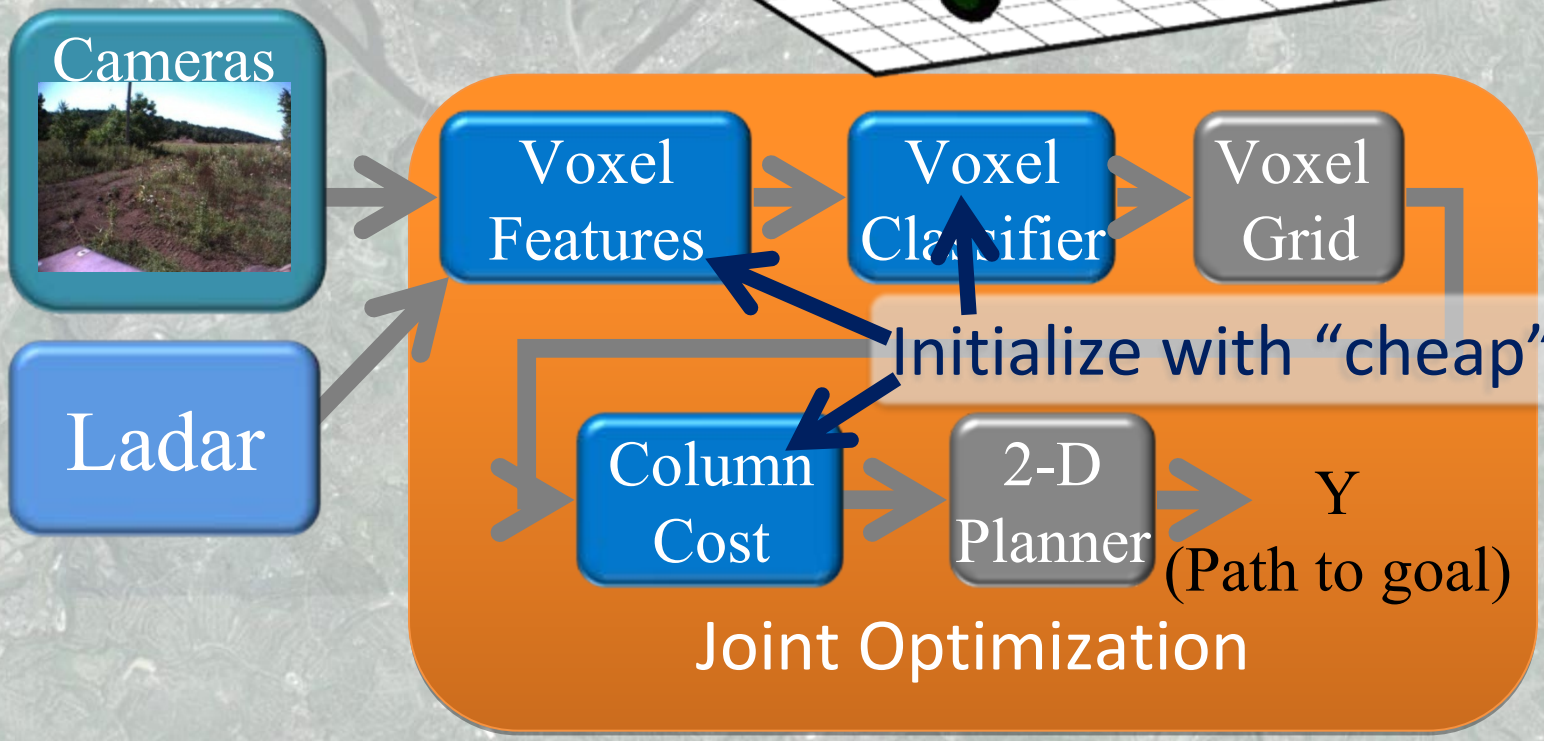
Differentiable Sparse Coding

David Bradley and J. Andrew Bagnell

NIPS 2008

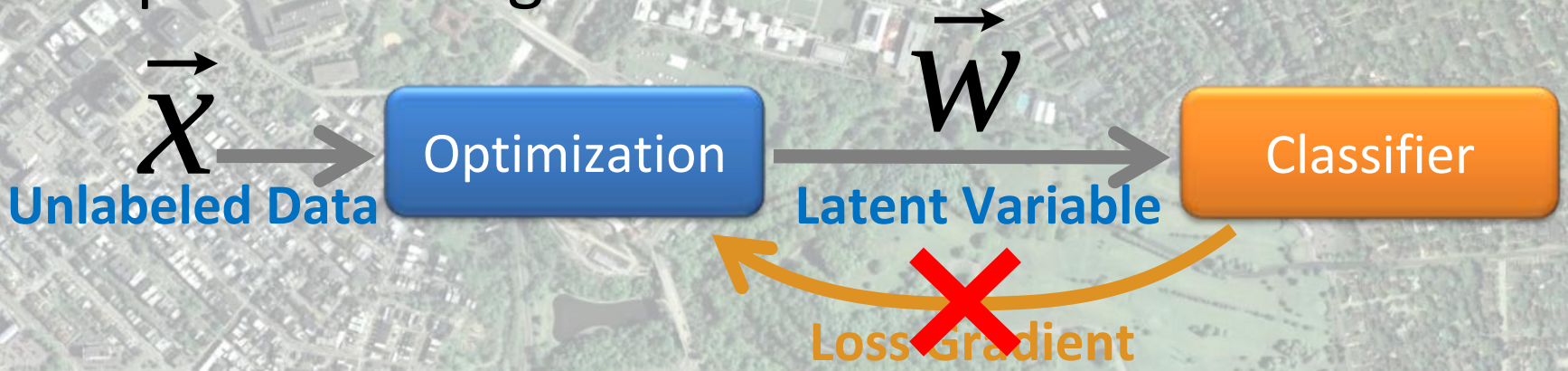
100,000 ft View

- Complex Systems



10,000 ft view

- Sparse Coding = Generative Model

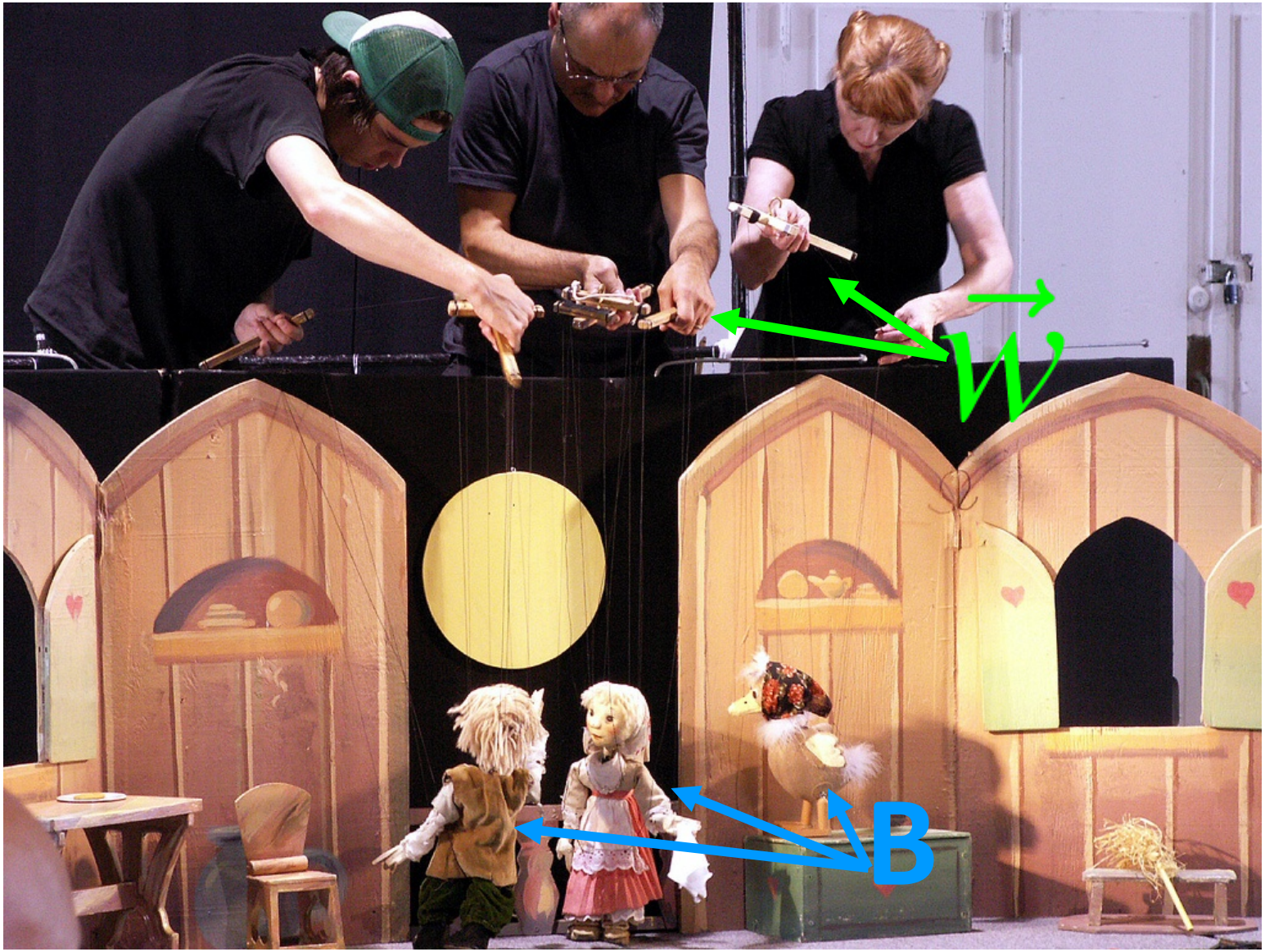


- Semi-supervised learning
- **KL-Divergence Regularization**
- **Implicit Differentiation**

Sparse Coding

Understand X

As a combination of factors



Sparse coding uses optimization

Reconstruction loss function

Some vector

$$\vec{x} \approx f(B\vec{w})$$

Want to use to classify x

Projection (feed-forward):

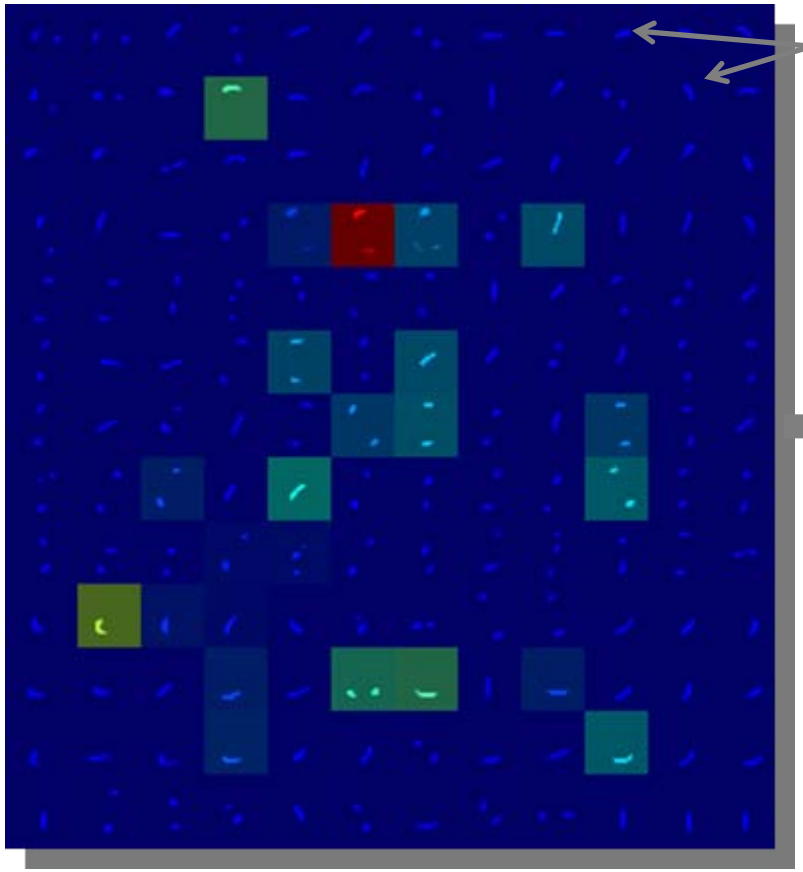
$$\vec{w} = f(\vec{x}^T B)$$

Sparse vectors:

