To Re(label), or Not To Re(label)

Christopher H. Lin

University of Washington

Mausam

IIT Delhi

Daniel S. Weld

University of Washington

Penguin

Bear

Giraffe

Triceratops

Bear

Giraffe

Penguin

Triceratops

Penguin

Penguin

~~Triceratops~~

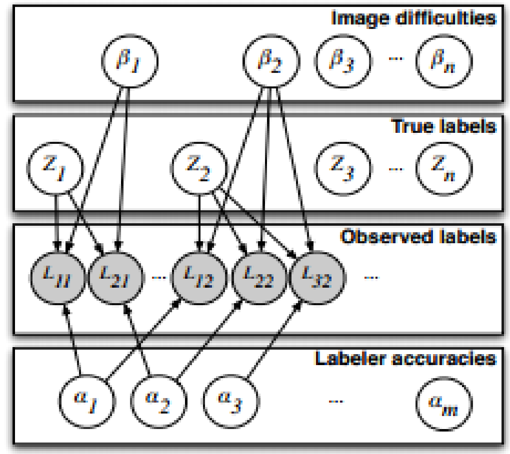
Penguin

Penguin

$$U(p(\beta_r|y_{i',l'})) = \left\| E_{p(\beta_r)}(\beta_r) - E_{p(\beta_r|y_{i',l'})}(\beta_r) \right\|_2 \tag{12}$$

$$\approx \left\| E\left(\frac{1}{S-1} \sum_{s=2}^{S} {Z_r^{s-1}}^{\top} \left[(\gamma|\gamma^{s-1}, Z^{s-1}) - (\gamma_{(i',l')}|\gamma^{s-1}, Z^{s-1}) \right] \right) \right\|_2 . \tag{13}$$

$$\begin{aligned} Q(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= E\left[\ln p(\mathbf{l}, \mathbf{z}|\boldsymbol{\alpha}, \boldsymbol{\beta})\right] \\ &= E\left[\ln \prod_j \left(p(z_j) \prod_i p(l_{ij}|z_j, \alpha_i, \beta_j) \right)\right] \\ & \quad \text{since } l_{ij} \text{ are cond. indep. given } \mathbf{z}, \boldsymbol{\alpha}, \boldsymbol{\beta} \\ &= \sum_j E\left[\ln p(z_j)\right] + \sum_{ij} E\left[\ln p(l_{ij}|z_j, \alpha_i, \beta_j)\right] \end{aligned}$$



$$p(z|L,\theta) = \int p(z,q|L,\theta)dq = \prod_{j\in[M]} \int_0^1 p(q_j|\theta) q_j^{c_j} (1-q_j)^{\gamma_j - c_j} dq_j \stackrel{def}{=} \prod_{j\in[M]} \psi_j(z_{\mathcal{N}_j}), \tag{4}$$

[Dawid et al 79, Whitehill et al 09, Welinder et al 10, Raykar et al 10, Wauthier et al 11, Karger et al 11, Kajino et al 12, Baba et al 13, Liu et al 12, Lin et al 12, etc, etc…]

How should we best spend a fixed budget *b*
when training a classifier?

budget b = 9

Unilabel?









9 examples with labels that are 75% accurate?

3 examples with labels that are 85% accurate?

1 example with a label that is 99% accurate?

How should we best spend a fixed budget *b*
when training a classifier?

How to relabel

[Donmez et al 2008, 2009, 2010
Yan et al 2011,
Dekel et al 2010,
Sheng et al 2008, Ipeirotis et al 2013,
Zhao et al 2011]

Dataset	# Features	# Examples
(a) Breast Cancer	9	699
(b) Bank Note Authentication	4	1372
(c) Seismic Bumps	18	2584
(d) EEG Eye State	14	14980
(e) Sonar	60	208
(f) Breast Cancer Diagnostic	30	569
(g) Hill-Valley	100	606
(h) Hill-Valley with Noise	100	606
(i) Internet Ads	1558	2359
(j) Gisette	5000	6000
(k) Farm Ads	54877	4143
(l) Spambase	57	4601

Unilabeling better in **over half** the datasets!!

Factors that Affect Relabeling Efficacy

Inductive Bias

Worker Accuracy

Budget

Assumptions

Passive Learning

Binary Classification

Identical Workers

Constant Cost

Majority Vote

j/k Relabeling

Penguin   Penguin   Triceratops

2/3 Relabeling

Penguin

Penguin

Factors that Affect Relabeling Efficacy

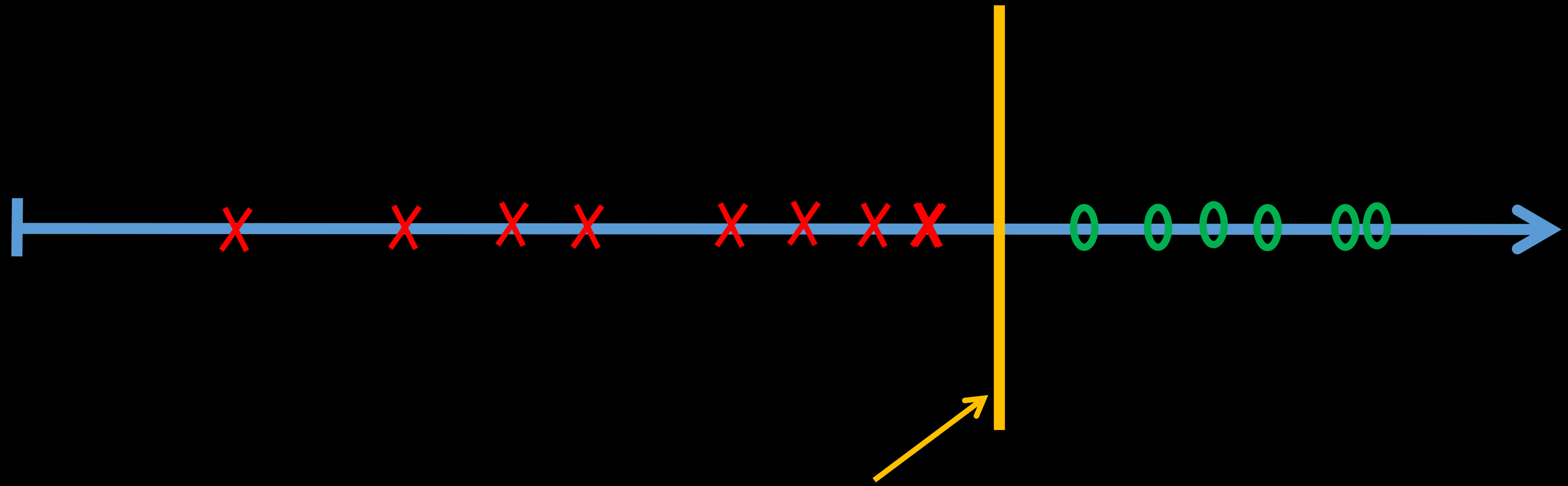
Inductive Bias

Worker Accuracy

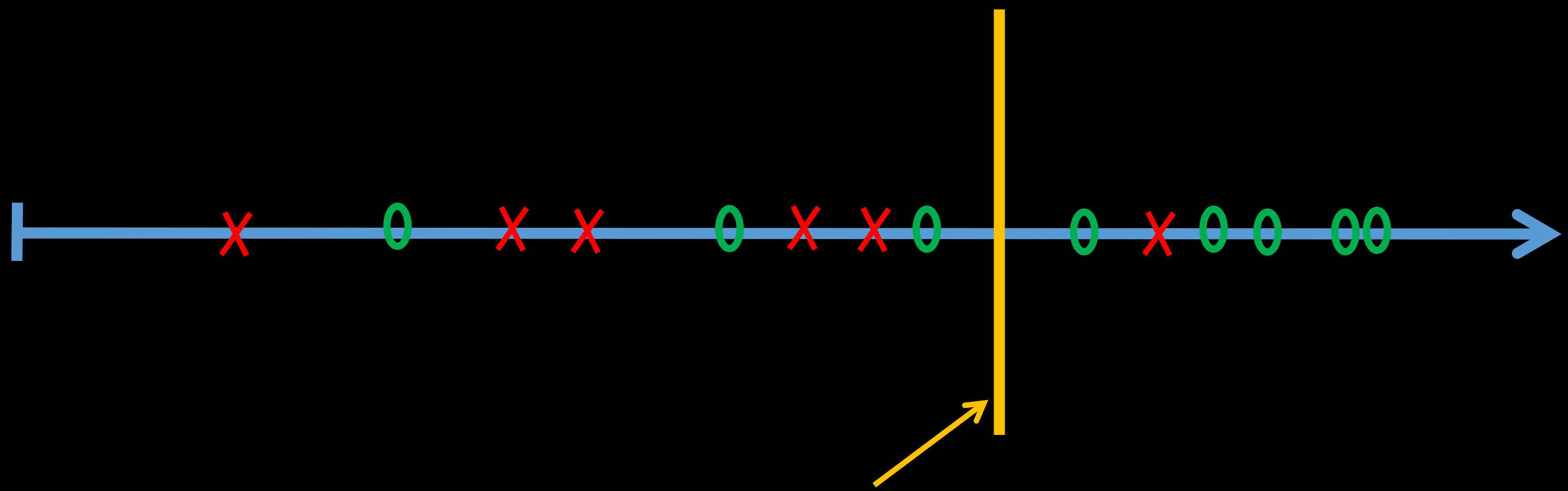
Budget

strong inductive bias → low expressiveness

weak inductive bias → high expressiveness



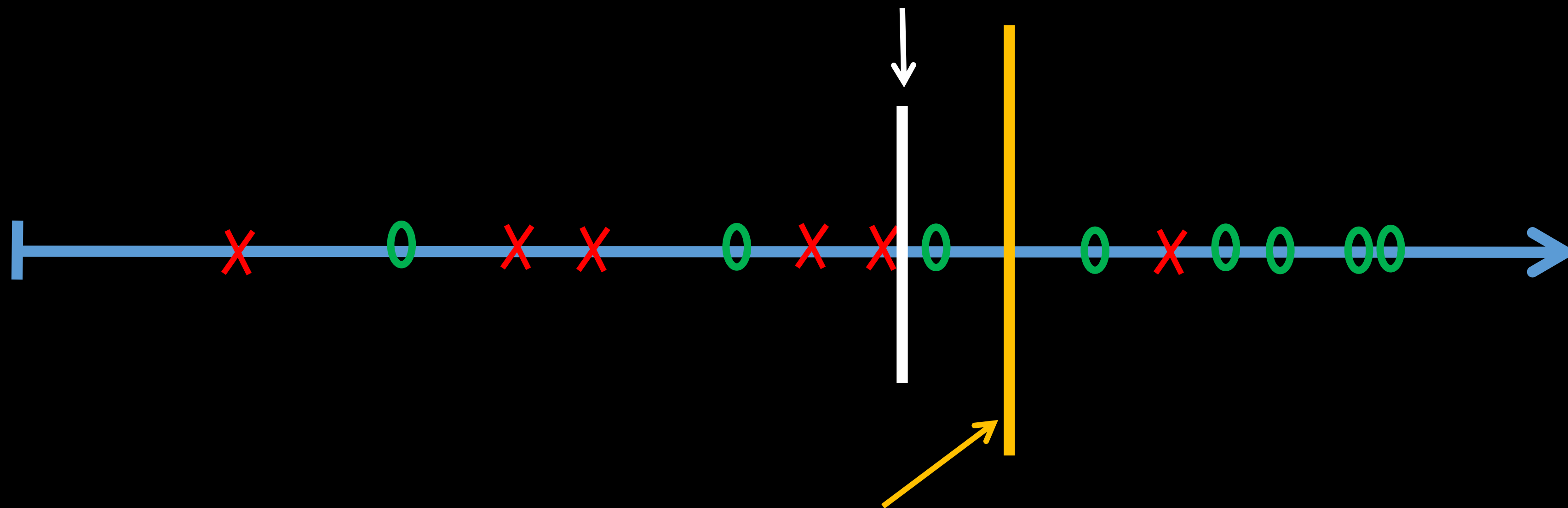
True Concept: 65 and older -> “Senior Citizen”



