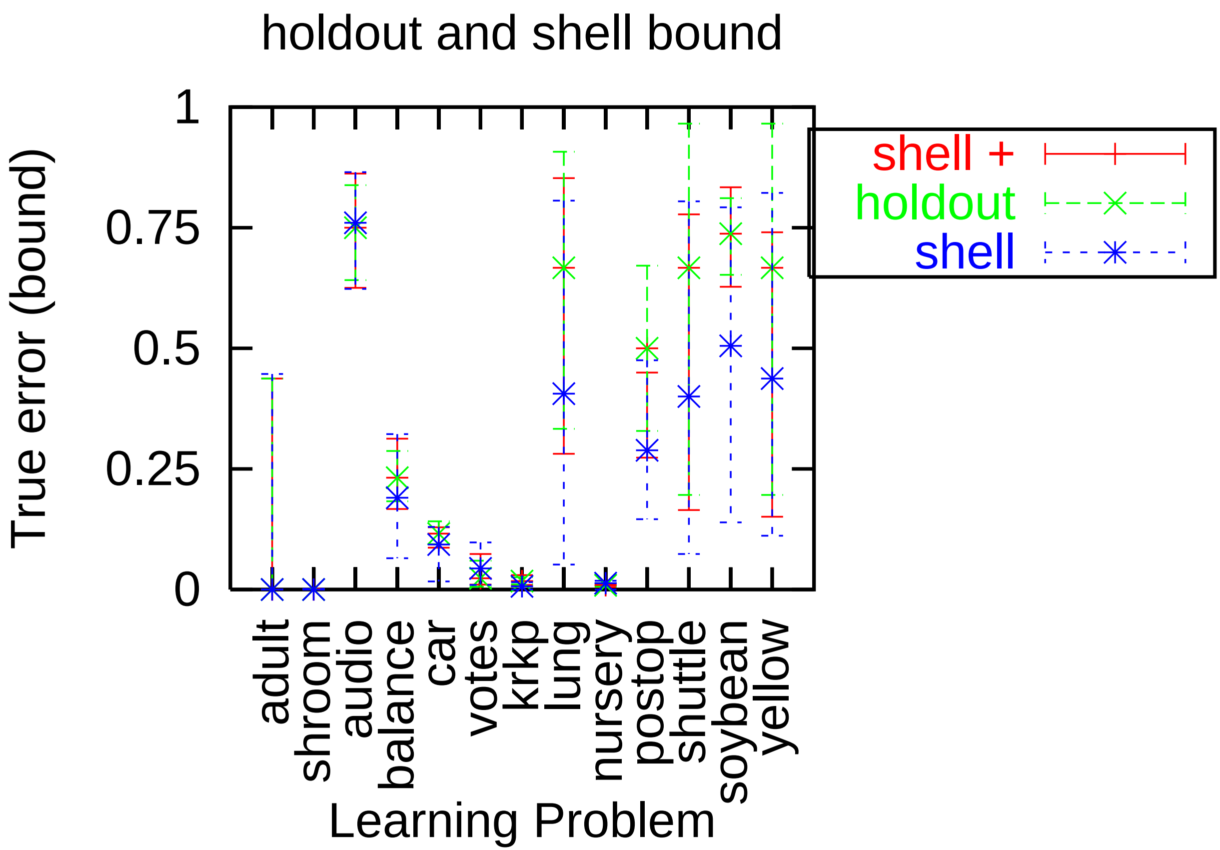
1. Use learning theory and standard practice

2. Often tighter bounds

Results: Test vs. Train



Results: Test vs. Combined



Future work

Proofs of learning for:

1. Cross validation

2. Continuous classifiers