Only a few significant elements

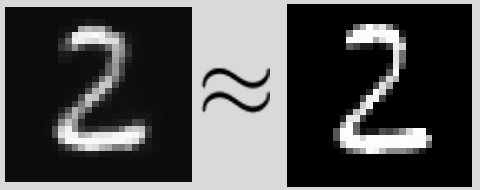
Example: X=Handwritten Digits

$\sum$



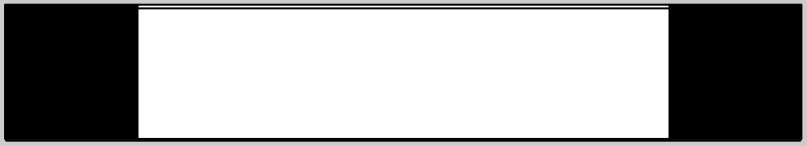
Basis vectors colored by w

$f(B\vec{w})$ X

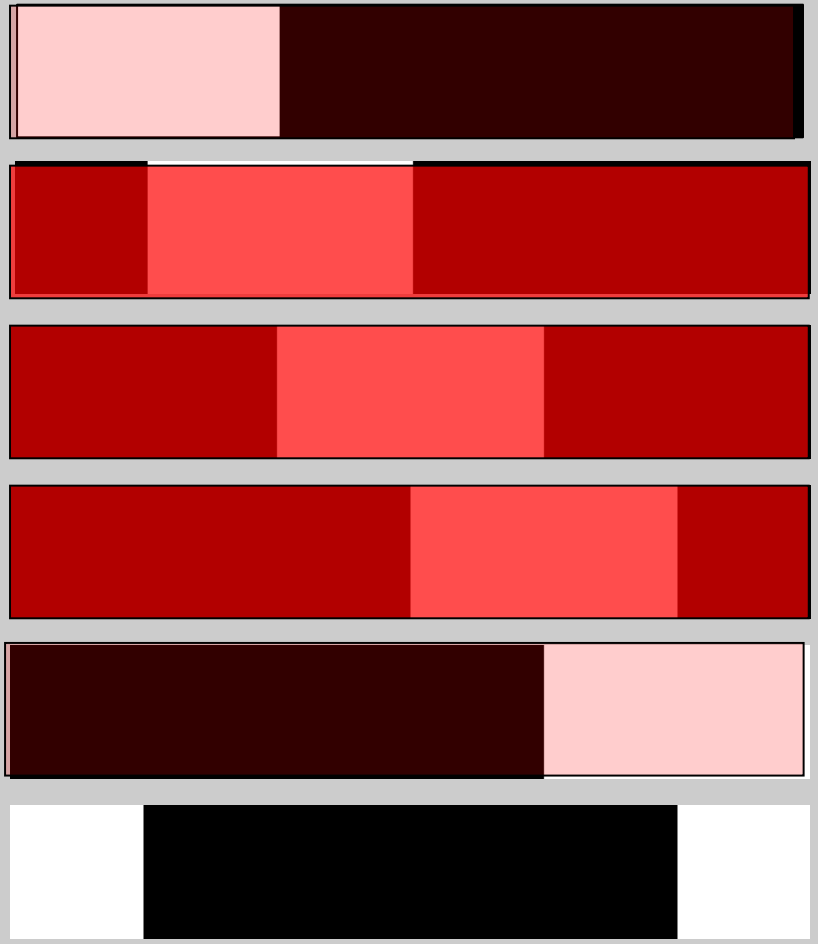


Optimization vs. Projection

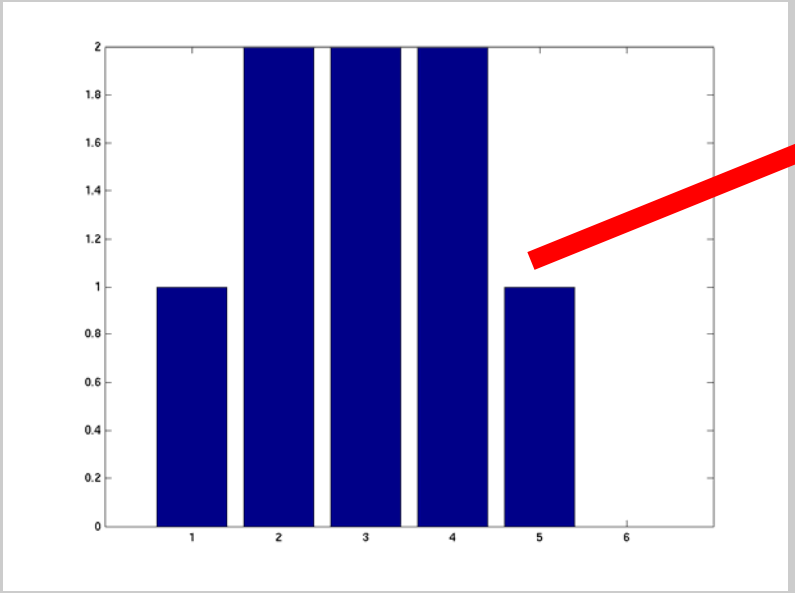
Input



Basis

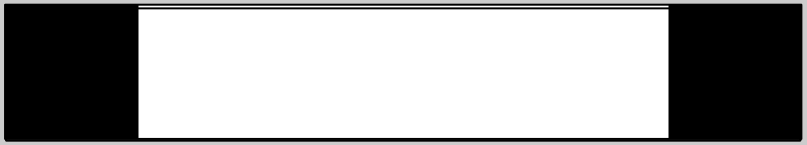


Projection



Optimization vs. Projection

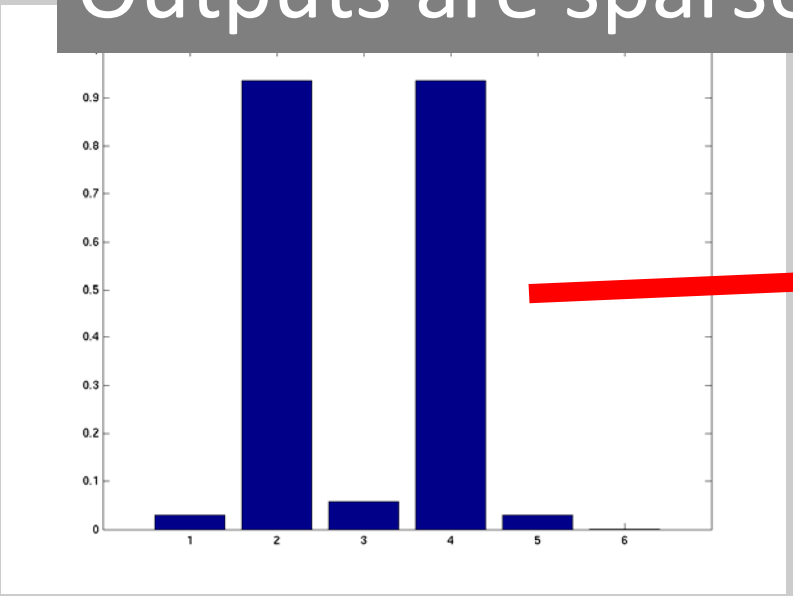
Input

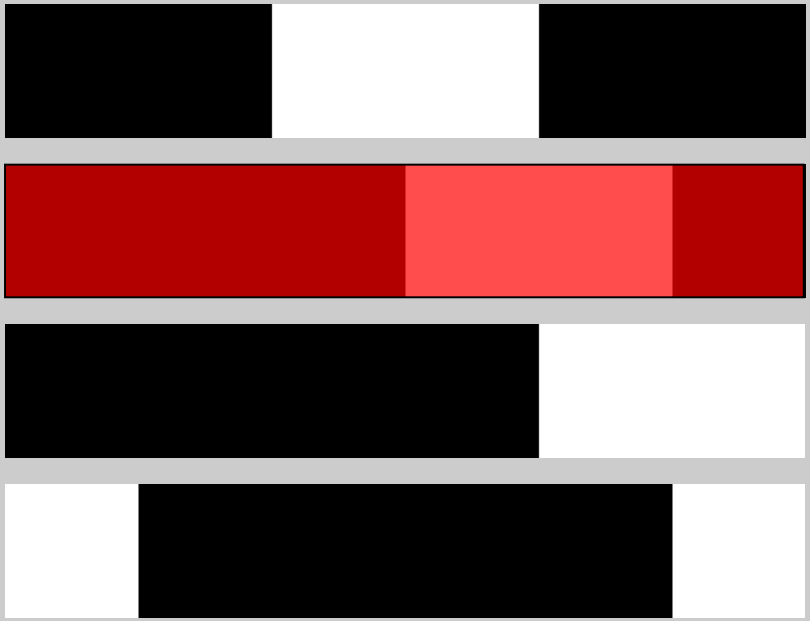


KL

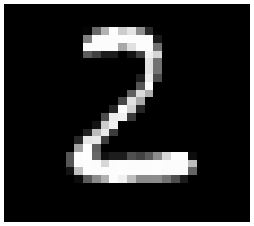
Basis

Outputs are sparse for each example





Generative Model

$X_i =$ 

Latent variables are Independent

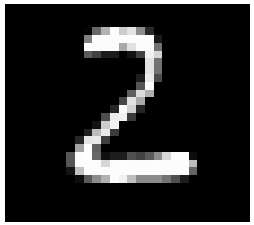
$$P(X) = \int_B \int_W P(X|W,B)P(W)P(B)dW\,dB$$

Likelihood Prior

$$P(X) = \int_B P(B) \int_W \prod_i P(X_i|W_i,B)P(W_i)dW\,dB$$

Examples are Independent

Sparse Approximation

$$X =$$ 

Likelihood Prior

$$P(\vec{x}) \approx \max_{\vec{w}} P(\vec{x}|\vec{w}, B) P(\vec{w})$$

$$\hat{w} = \arg\max_{\vec{w}} P(\vec{x}|\vec{w}, B) P(\vec{w})$$

MAP Estimate

Sparse Approximation

$$\arg\max_{\vec{w}} P(\vec{x}) = \arg\min_{\vec{w}} \left(-\log P(\vec{x})\right)$$

Distance between reconstruction and input

Distance between weight vector and prior mean

$$-\log P(\vec{x}) \propto \text{Loss}\left(\vec{x} \,\|\, f(B\vec{w})\right) + \lambda \text{Prior}\left(\vec{w} \,\|\, \vec{p}\right)$$

Regularization Constant

Example: Squared Loss + L1

$$\hat{w} = \arg\min_{\vec{w}} \sum_i (x_i - B^i \vec{w})^2 + \lambda \sum_i |\vec{w}_i|$$

- Convex + sparse (widely studied in engineering)
- Sparse coding solves for B as well (non-convex for now…)
- Shown to generate useful features on diverse problems