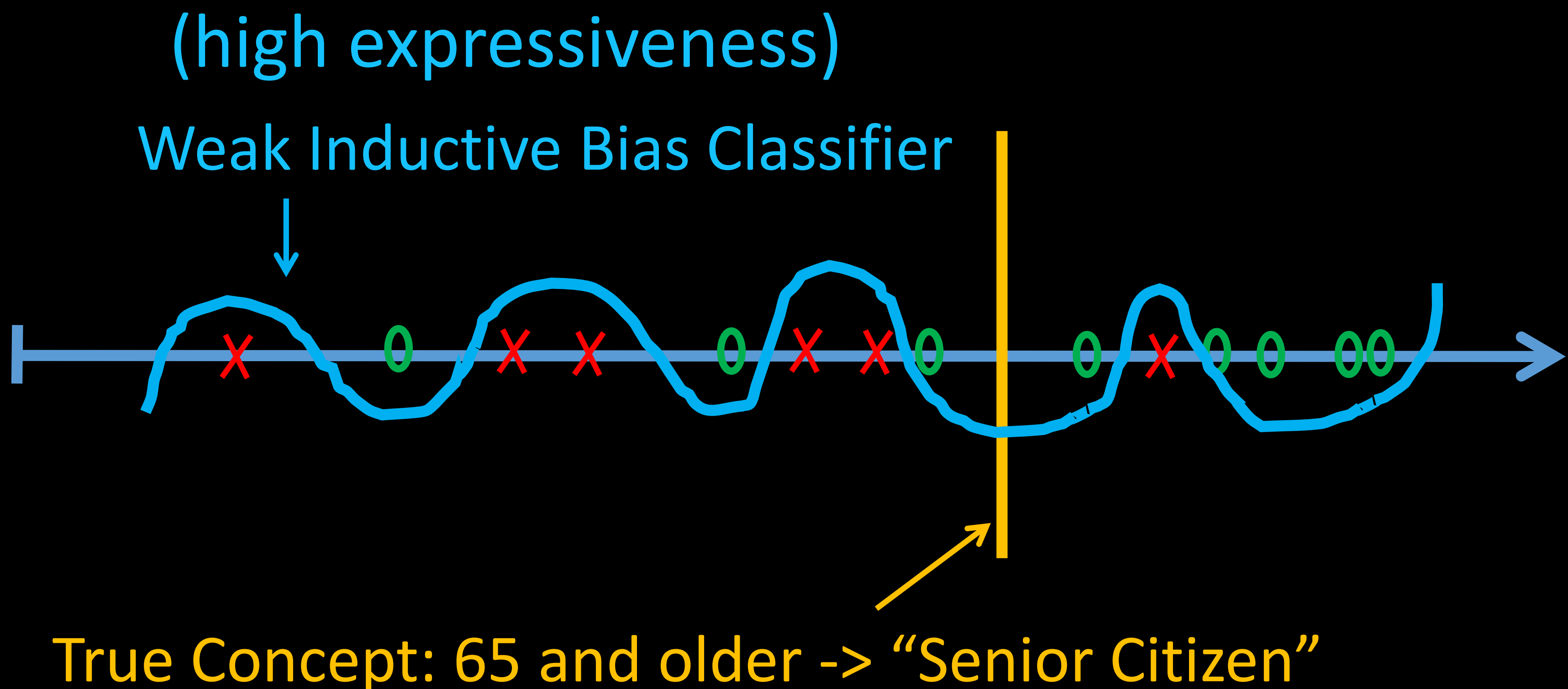
True Concept: 65 and older -> “Senior Citizen”

(low expressiveness)

Strong Inductive Bias Classifier



True Concept: 65 and older -> “Senior Citizen”

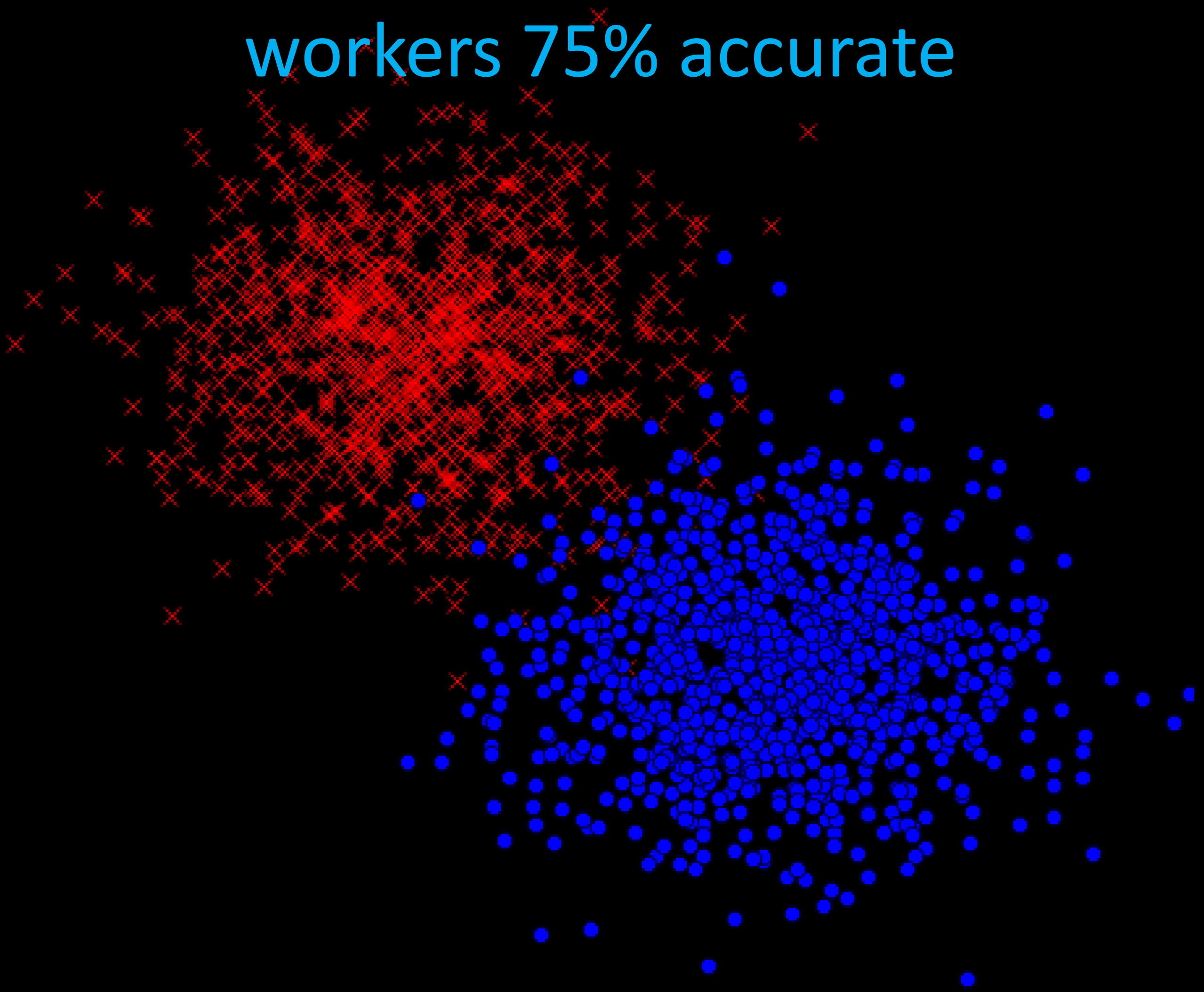

(high expressiveness)
Weak Inductive Bias Classifier
True Concept: 65 and older -> “Senior Citizen”