Tropp, *Signal Processing*, 2006

Donoho and Elad, *Proceedings of the National Academy of Sciences*, 2002

Raina, Battle, Lee, Packer, Ng, ICML, 2007

L1 Sparse Coding

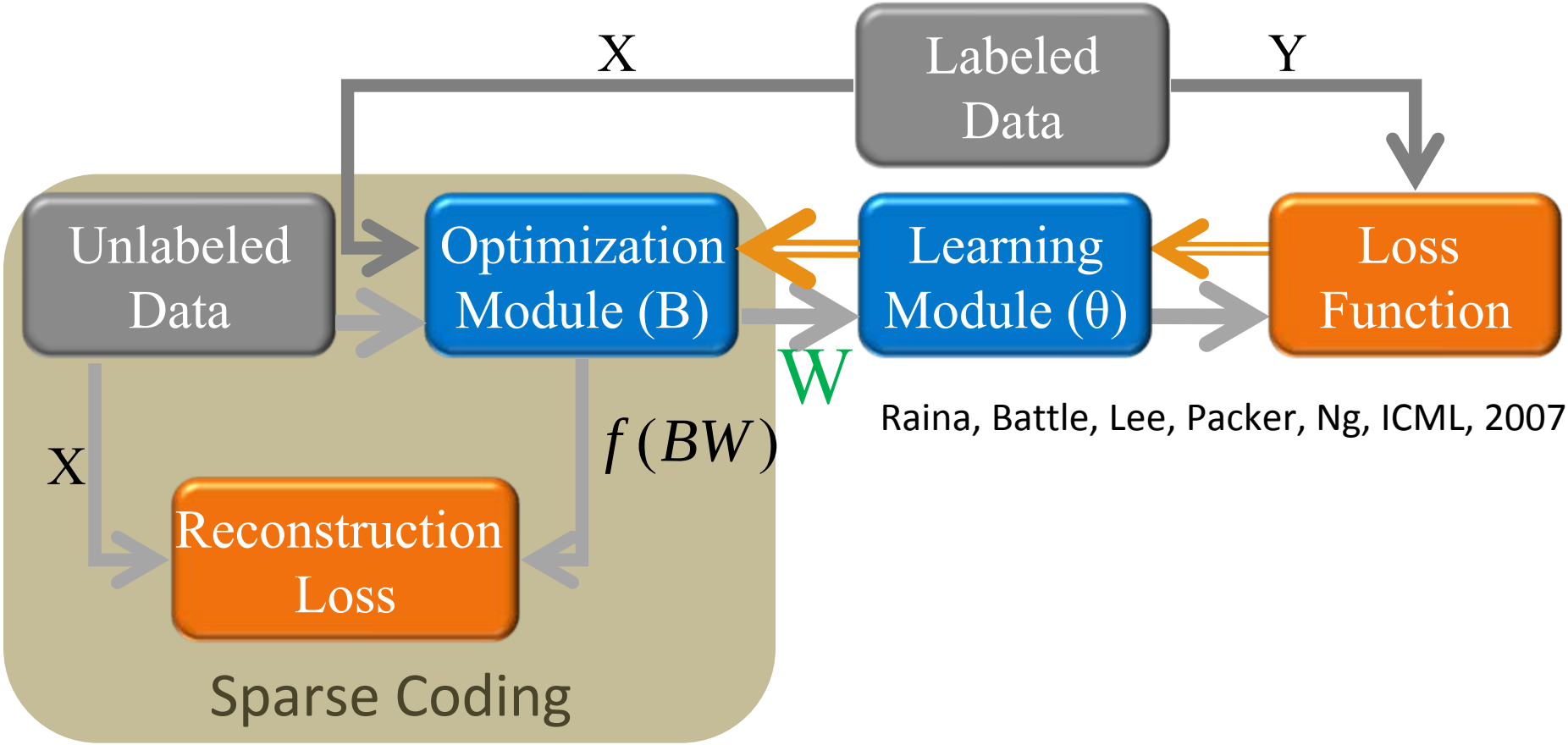
$$\arg\min_{W,B} \sum_j \frac{1}{2}\|Bw_j - x\|_2^2 + \lambda\|w_j\|_1$$

Optimize B over all examples

$$s.t. \qquad \|b_j\|_2 = 1$$

Shown to generate useful features on diverse problems

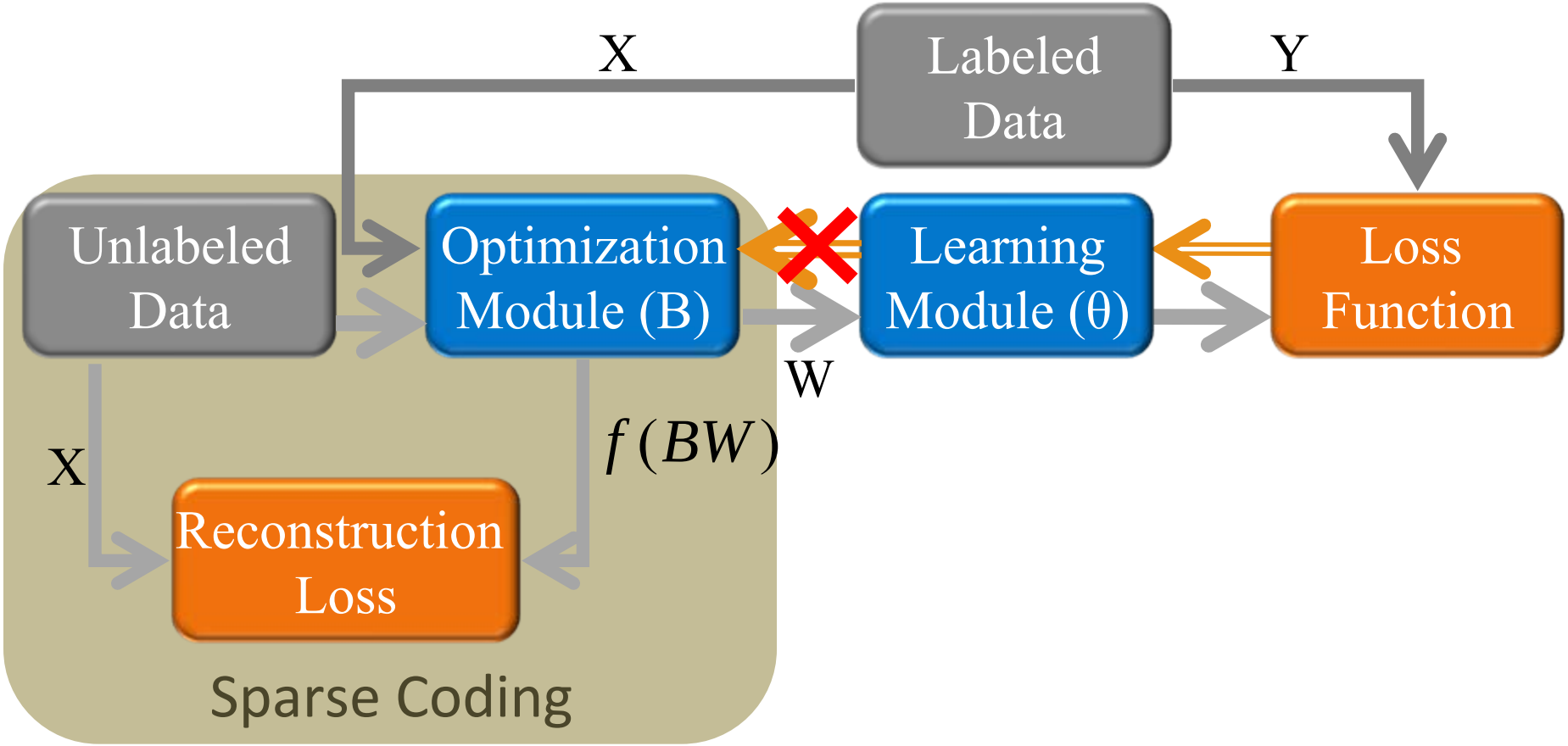
Differentiable Sparse Coding



Raina, Battle, Lee, Packer, Ng, ICML, 2007

Bradley & Bagnell, "Differentiable Sparse Coding", NIPS 2008

L1 Regularization is Not Differentiable



Bradley & Bagnell, "Differentiable Sparse Coding", NIPS 2008

Why is this unsatisfying?

Problem #1: Instability

- L1 Map Estimates are discontinuous
- Outputs are not differentiable
- Instead use KL-divergence

Proven to compete with L1 in online learning

Problem #2:
No closed-form Equation

$$\hat{w} = \arg\min_{\vec{w}} \text{Loss}\left(\vec{x} \,||\, f(B\vec{w})\right) + \lambda \text{Prior}\left(\vec{w} \,||\, \vec{p}\right)$$

At the MAP estimate:

$$\nabla_{\vec{w}} \text{Loss}\left(\vec{x} \,||\, f(B\hat{w})\right) + \lambda \nabla_{\vec{w}} \text{Prior}\left(\hat{w} \,||\, \vec{p}\right) = 0$$

Solution: Implicit Differentiation

Differentiate both sides with respect to an element of B:

$$\frac{\partial}{\partial B^{\mathrm{i}}_{\mathrm{j}}}\left(\nabla_{\vec{w}}\mathrm{Loss}\left(\vec{x}\,\|\,f(B\hat{w})\right)+\lambda\nabla_{\vec{w}}\mathrm{Prior}\left(\hat{w}\,\|\,\vec{p}\right)=0\right)$$

Since $\hat{w}$ is a function of B:

Solve for this

$$\frac{\partial}{\partial B^{i}_{j}}(\hat{w})=\frac{\partial \hat{w}}{\partial B^{i}_{j}}$$

Example: Squared Loss, KL prior

$$\hat{w} = \arg\min_{w} \frac{1}{2}\|x - Bw\|_2^2 + \lambda \sum_i w_i \log \frac{w_i}{p_i} - w_i + p_i$$

KL-Divergence

$$\frac{\partial \hat{w}}{\partial B_i^k} = -\left(B^T B + \mathrm{diag}(\frac{\lambda}{\hat{w}})\right)^{-1}\left((B^k \hat{w}_i)^T + \vec{e}_i(f(B^k \hat{w}) - x_k)\right)$$

Handwritten Digit Recognition

50,000 digit training set
10,000 digit validation set
10,000 digit test set

Handwritten Digit Recognition

Step #1:

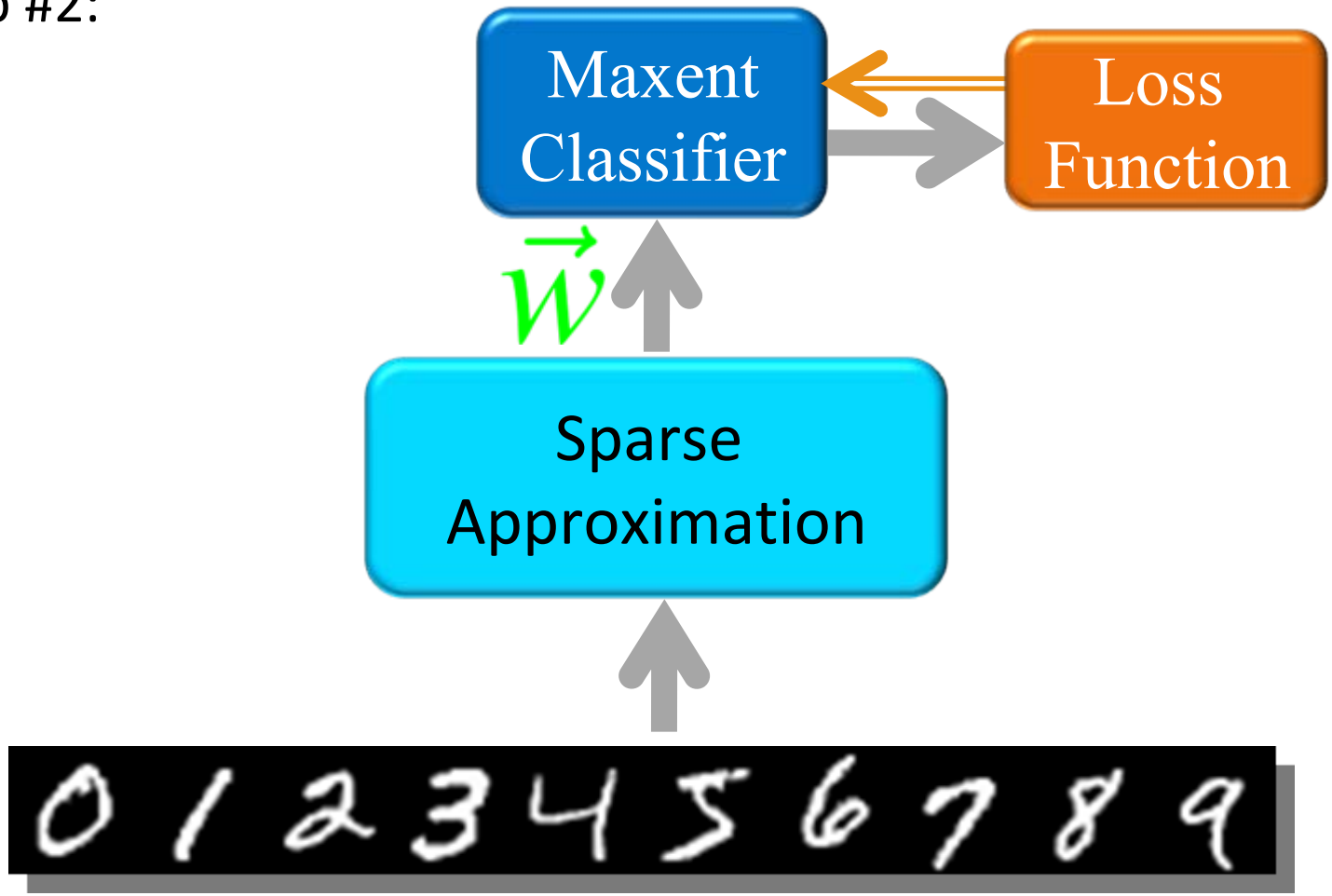
Unsupervised
Sparse Coding
L_2 loss and L_1 prior

Training Set

Raina, Battle, Lee, Packer, Ng, ICML, 2007

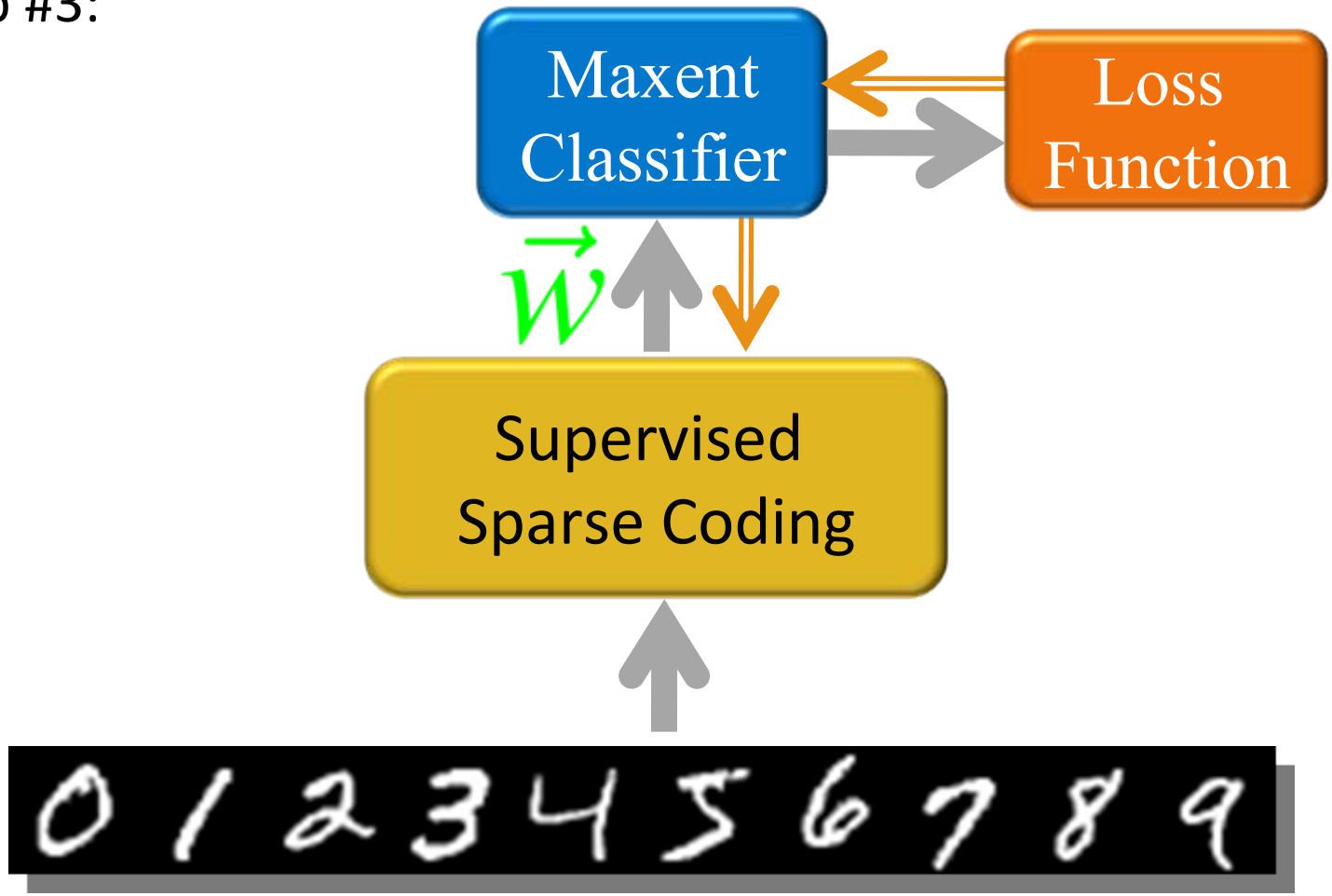
Handwritten Digit Recognition

Step #2:

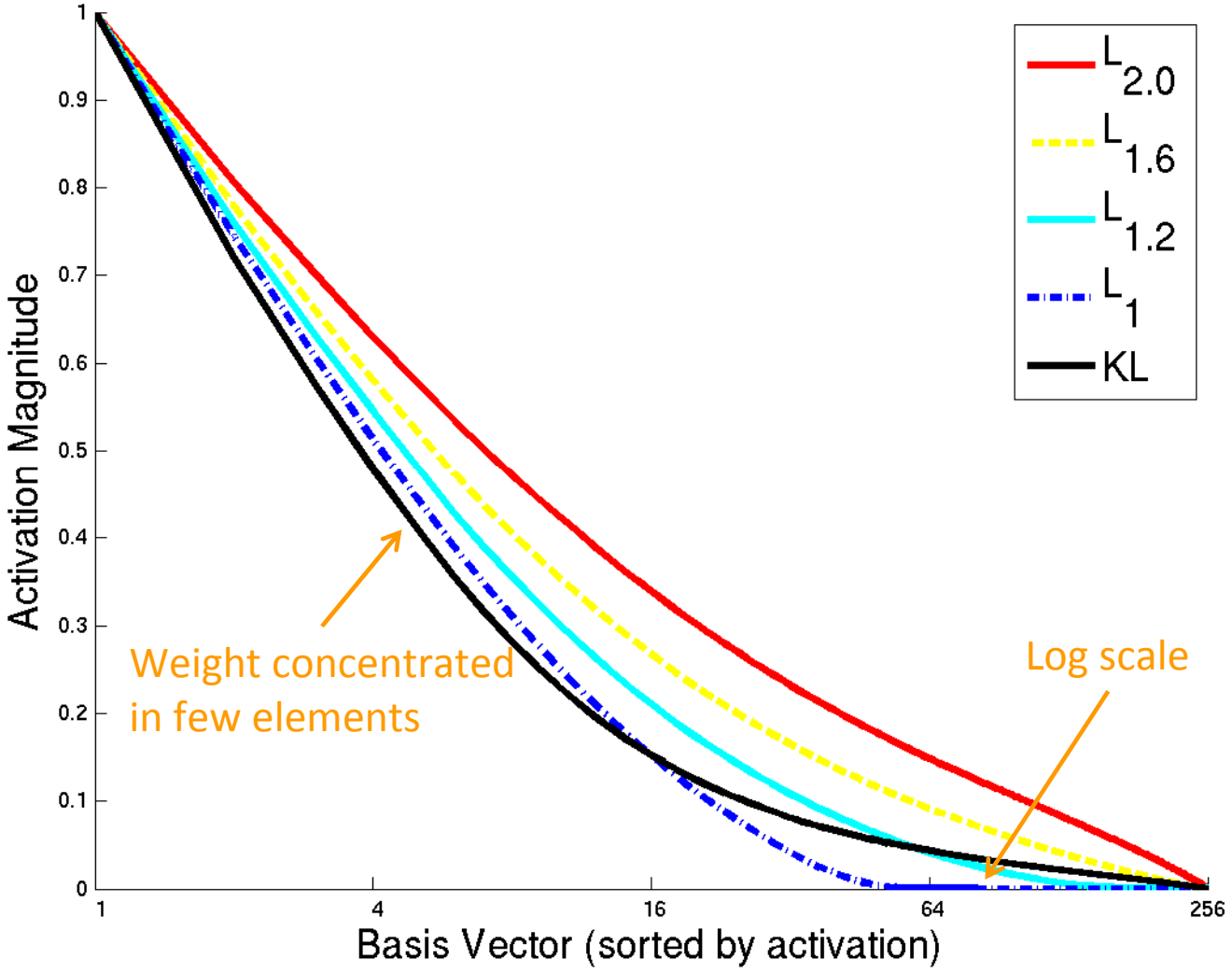


Handwritten Digit Recognition

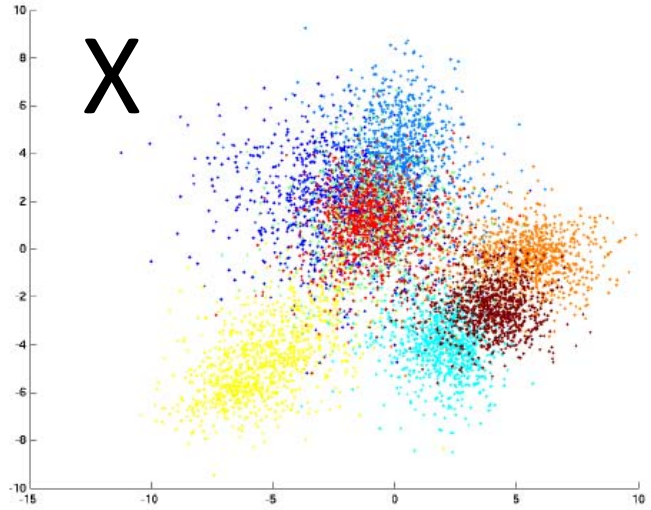
Step #3:

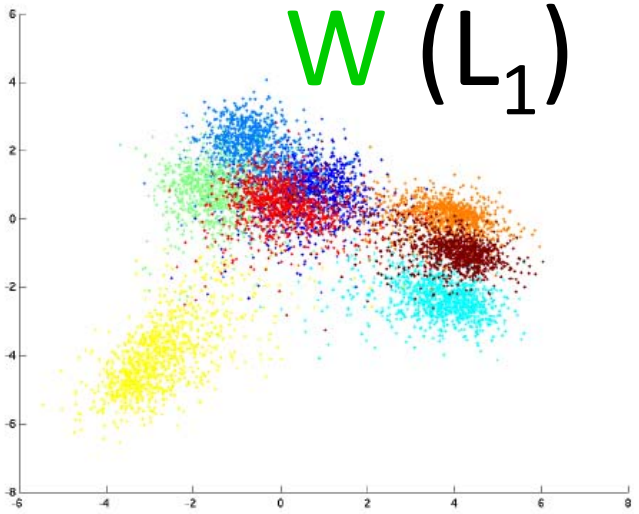


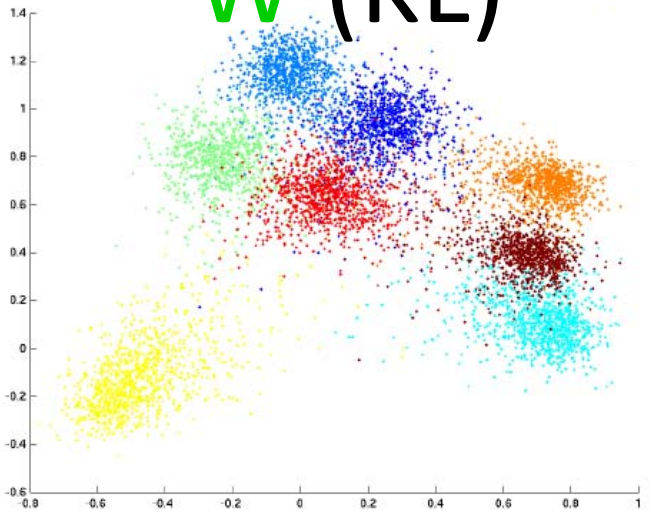
KL Maintains Sparsity

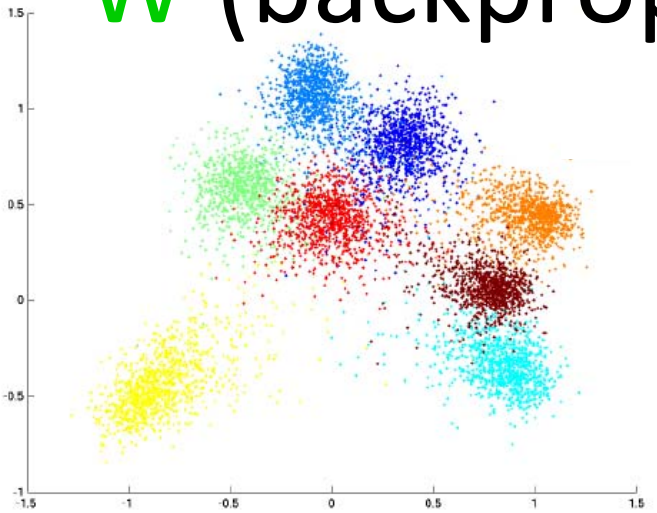


KL adds Stability



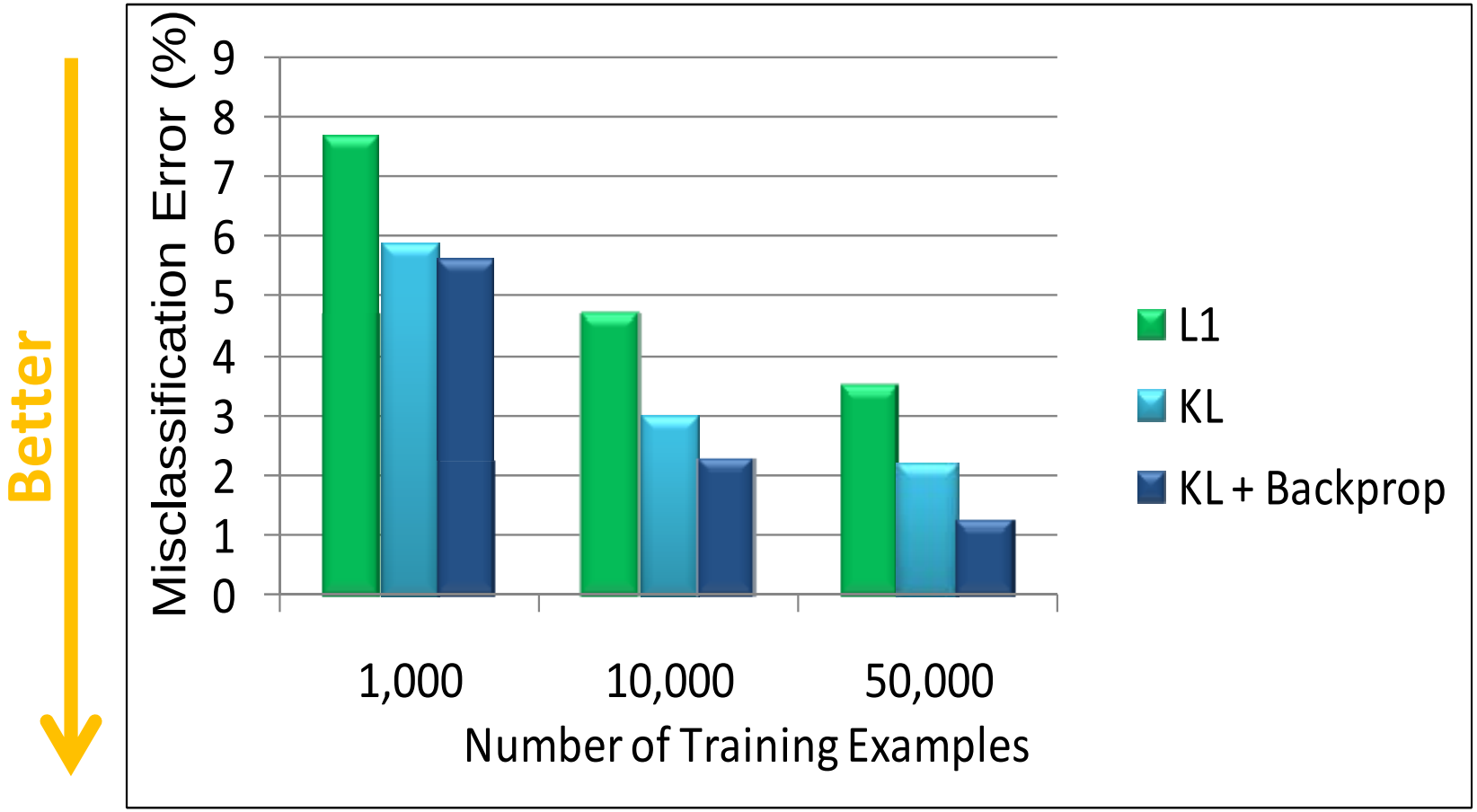




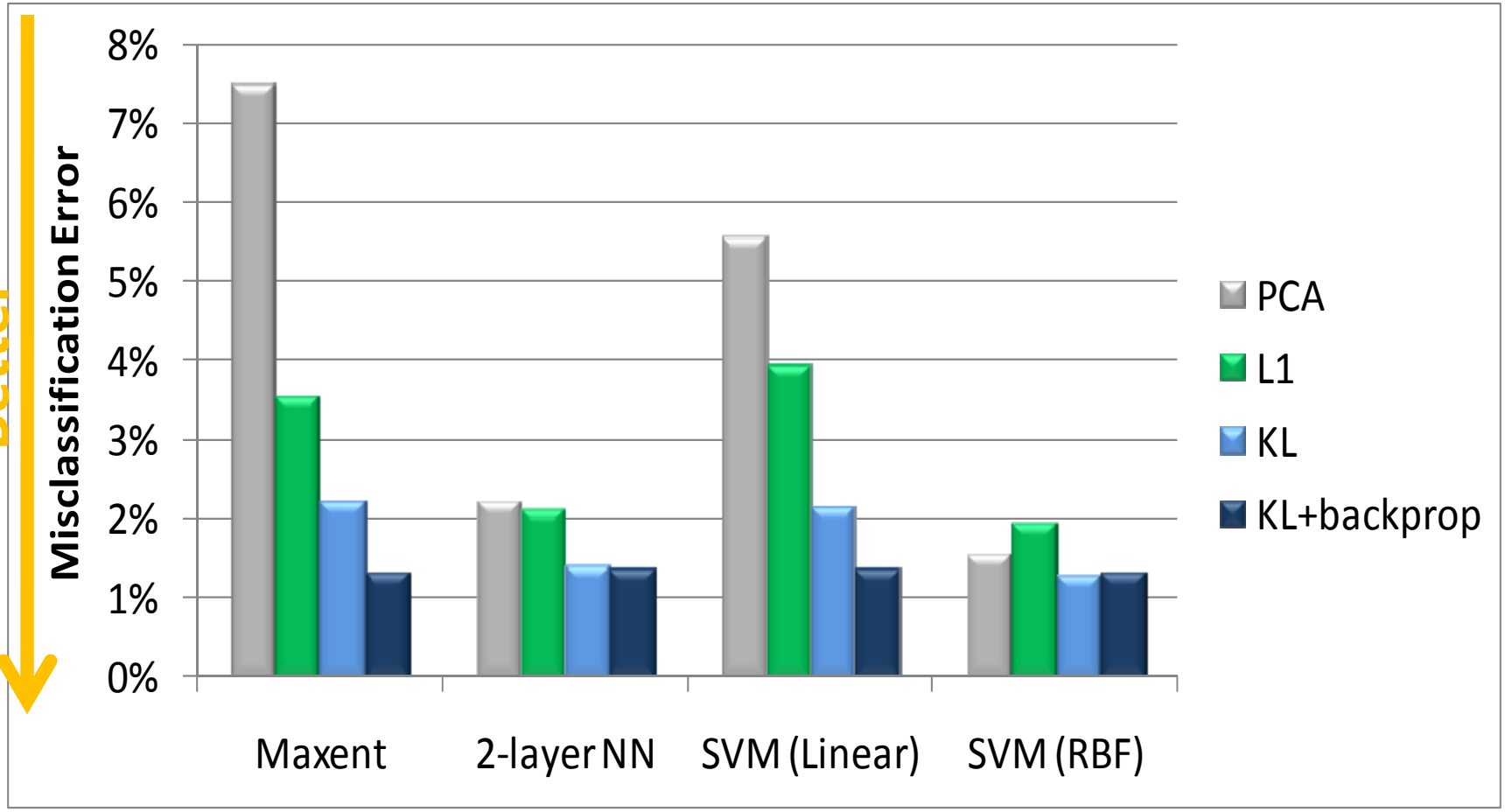


Duda, Hart, Stork. *Pattern classification.* 2001

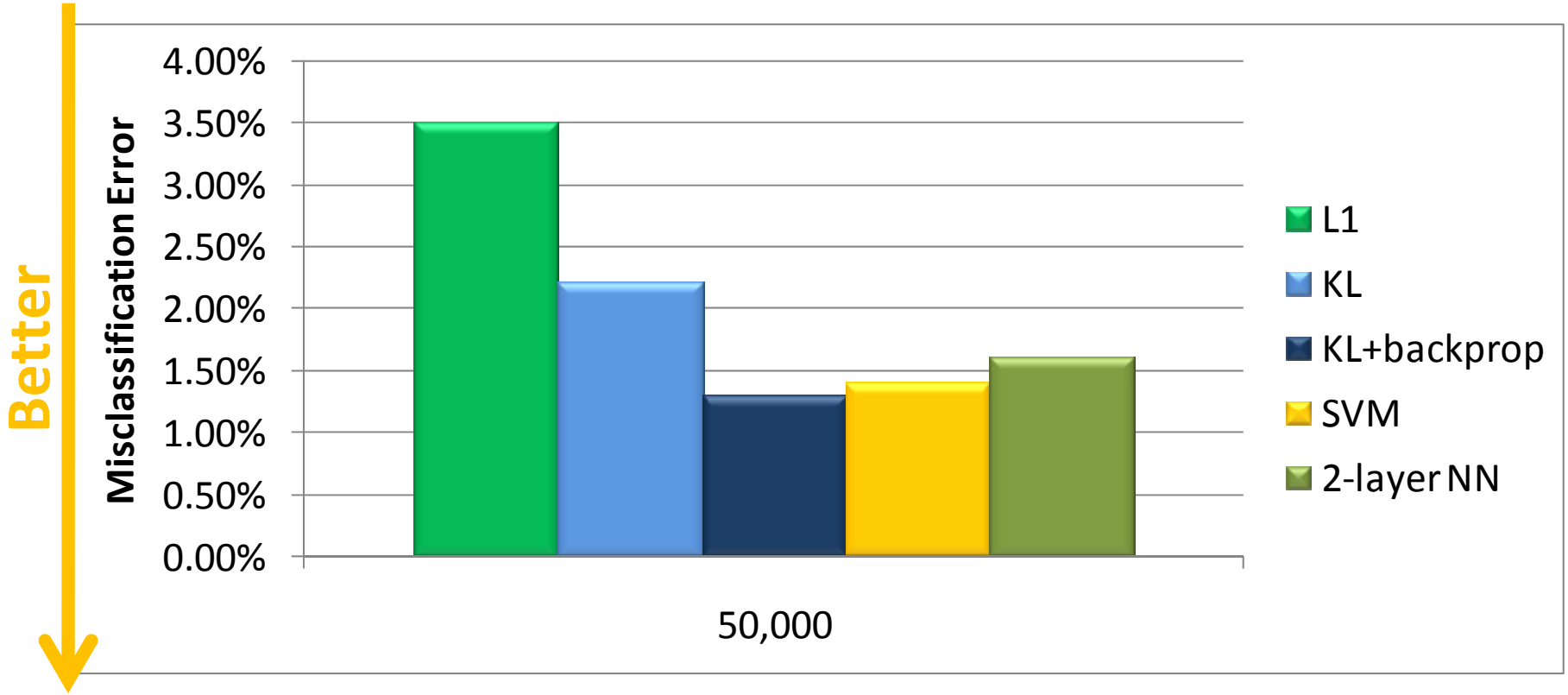
Performance vs. Prior



Classifier Comparison