Weaker Inductive Bias

Increases Relabeling Power

budget b = 500

workers 75% accurate

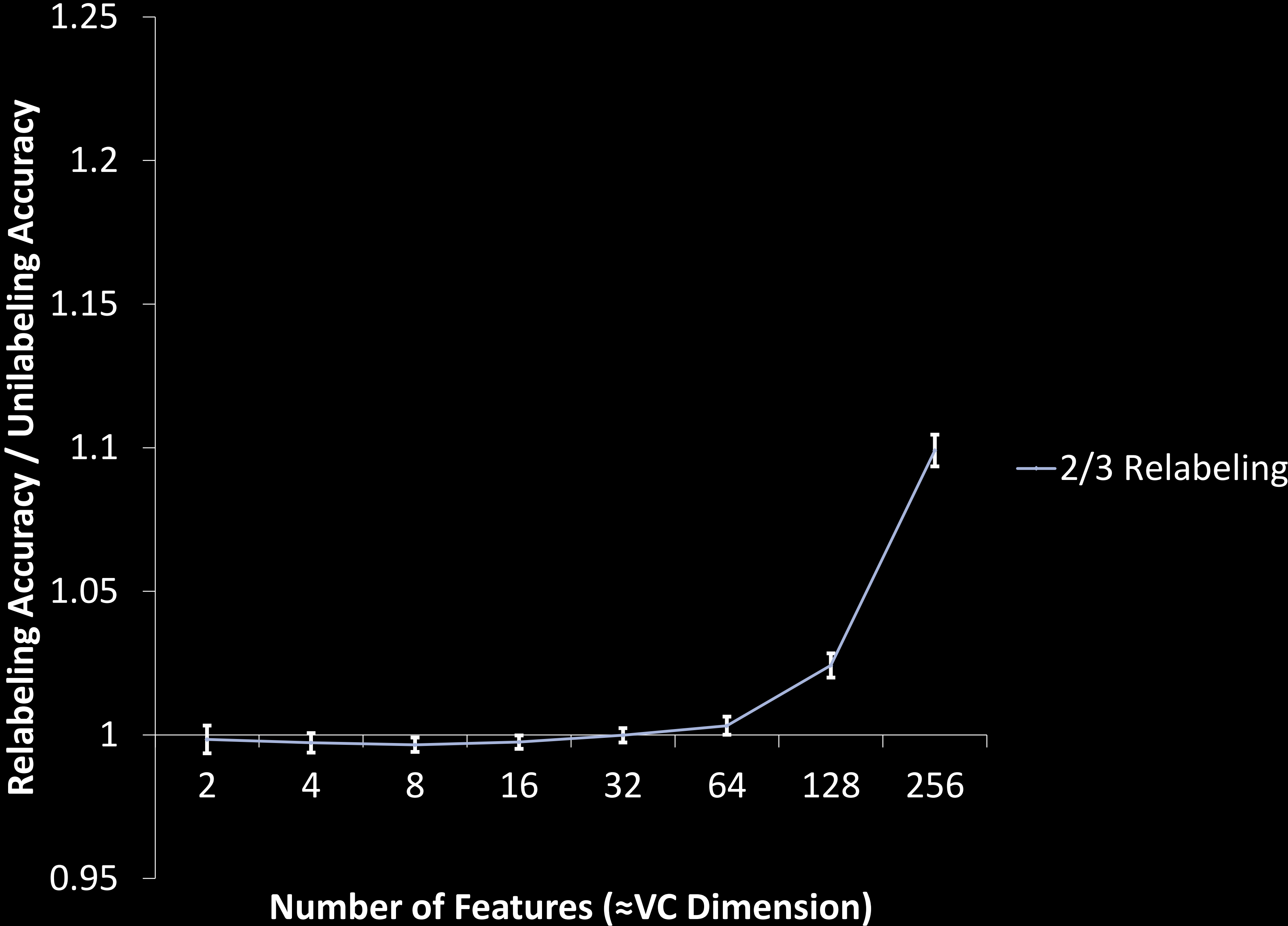


Controlling Inductive Bias via

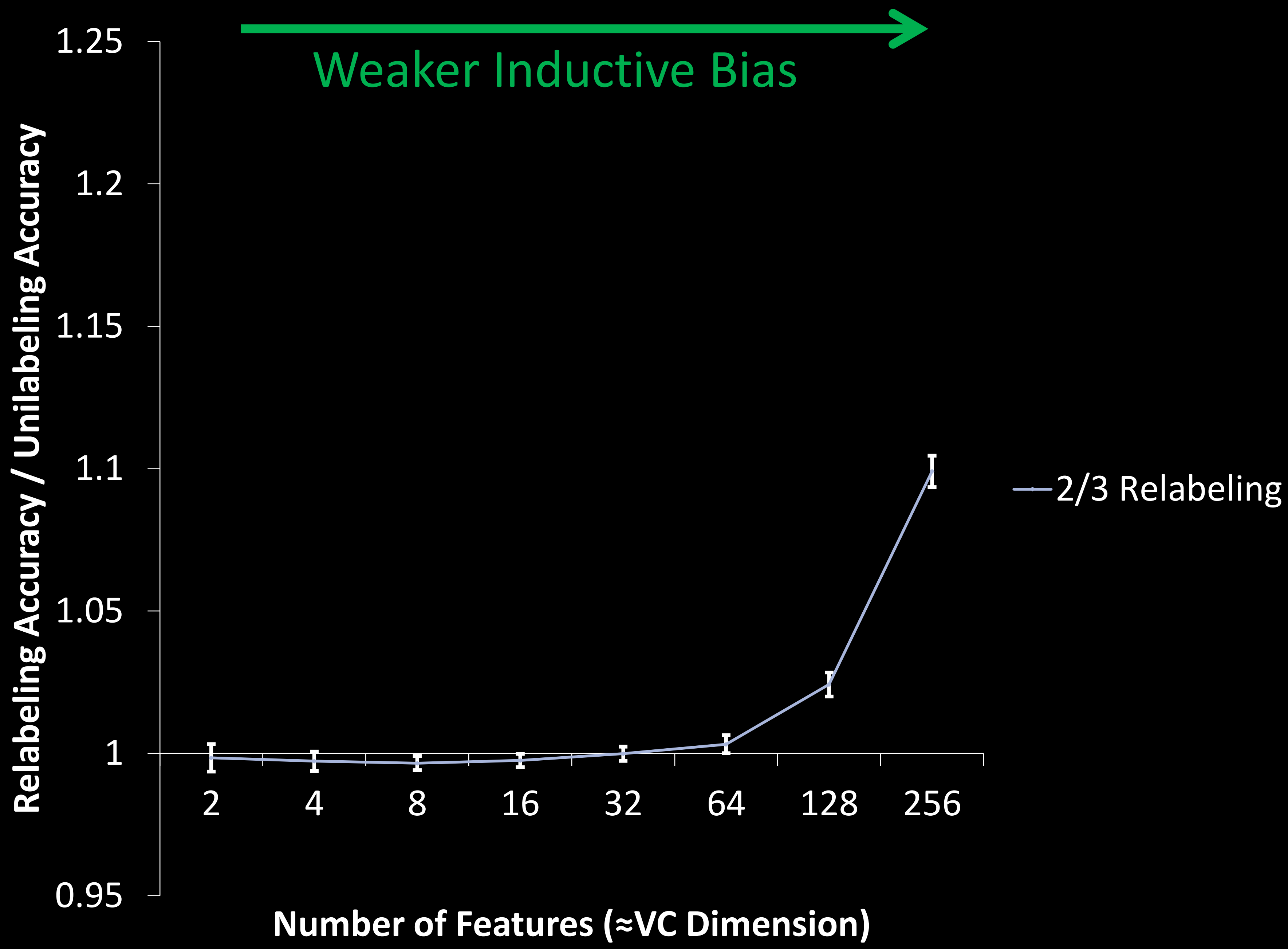
Features

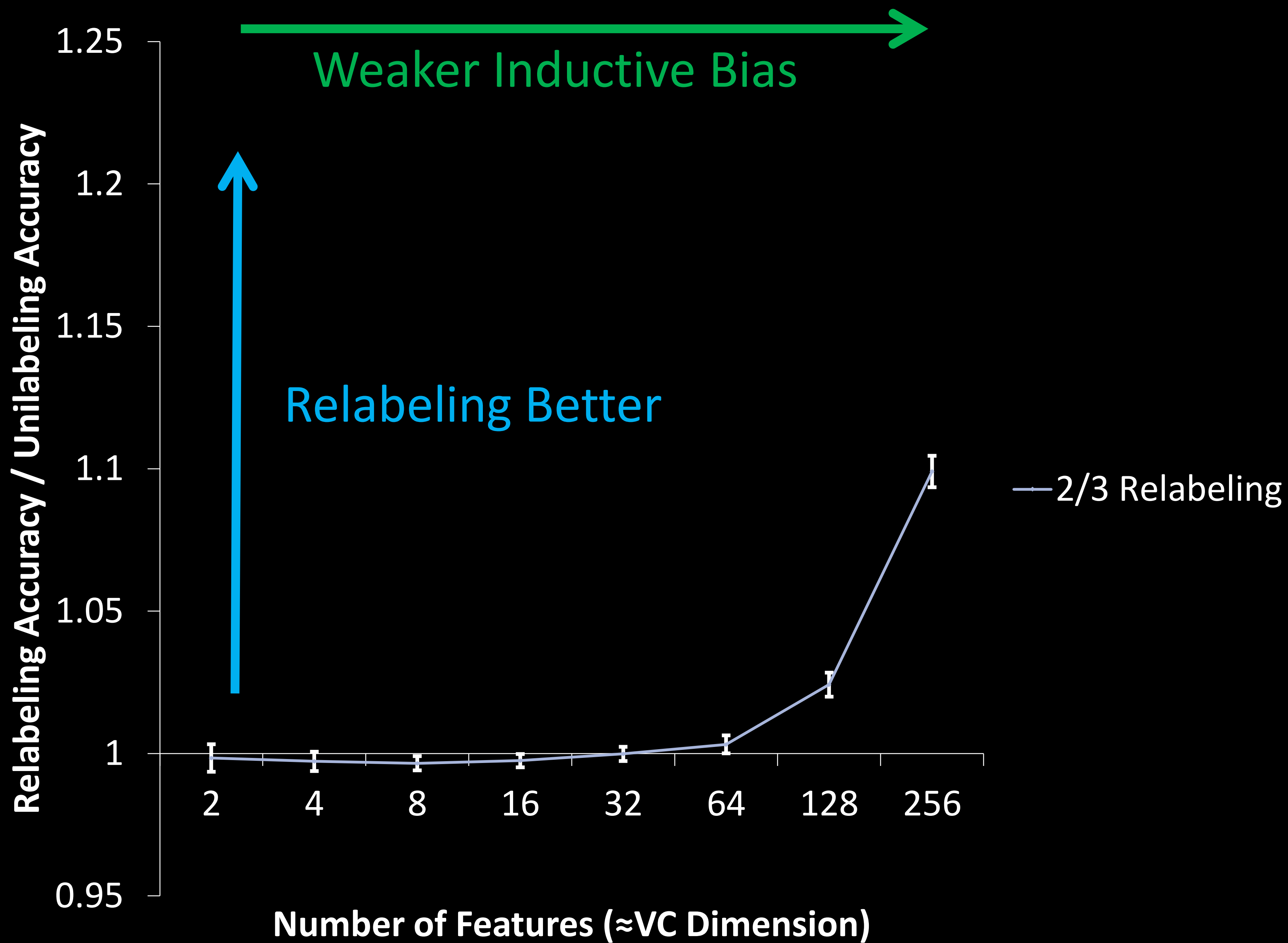
Type of Classifier

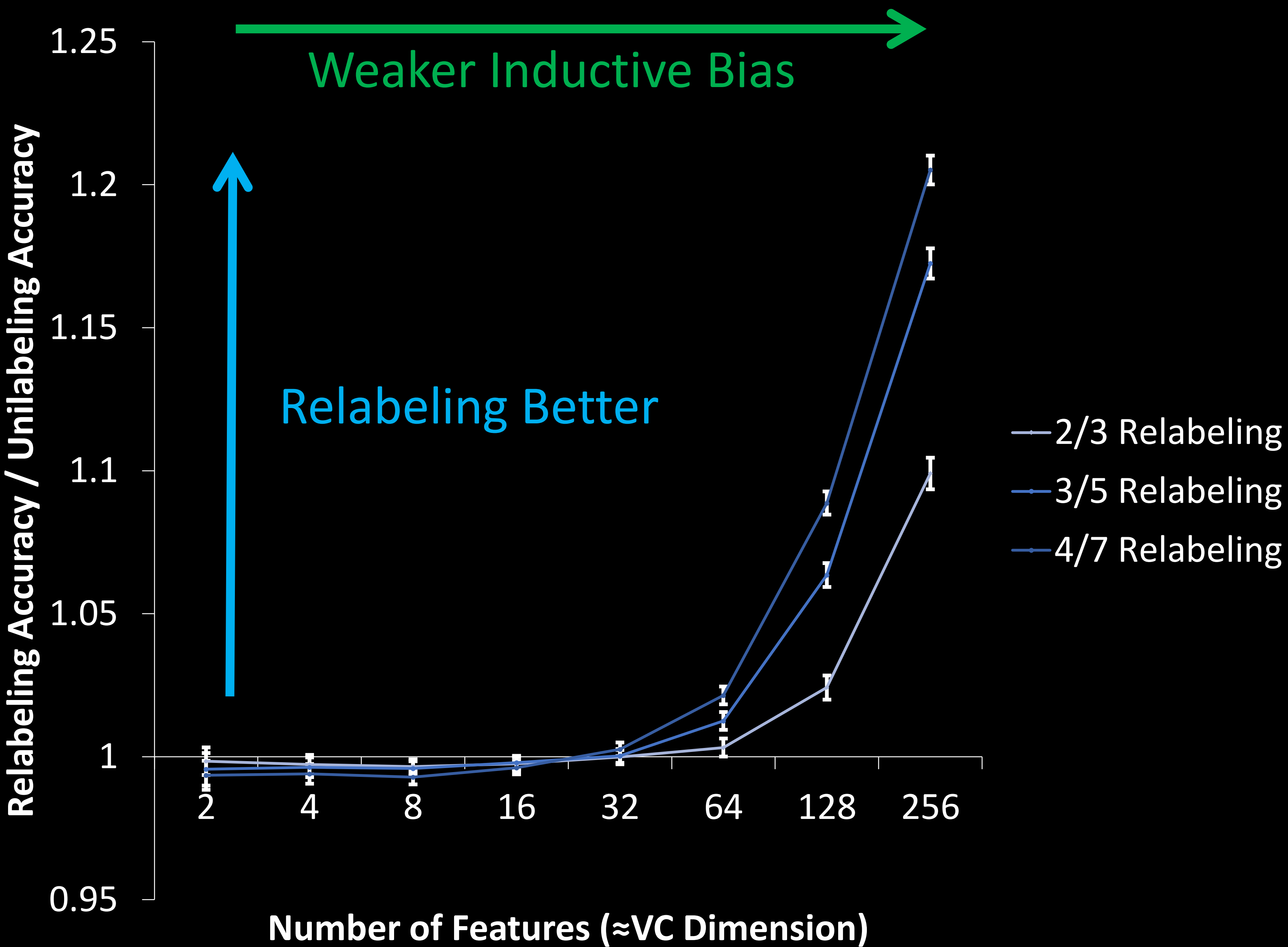
Regularization



Relabeling Accuracy / Unilabeling Accuracy
1.25
1.2
1.15
1.1
1.05
1
0.95
2
4
8
16
32
64
128
256
Number of Features (≈VC Dimension)
2/3 Relabeling