Comparison to other algorithms



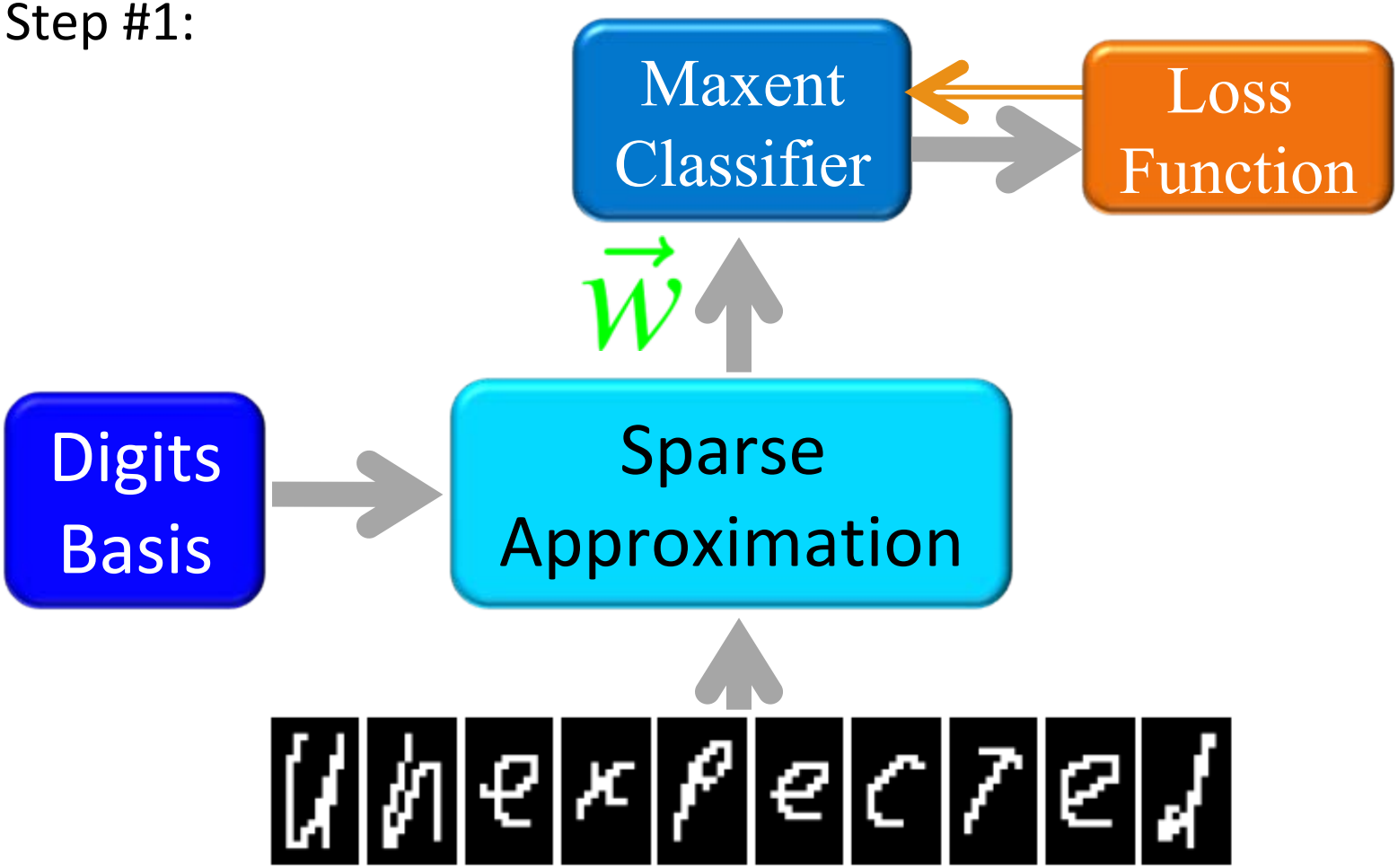
Algorithm	L1	KL	KL+backprop	SVM	2-layer NN [15]
Test Set Error	3.53%	2.21%	**1.30%**	1.4%	1.6%

Transfer to English Characters

24,000 character training set
12,000 character validation set
12,000 character test set

Transfer to English Characters

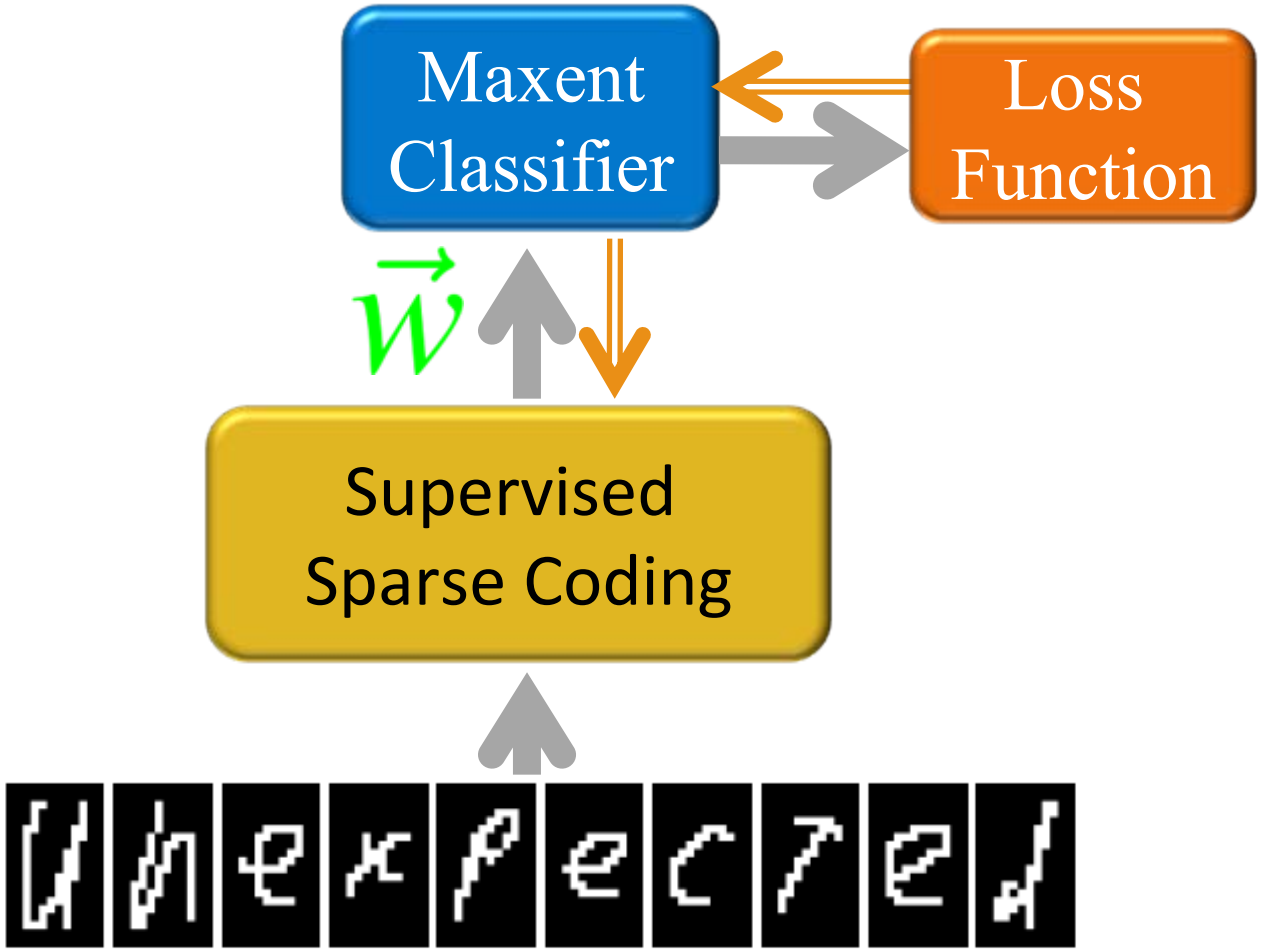
Step #1:



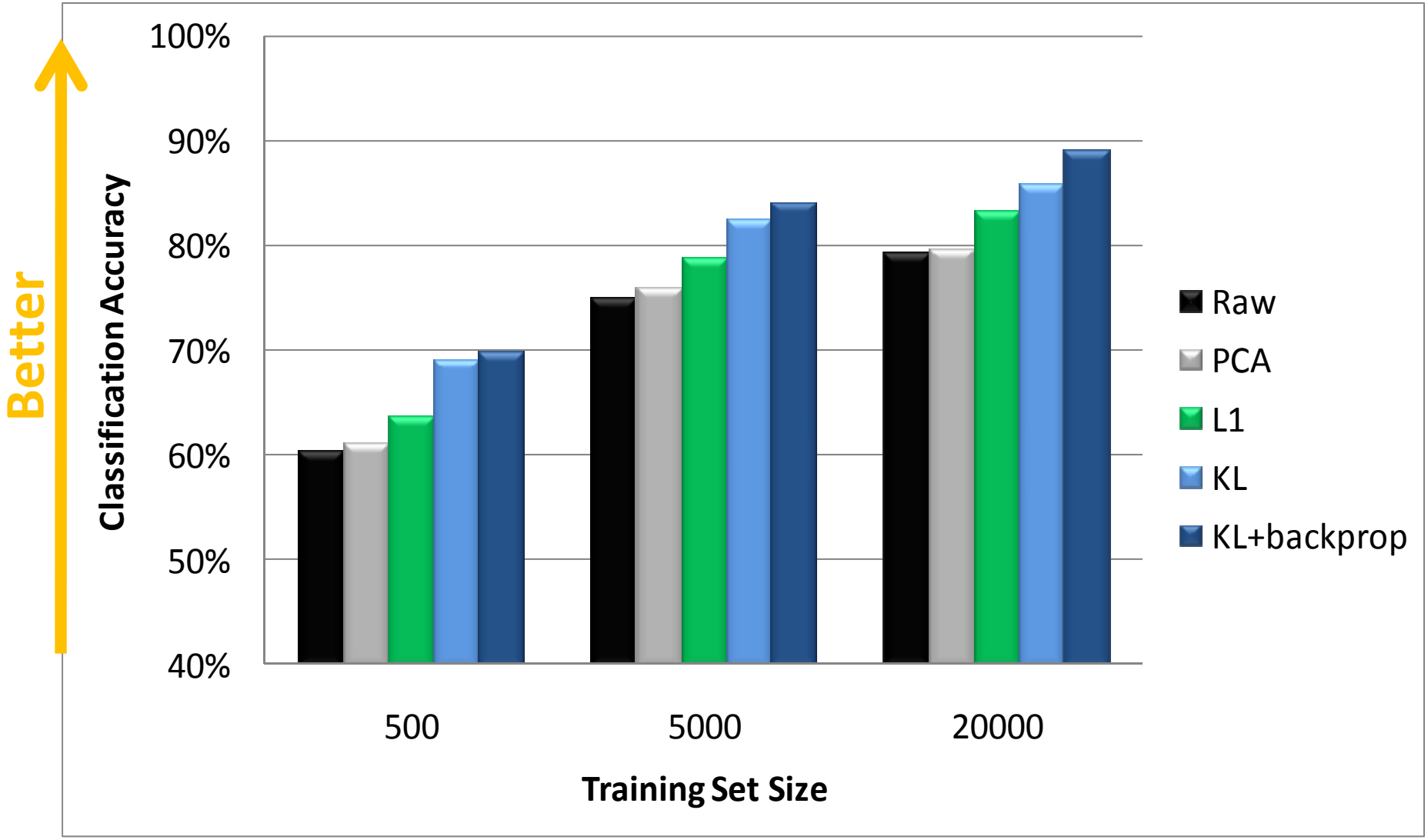
Raina, Battle, Lee, Packer, Ng, ICML, 2007

Transfer to English Characters

Step #2:



Transfer to English Characters



Text Application

Step #1:

5,000 movie reviews


Tuesday, July 5, 1977

Reviews

'Star Wars'

By BOB SMITH
Staff Writer

Not since I was a child have I experienced this phenomenon—a movie audience spontaneously cheering when the hero finally zapped the villain.

And, except for the recent film, "Rocky," I had not witnessed the other phenomenon—a spontaneous applause inside a movie theatre when the show was over.

But these things happened when I first saw "Star Wars" and again recently when I reviewed it at Longview's Cinema Theatres, where it is billed for four weeks, with a possibility of a longer stay.

There must be a reason why all ages respond as to "Star Wars," and one reason must be that movie audiences are tiring of the complicated thrillers with convoluted psychologies and no longer a distinction between heroes and villains. There is a thirst for the days when one could believe that the world really is divided between the good guys and the bad guys.

One thing "Star Wars" is not, is science fiction, although "sci-fi" buffs certainly will get a bang out of it. What it is, is a mixture of all those grand old movies of a simpler day.

What "Star Wars" is, is just pure fun.

Recipe: A bit of Wizard of Oz, a dose of the old-time Saturday matinee Western, a chapter from Buck Rogers, an idea or two

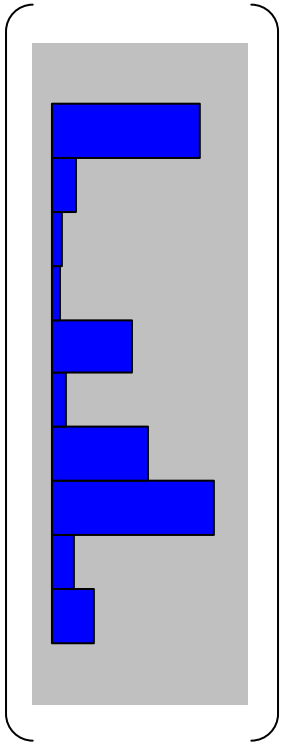
Finally, "2001: A Space Odyssey" must take a back seat—the special effects in "Star Wars" are, well, out of this world.

In spite of the simplicity of the story, there is yet another surprise—the characterizations and mood settings are superb. Especially noteworthy are Carrie Fisher (daughter of Eddie Fisher and Debbie Reynolds) as Princess Leia, Mark Hamill as Luke Skywalker, Alec Guinness as the last Jedi Knight, two comical robots and a cowardly alien.

There are many more things to be said in favor of "Star Wars" but we don't have the space here. We can promise you the best fun since Errol Flynn speared Blackbeard on his cutlass.

10 point sentiment scale
1=hated it, 10=loved it

X



Unsupervised
Sparse Coding
KL loss + KL prior

Pang, Lee, *Proceeding of the ACL*, 2005

Text Application

Step #2:

5-fold Cross Validation


Tuesday, July 5, 1977

Reviews

'Star Wars'

By BOB SMITH
Staff Writer

Not since I was a child have I experienced this phenomenon—a movie audience spontaneously cheering when the hero finally zapped the villain.

And, except for the recent film, "Rocky," I had not witnessed the other phenomenon—a spontaneous applause inside a movie theatre when the show was over.

But these things happened when I first saw "Star Wars" and again recently when I reviewed it at Longview's Cinema Theatres, where it is billed for four weeks, with a possibility of a longer stay.

There must be a reason why all ages respond as to "Star Wars," and one reason must be that movie audiences are tiring of the over-plotted thrillers with convoluted psychologies and no longer a distinction between heroes and villains. There is a thirst for the days when one could believe that the world really is divided between the good guys and the bad guys.

One thing "Star Wars" is not, in science fiction, although "sci-fi" buffs certainly will get a bang out of it. What it is, is a mixture of all those grand old movies of a simpler day.

What "Star Wars" is, is just pure fun.

Recipe: A bit of Wizard of Oz, a dose of the old-time Saturday matinee Western, a chapter from Buck Rogers, an idea or two

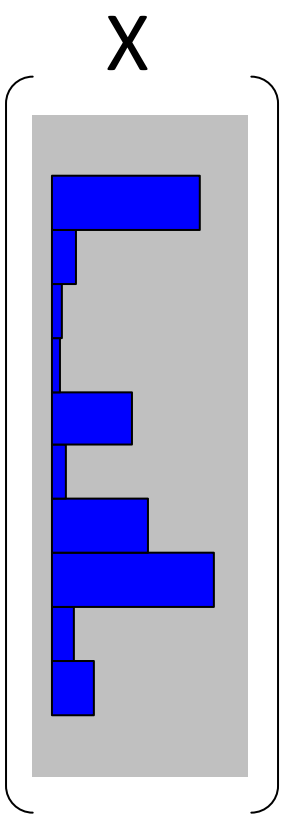
Finally, "2001: A Space Odyssey" must take a back seat—the special effects in "Star Wars" are, well, out of this world.

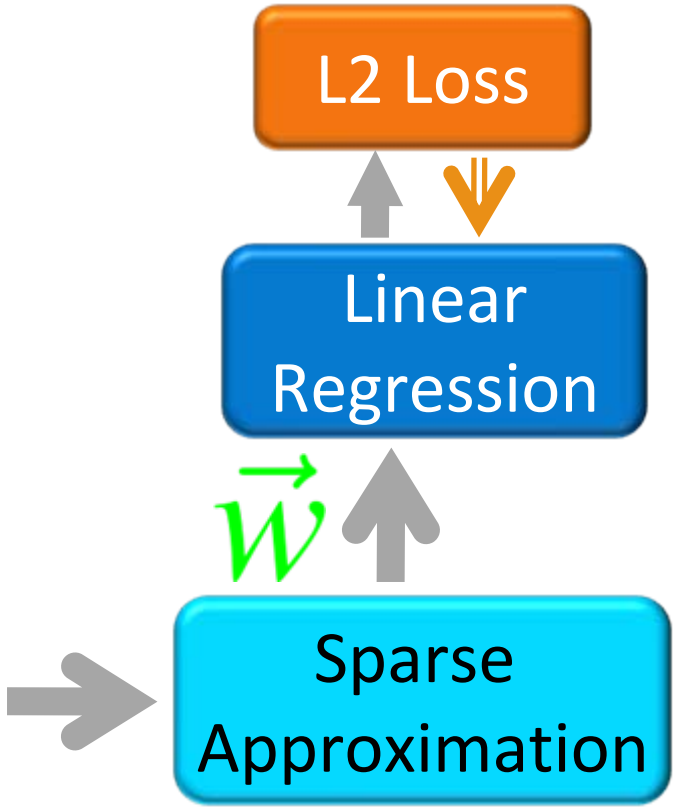
In spite of the simplicity of the story, there is yet another surprise—the characterizations and mood settings are superb. Especially noteworthy are Carrie Fisher (daughter of Eddie Fisher and Debbie Reynolds) as Princess Leia, Mark Hamill as Luke Skywalker, Alec Guinness as the last Jedi Knight, two comical robots and a cowardly alien.

There are many more things to be said in favor of "Star Wars" but we don't have the space here. We can promise you the best fun since Errol Flynn speared Blackbeard on his cutlass.

10 point sentiment scale

1=hated it, 10=loved it





Text Application

Step #3:

5-fold Cross Validation


Tuesday, July 5, 1977

Reviews

'Star Wars'

By BOB SMITH
Staff Writer

Not since I was a child have I experienced this phenomenon—a movie audience spontaneously cheering when the hero finally zapped the villain.

And, except for the recent film, "Rocky," I had not witnessed the other phenomenon—a spontaneous applause inside a movie theatre when the show was over.

But these things happened when I first saw "Star Wars" and again recently when I reviewed it at Longview's Cinema Theatres, where it is billed for four weeks, with a possibility of a longer stay.

There must be a reason why all ages respond as to "Star Wars," and one reason must be that movie audiences are tiring of the complicated thrillers with convoluted psychologies and no longer a distinction between heroes and villains. There is a thirst for the days when one could believe that the world really is divided between the good guys and the bad guys.

One thing "Star Wars" is not, is science fiction, although "sci-fi" buffs certainly will get a bang out of it. What it is, is a mixture of all those grand old movies of a simpler day.

What "Star Wars" is, is just pure fun.

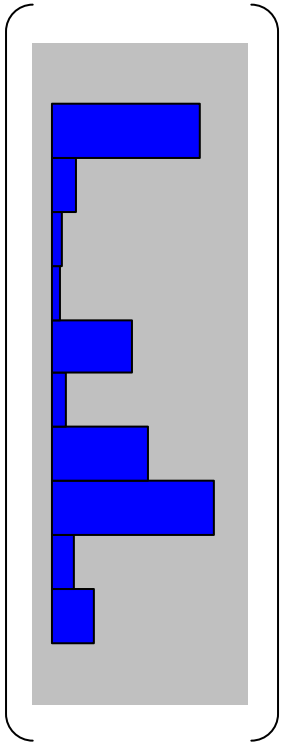
Recipe: A bit of Wizard of Oz, a dose of the old-time Saturday matinee Western, a chapter from Buck Rogers, an idea or two

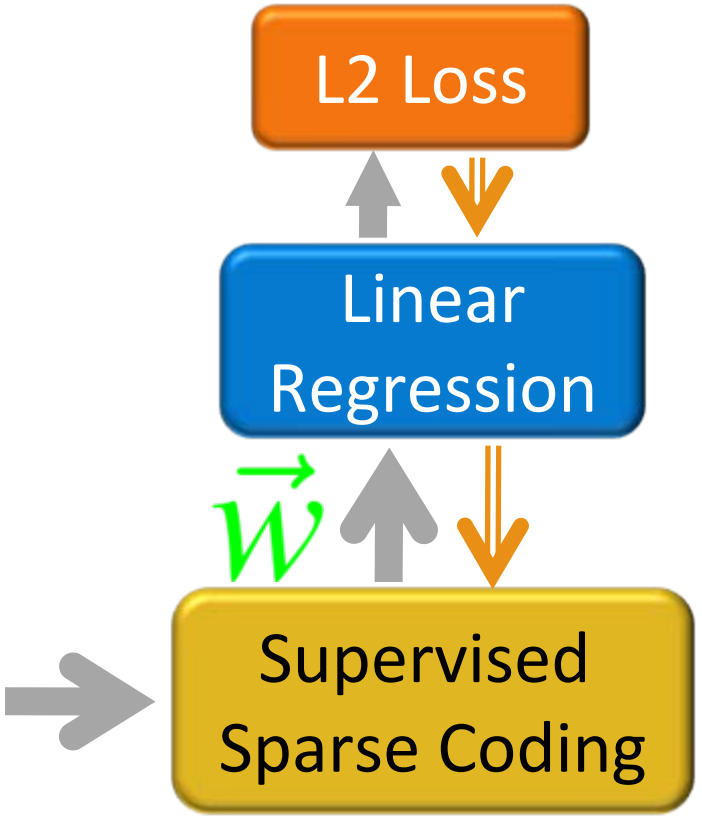
Finally, "2001: A Space Odyssey" must take a back seat—the special effects in "Star Wars" are, well, out of this world.

In spite of the simplicity of the story, there is yet another surprise—the characterizations and mood settings are superb. Especially noteworthy are Carrie Fisher (daughter of Eddie Fisher and Debbie Reynolds) as Princess Leia, Mark Hamill as Luke Skywalker, Alec Guinness as the last Jedi Knight, two comical robots and a cowardly alien.

There are many more things to be said in favor of "Star Wars" but we don't have the space here. We can promise you the best fun since Errol Flynn speared Blackbeard on his cutlass.

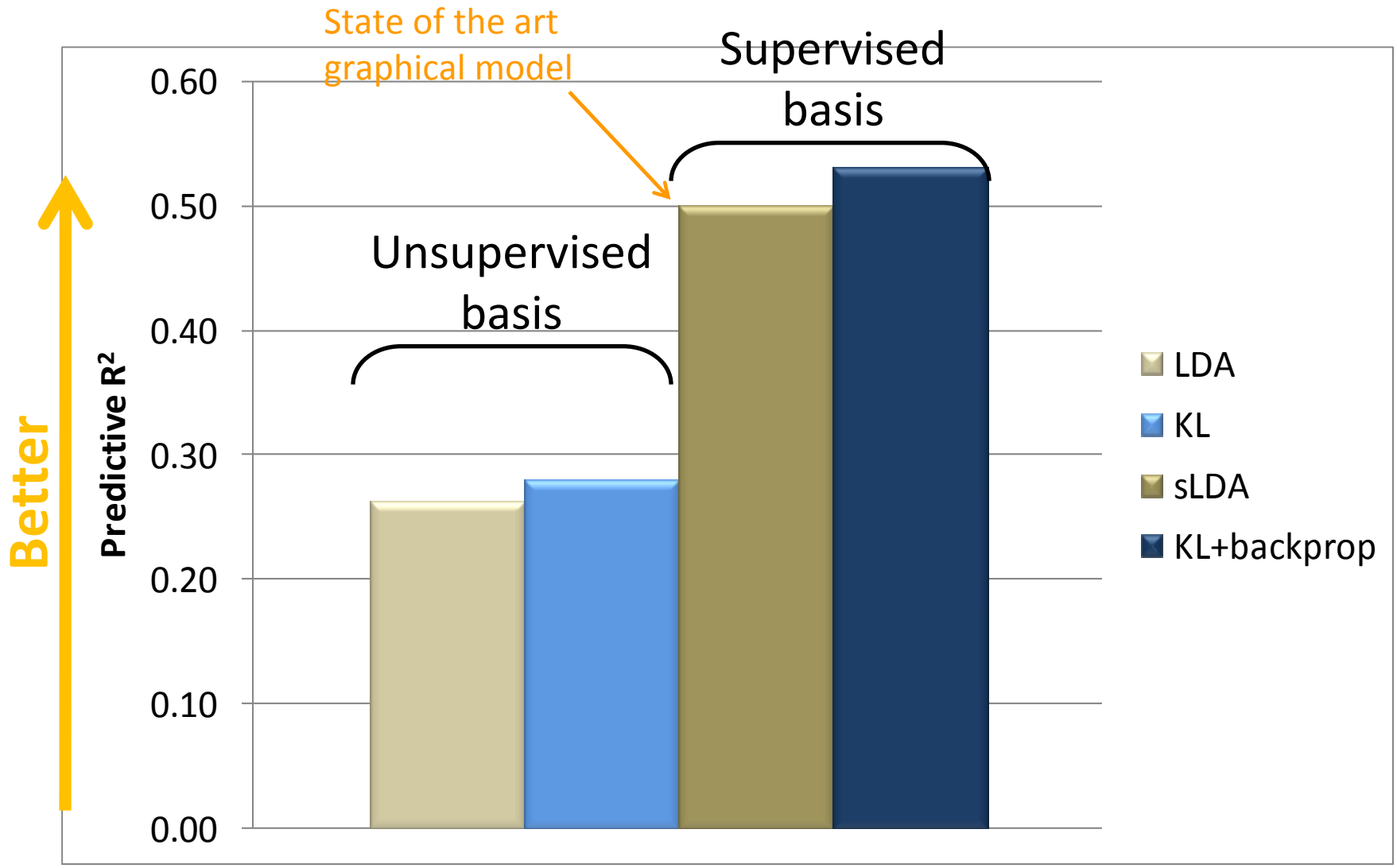
X





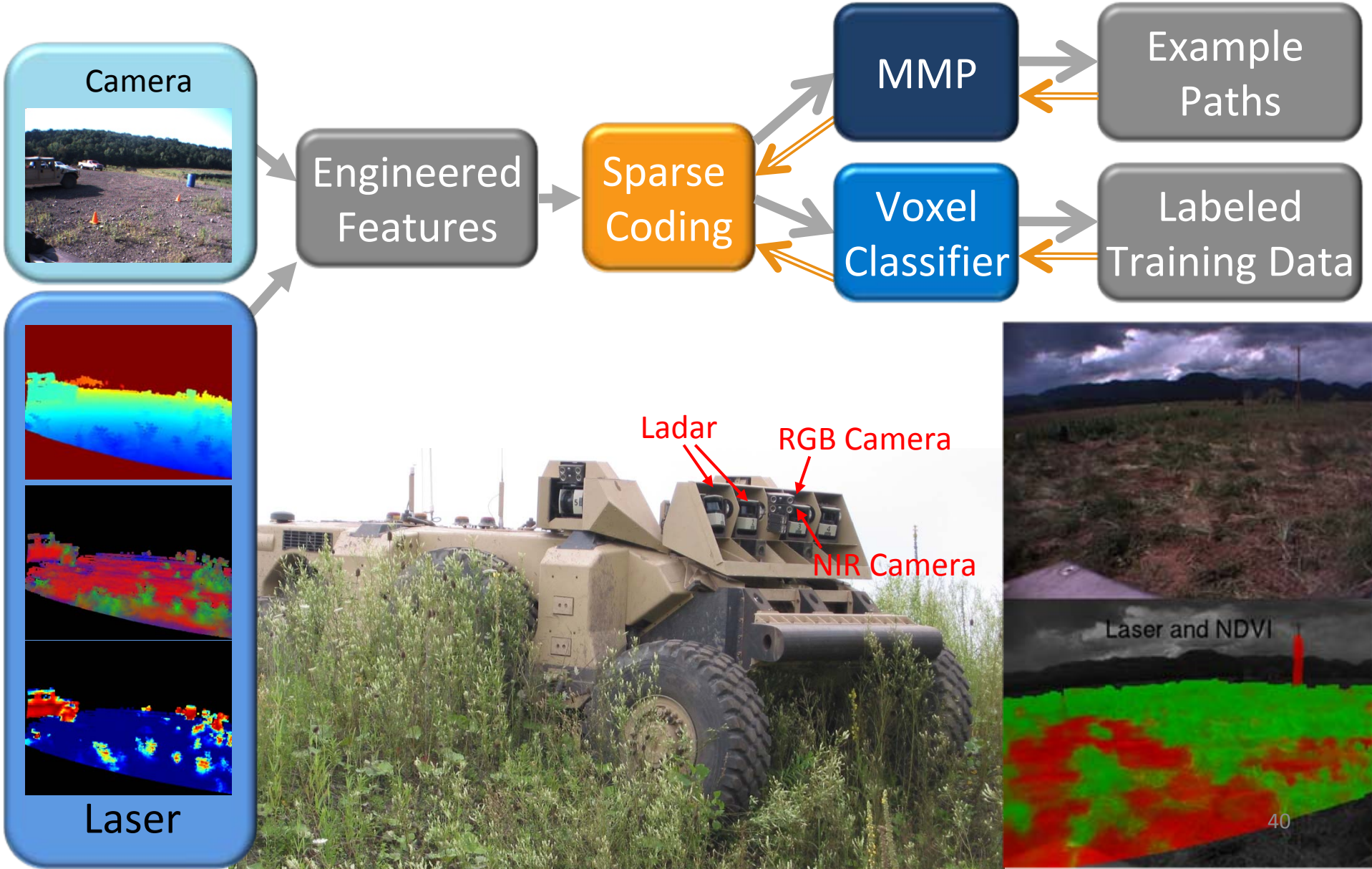
10 point sentiment scale
1=hated it, 10=loved it

Movie Review Sentiment



Blei, McAuliffe, NIPS, 2007

Future Work


Camera
Laser
Engineered Features
Sparse Coding
MMP
Voxel Classifier
Example Paths
Labeled Training Data
Ladar
RGB Camera
NIR Camera
Laser and NDVI

Future Work:
Convex Sparse Coding

- Sparse approximation is convex
- Sparse coding is not because fixed-size basis is a non-convex constraint
- Sparse coding $\leftrightarrow$ sparse approximation on infinitely large basis + non-convex rank constraint
  - Relax to a convex L_1 rank constraint
- Use boosting for sparse approximation directly on infinitely large basis

Bengio, Le Roux, Vincent, Dellalleau, Marcotte, NIPS, 2005

Zhao, Yu. *Feature Selection for Data Mining*, 2005

Riffkin, Lippert. *Journal of Machine Learning Research*, 2007

Questions?