Weaker Inductive Bias
Relabeling Accuracy / Unilabeling Accuracy
1.25
1.2
1.15
1.1
1.05
1
0.95
2
4
8
16
32
64
128
256
Number of Features (≈VC Dimension)
2/3 Relabeling


Weaker Inductive Bias
Relabeling Better
Relabeling Accuracy / Unilabeling Accuracy
1.25
1.2
1.15
1.1
1.05
1
0.95
2
4
8
16
32
64
128
256
Number of Features (≈VC Dimension)
2/3 Relabeling

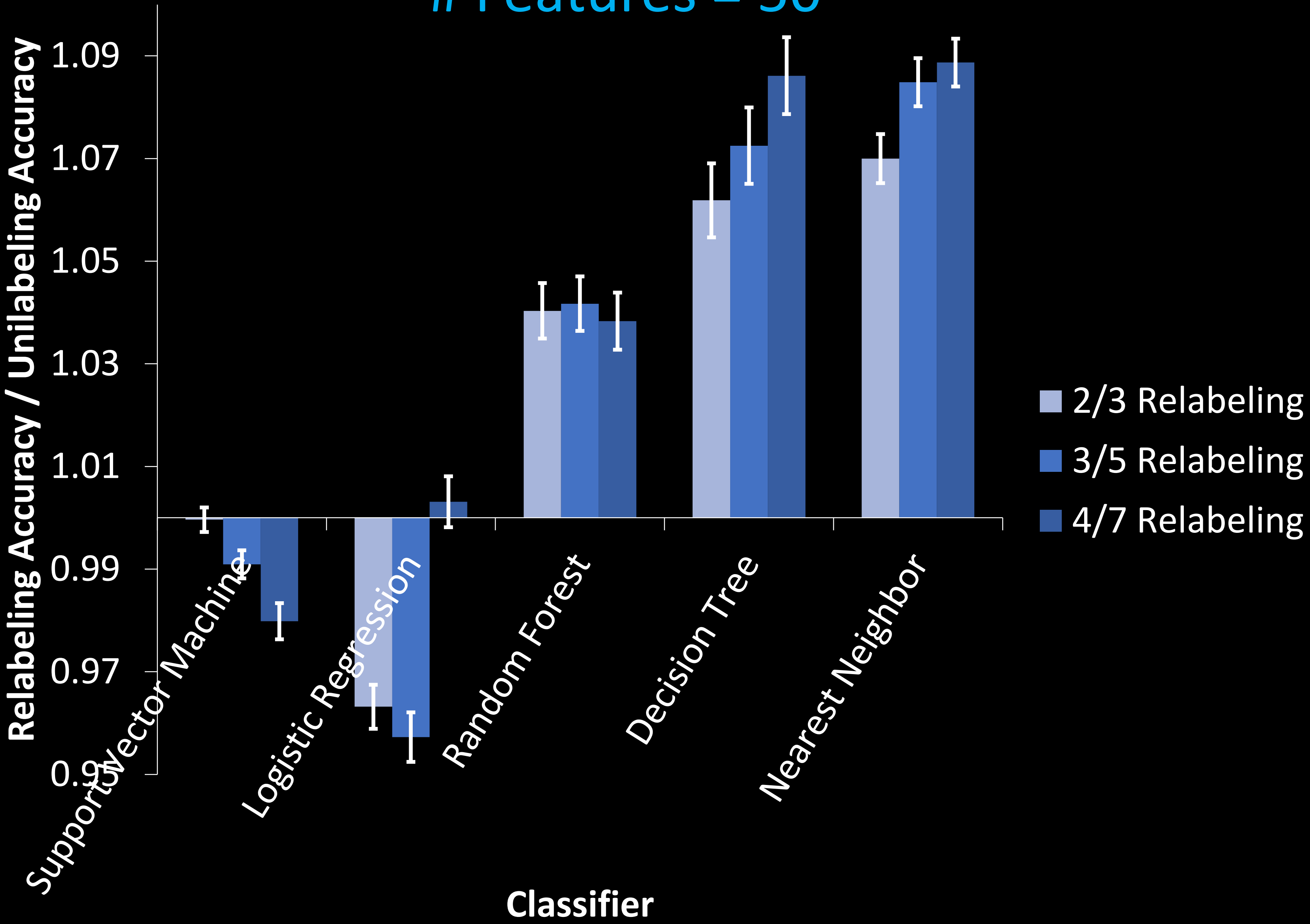


Weaker Inductive Bias
Relabeling Better
Relabeling Accuracy / Unilabeling Accuracy
1.25
1.2
1.15
1.1
1.05
1
0.95
2
4
8
16
32
64
128
256
Number of Features (≈VC Dimension)
2/3 Relabeling
3/5 Relabeling
4/7 Relabeling

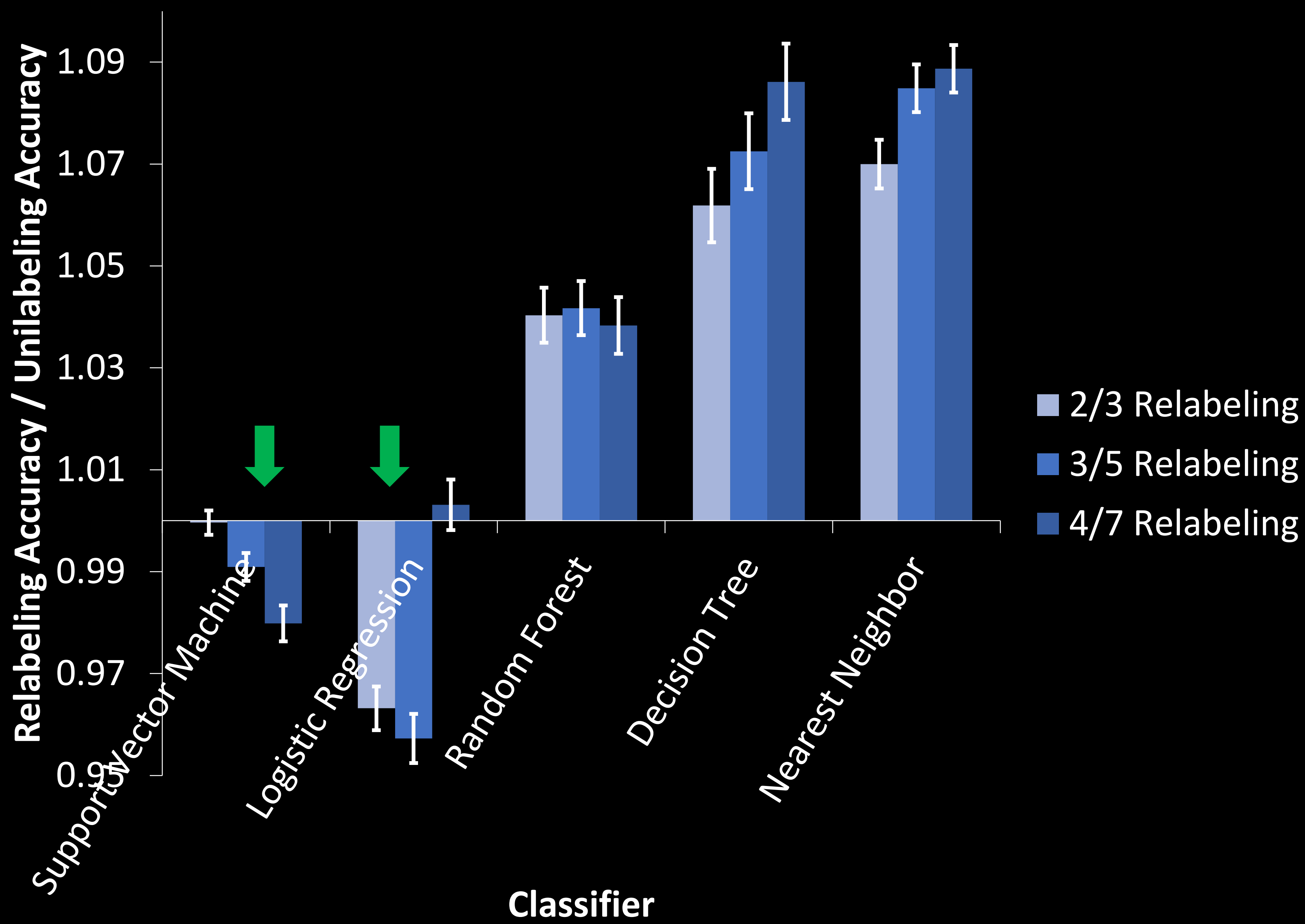
Weaker Inductive Bias

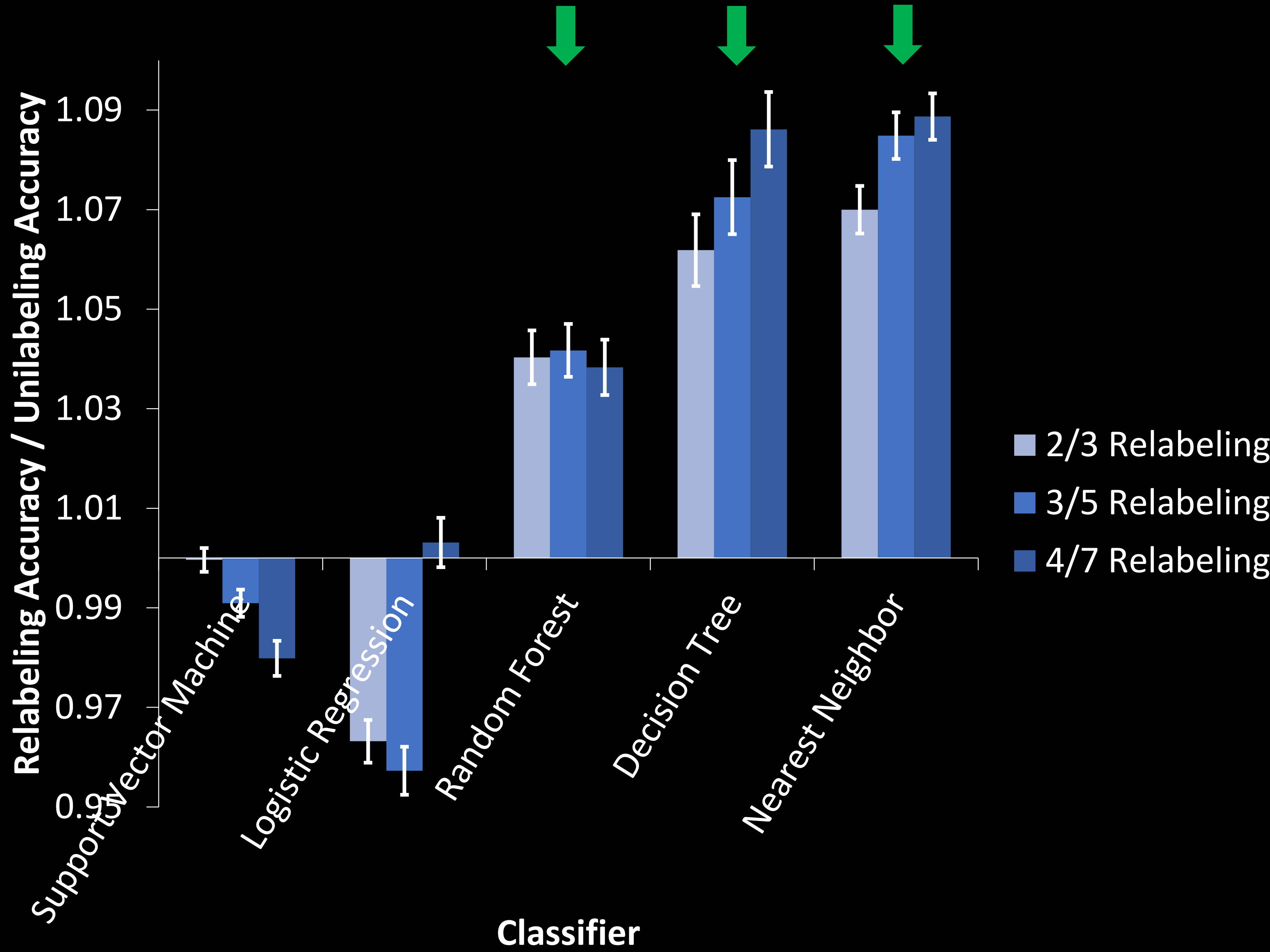
Increases Relabeling Power

Features = 50



Relabeling Accuracy / Unilabeling Accuracy
1.09
1.07
1.05
1.03
1.01
0.99
0.97
0.95
Support Vector Machine
Logistic Regression
Random Forest
Decision Tree
Nearest Neighbor
Classifier
2/3 Relabeling
3/5 Relabeling
4/7 Relabeling


Relabeling Accuracy / Unilabeling Accuracy
1.09
1.07
1.05
1.03
1.01
0.99
0.97
0.95
Support Vector Machine
Logistic Regression
Random Forest
Decision Tree
Nearest Neighbor
Classifier
2/3 Relabeling
3/5 Relabeling
4/7 Relabeling


Relabeling Accuracy / Unilabeling Accuracy
1.09
1.07
1.05
1.03
1.01
0.99
0.97
0.95
Support Vector Machine
Logistic Regression
Random Forest
Decision Tree
Nearest Neighbor
Classifier
2/3 Relabeling
3/5 Relabeling
4/7 Relabeling

Factors that Affect Relabeling Efficacy

Inductive Bias

Weaker Inductive Bias Increases Relabeling Power

Worker Accuracy

Budget

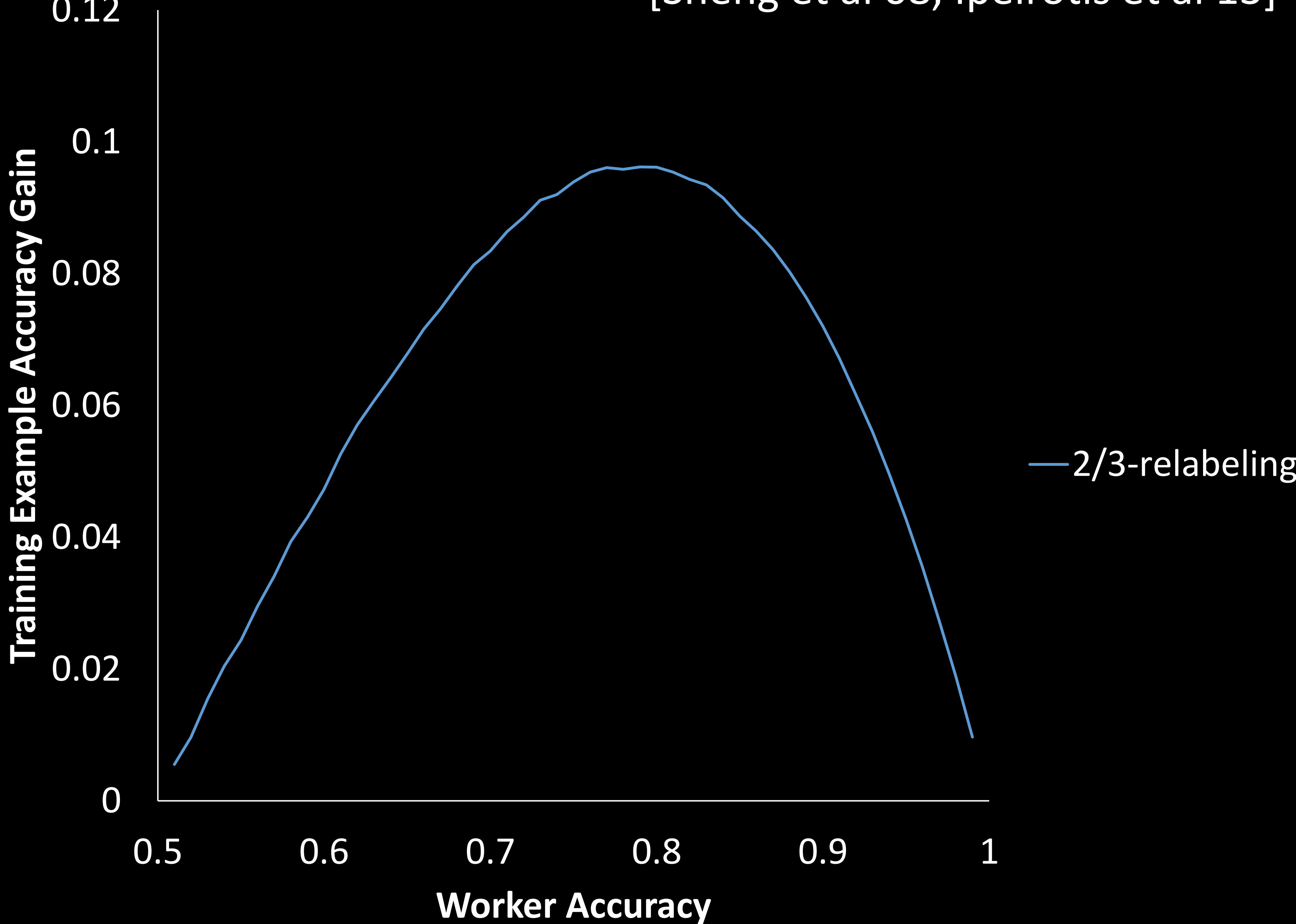
Factors that Affect Relabeling Efficacy

Inductive Bias

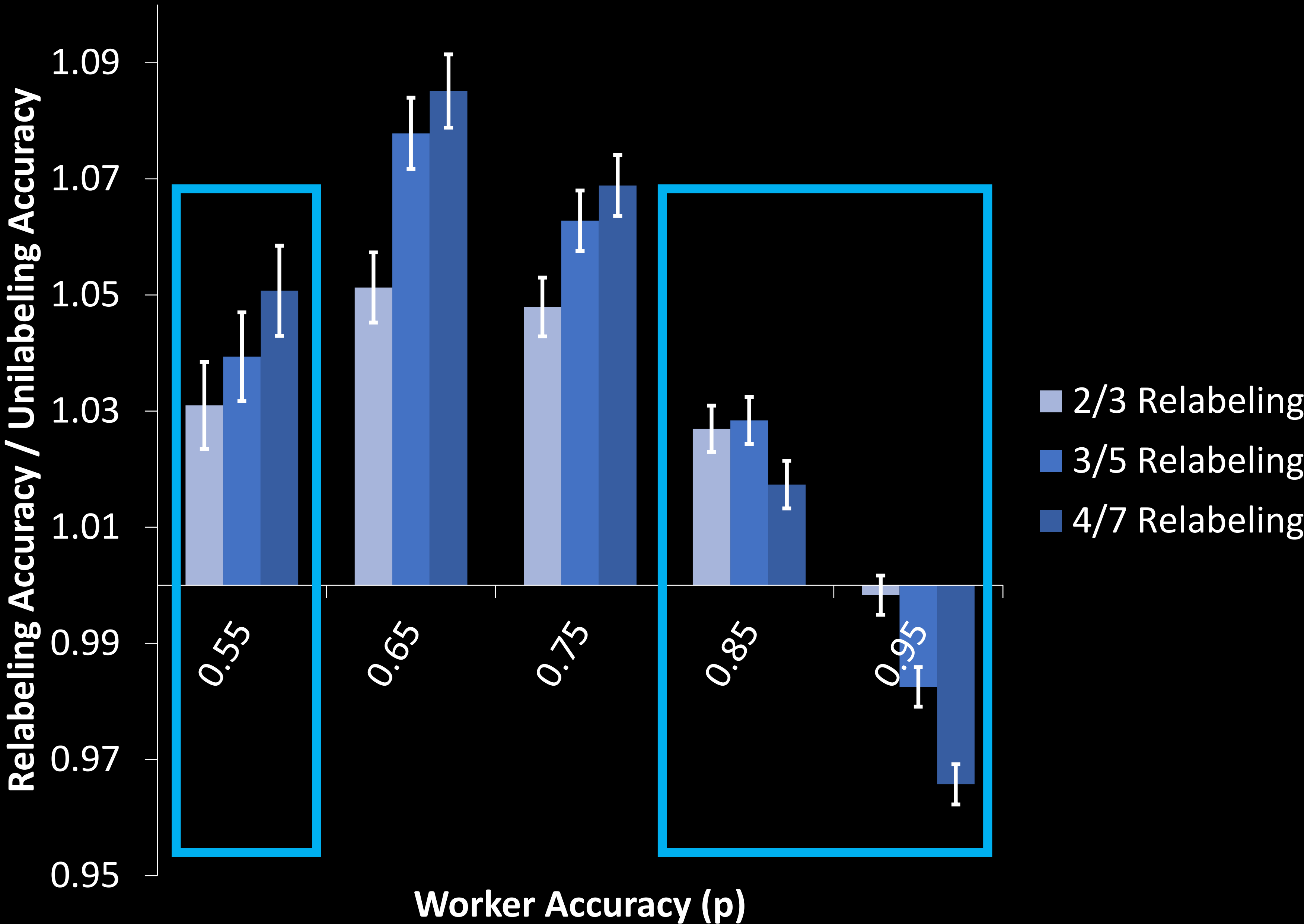
Worker Accuracy

Budget

[Sheng et al 08, Ipeirotis et al 13]




Relabeling Accuracy / Unilabeling Accuracy
1.09
1.07
1.05
1.03
1.01
0.99
0.97
0.95
0.55
0.65
0.75
0.85
0.95
Worker Accuracy (p)
2/3 Relabeling
3/5 Relabeling
4/7 Relabeling



1.09
1.07
1.05
1.03
1.01
0.99
0.97
0.95
Relabeling Accuracy / Unilabeling Accuracy
0.55
0.65
0.75
0.85
0.95
Worker Accuracy (p)
2/3 Relabeling
3/5 Relabeling
4/7 Relabeling

Moderately Accurate Workers Maximize Relabeling Power

Factors that Affect Relabeling Efficacy

Inductive Bias

Weaker Inductive Bias Increases Relabeling Power

Worker Accuracy

Moderately Accurate Workers Maximize Relabeling Power

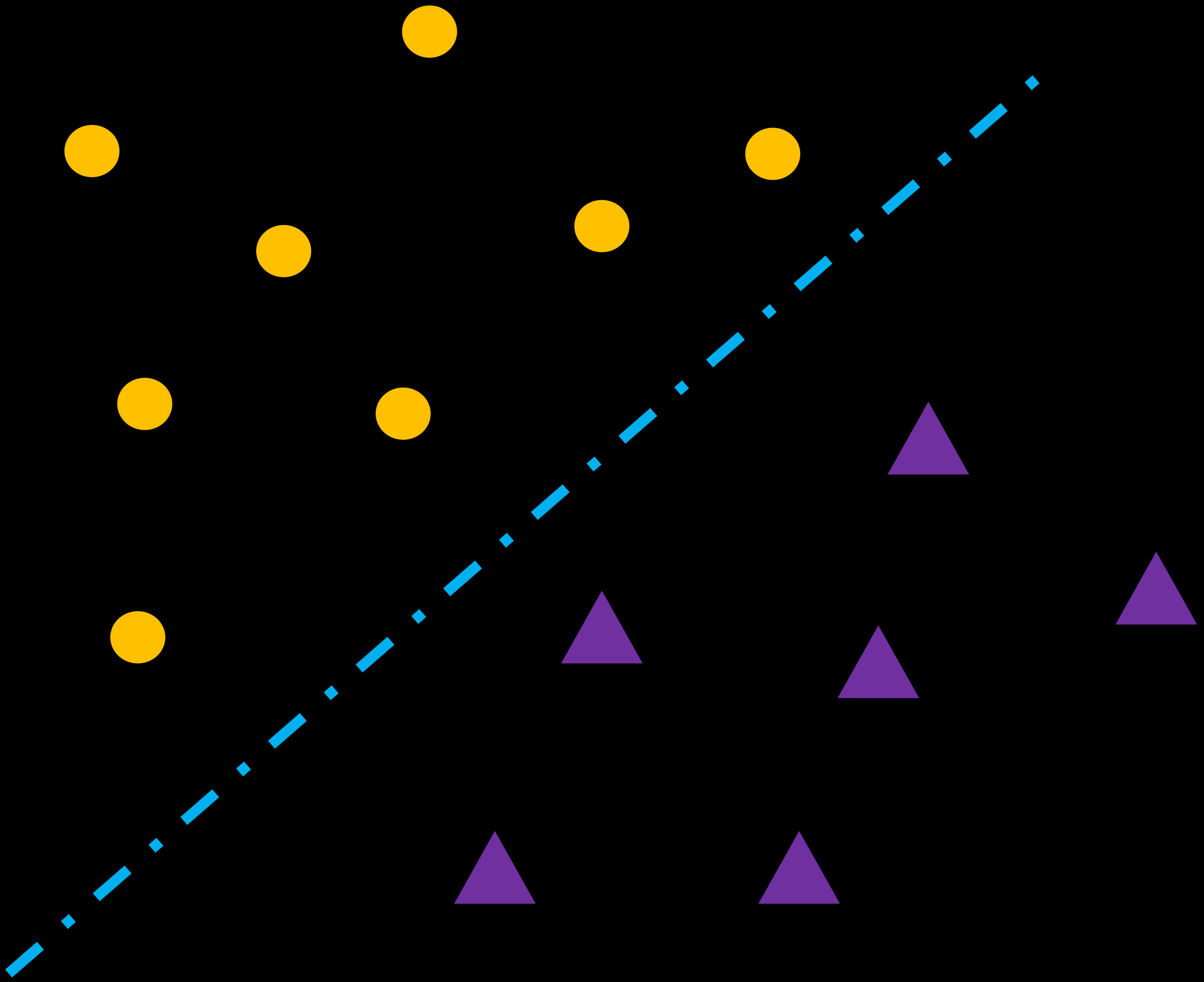
Budget

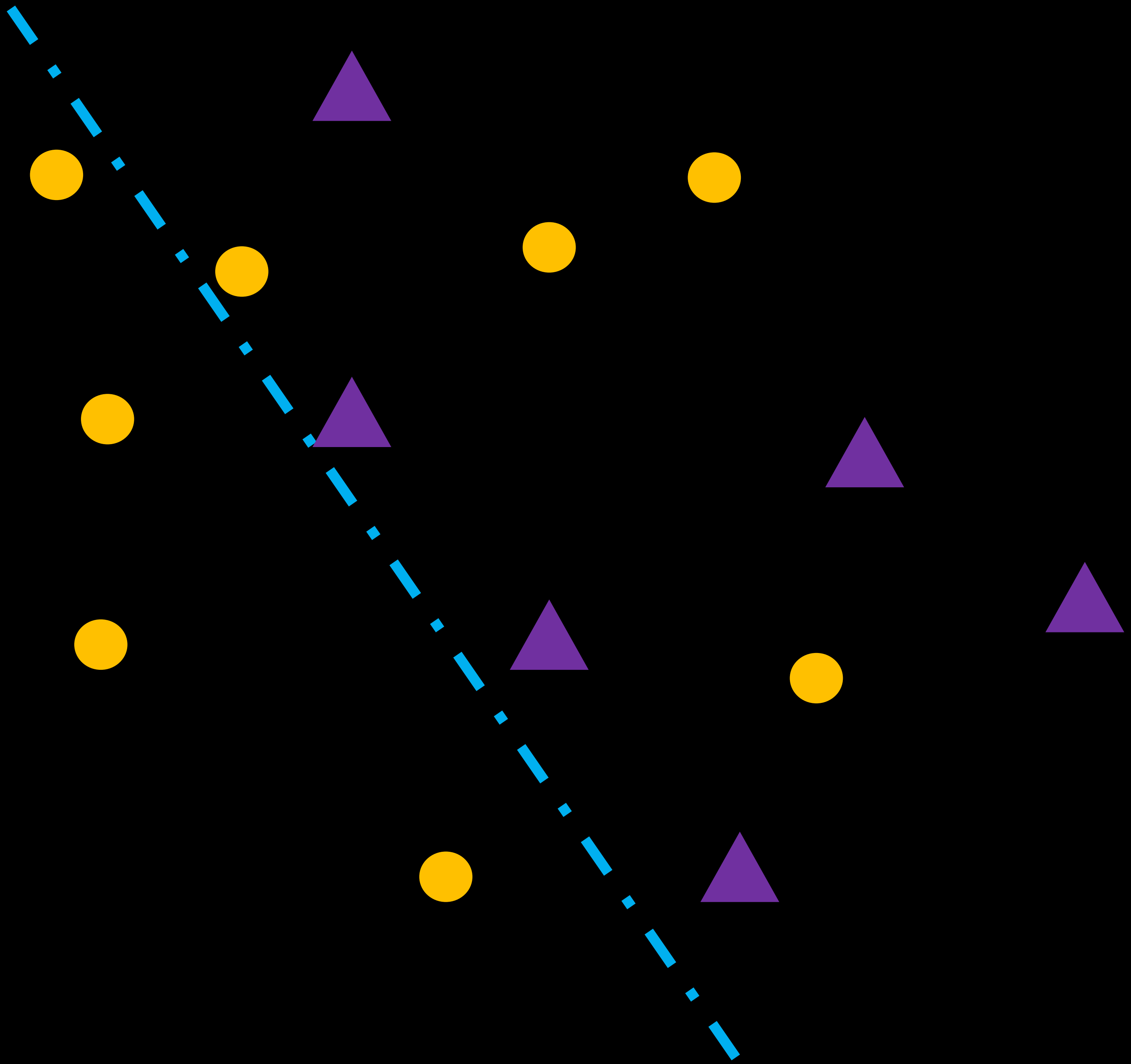
Factors that Affect Relabeling Efficacy

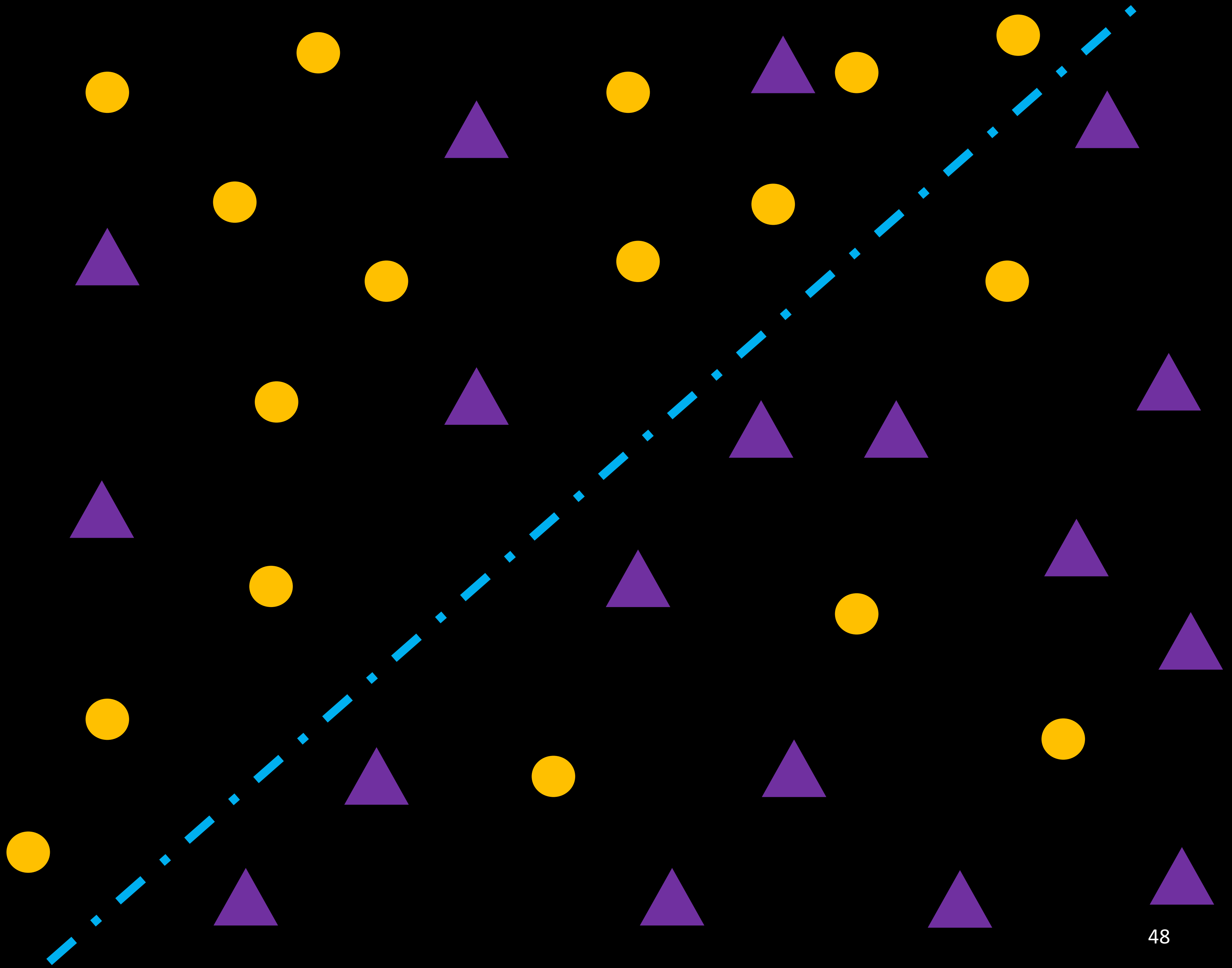
Inductive Bias

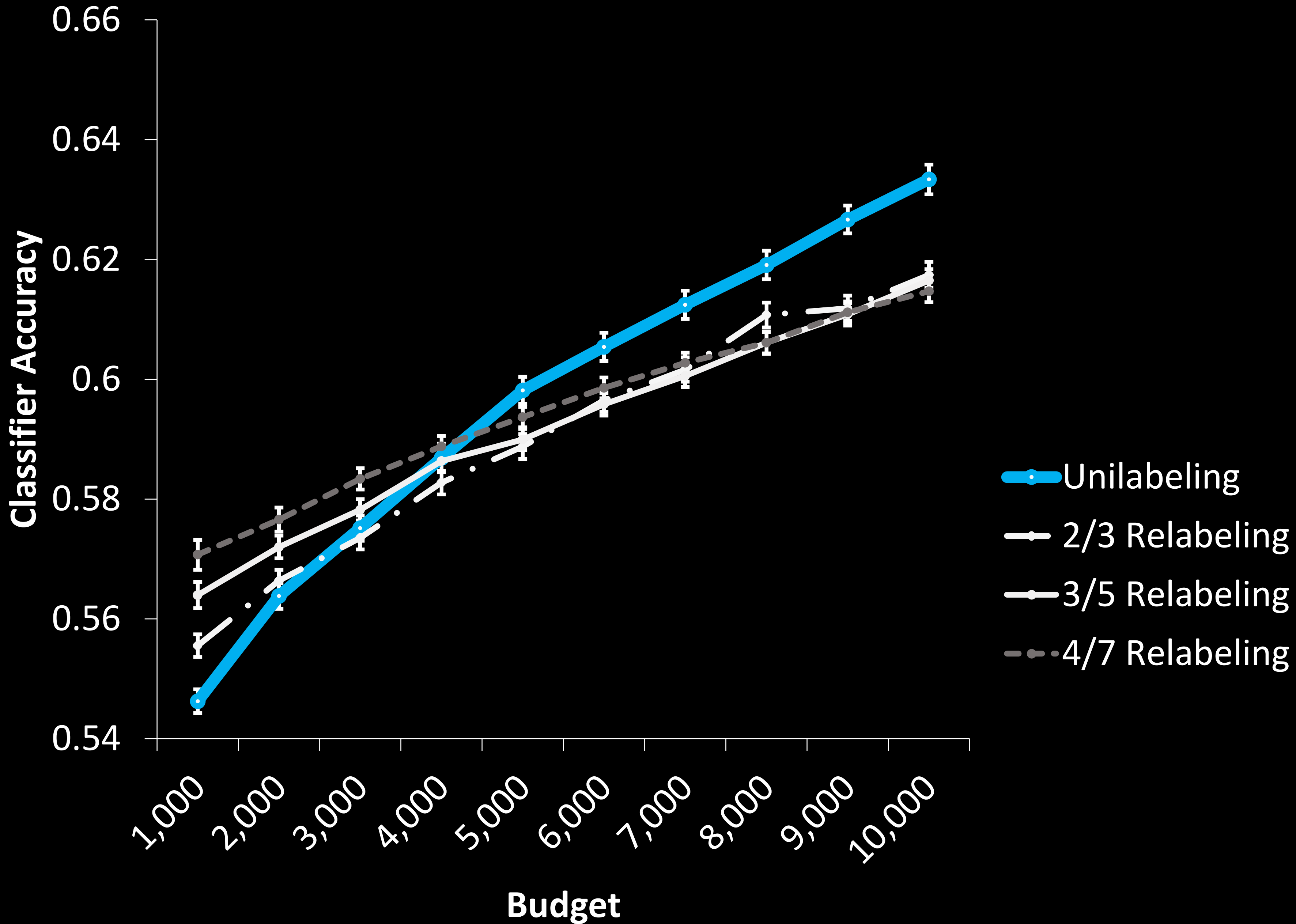
Worker Accuracy

Budget



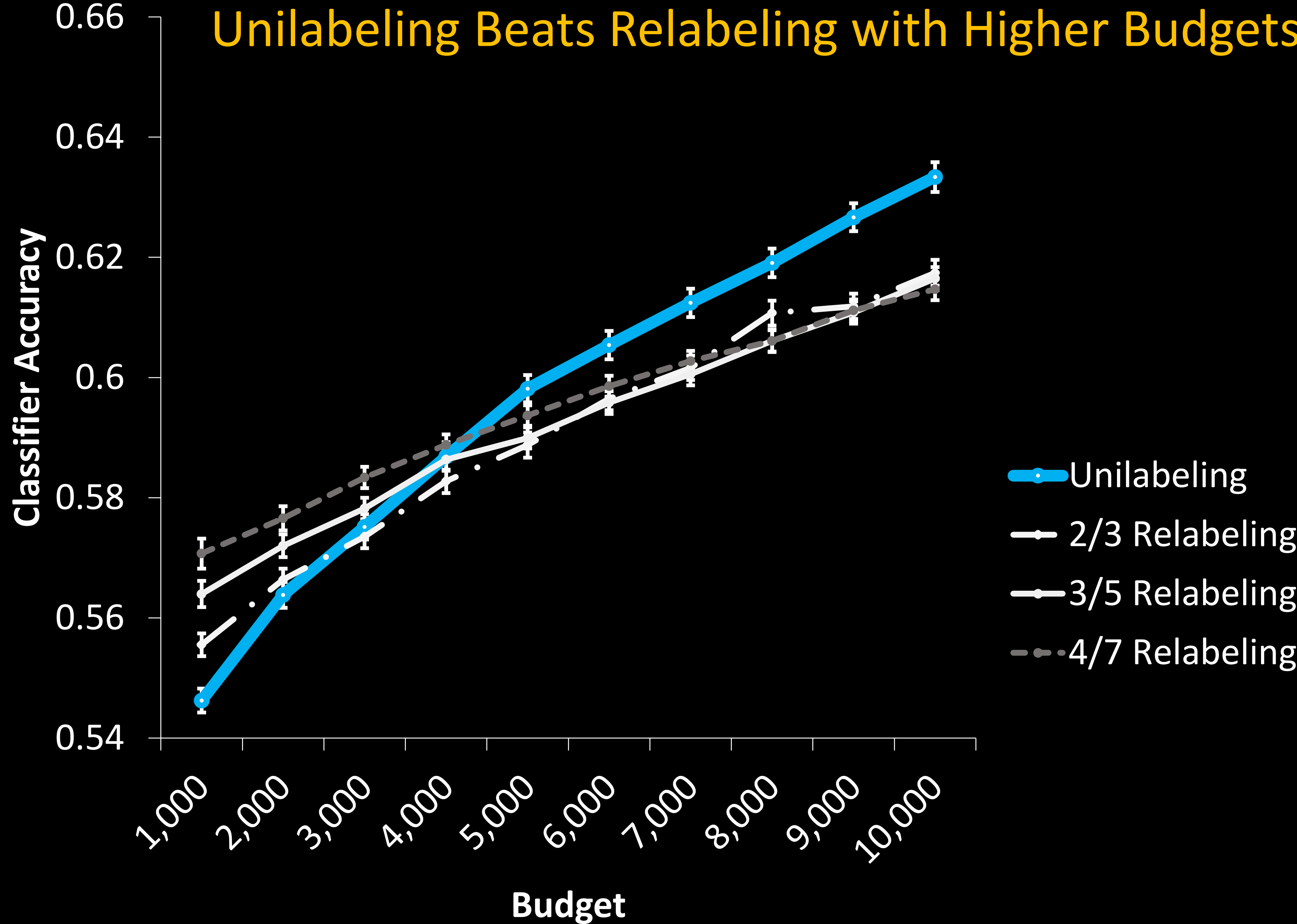






Classifier Accuracy
0.66
0.64
0.62
0.6
0.58
0.56
0.54
1,000
2,000
3,000
4,000
5,000
6,000
7,000
8,000
9,000
10,000
Budget
Unilabeling
2/3 Relabeling
3/5 Relabeling
4/7 Relabeling

Unilabeling Beats Relabeling with Higher Budgets



Factors that Affect Relabeling Efficacy

Inductive Bias

- Weaker Inductive Bias Increases Relabeling Power

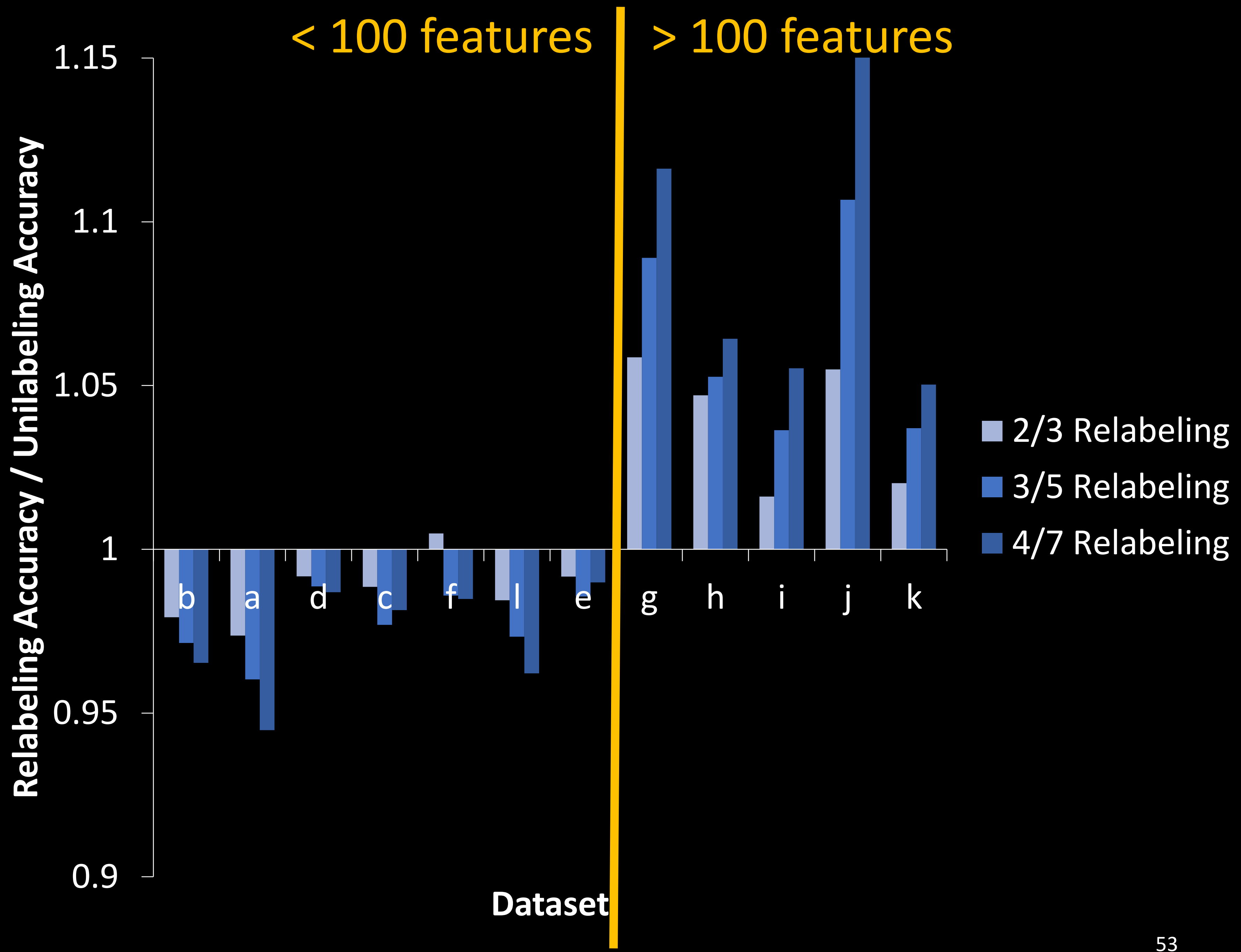
Worker Accuracy

- Moderately Accurate Workers Maximize Relabeling Power

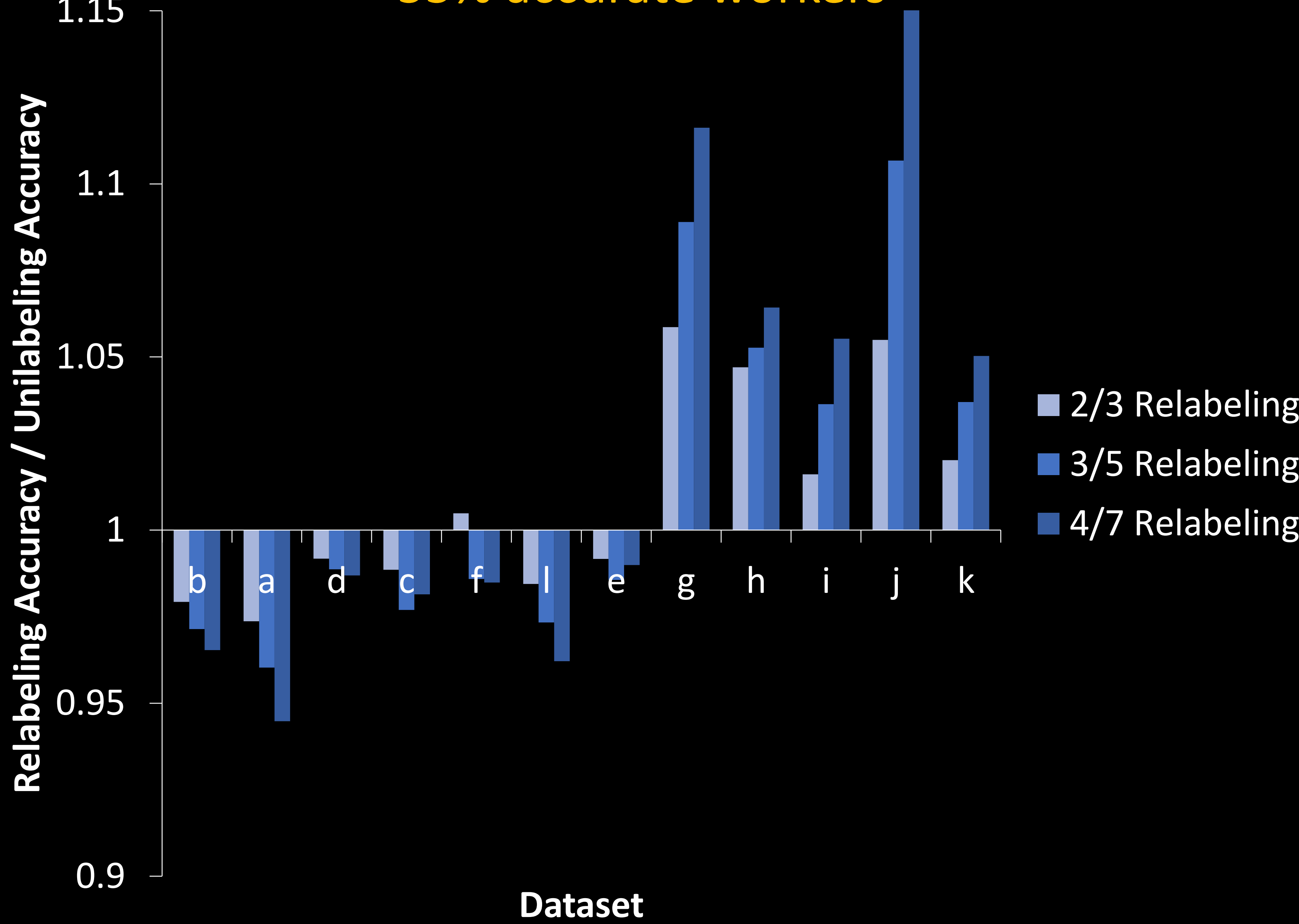
Budget

- Smaller Budgets Increase Relabeling Power

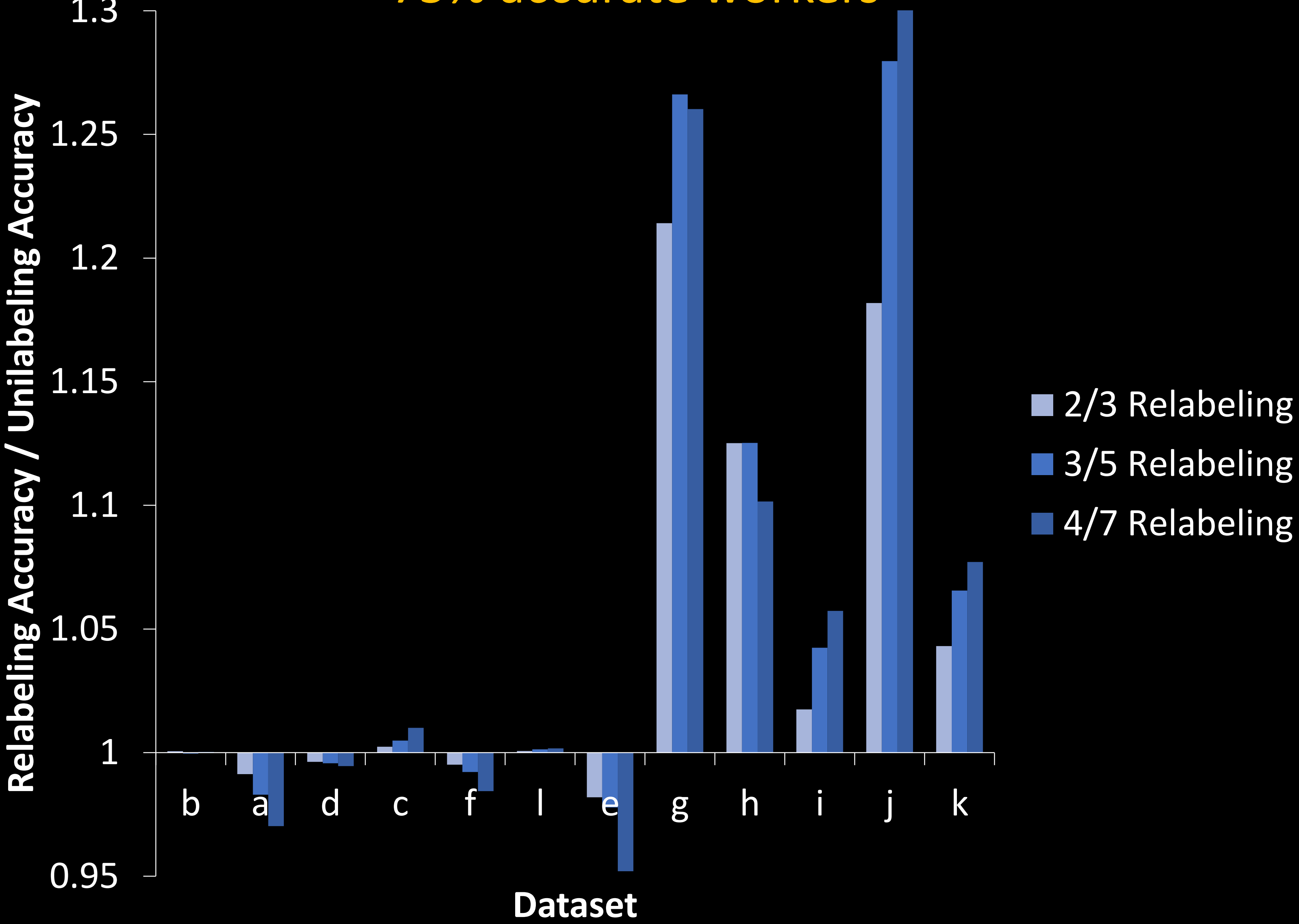
Dataset	# Features	# Examples
(a) Breast Cancer	9	699
(b) Bank Note Authentication	4	1372
(c) Seismic Bumps	18	2584
(d) EEG Eye State	14	14980
(e) Sonar	60	208
(f) Breast Cancer Diagnostic	30	569
(g) Hill-Valley	100	606
(h) Hill-Valley with Noise	100	606
(i) Internet Ads	1558	2359
(j) Gisette	5000	6000
(k) Farm Ads	54877	4143
(l) Spambase	57	4601


< 100 features
> 100 features
Relabeling Accuracy / Unilabeling Accuracy
1.15
1.1
1.05
1
0.95
0.9
b
a
d
c
f
l
e
g
h
i
j
k
Dataset
2/3 Relabeling
3/5 Relabeling
4/7 Relabeling

55% accurate workers



75% accurate workers



Factors that Affect Relabeling Efficacy

Inductive Bias

Weaker Inductive Bias Increases Relabeling Power

Worker Accuracy

Moderately Accurate Workers Maximize Relabeling Power

Budget

Smaller Budgets Increase Relabeling Power

Future Work

Assumptions

Passive Learning

Identical Workers

Majority Vote

Binary Classification

Constant Cost

~~Assumptions~~

Passive Learning →	Active Learning
Identical Workers →	Real Workers
Majority Vote →	Sophisticated AI
Binary Classification →	General Classification
Constant Cost →	Varying Costs

~~Assumptions~~

Passive Learning	Active Learning
Identical Workers	Real Workers
Majority Vote	Sophisticated AI
Binary Classification	General Classification
Constant Cost	Varying Costs

End-to-end Decision Theoretic System

Factors that Affect Relabeling Efficacy

Inductive Bias

Weaker Inductive Bias Increases Relabeling Power

Worker Accuracy

Moderately Accurate Workers Maximize Relabeling Power

Budget

Smaller Budgets Increase Relabeling Power