DEEP LEARNING

Philipp Krähenbühl

ADMINISTRATIVE

- Presentation list (soon) online

THE DEEP LEARNING CONSPIRACY

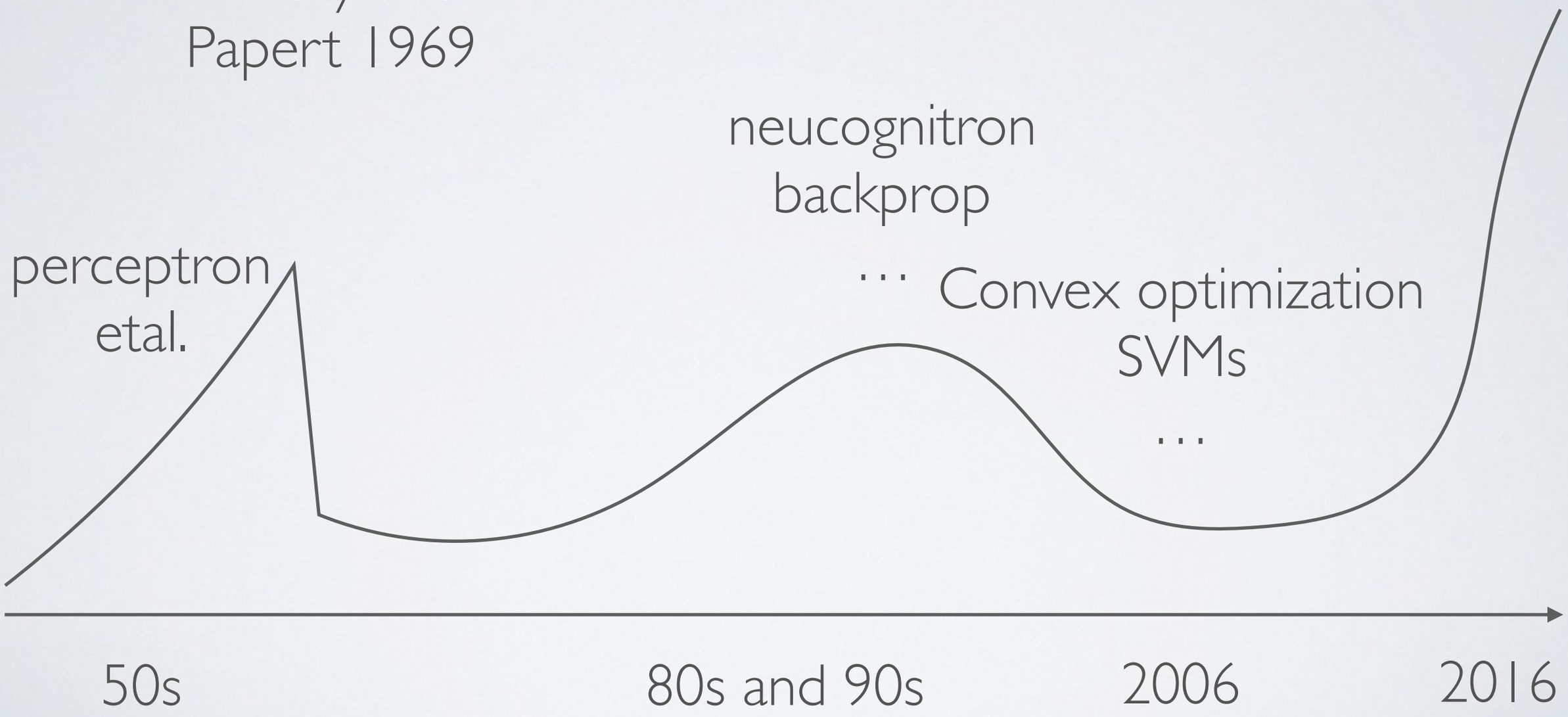
REVIEW

doi:10.1038/nature14539

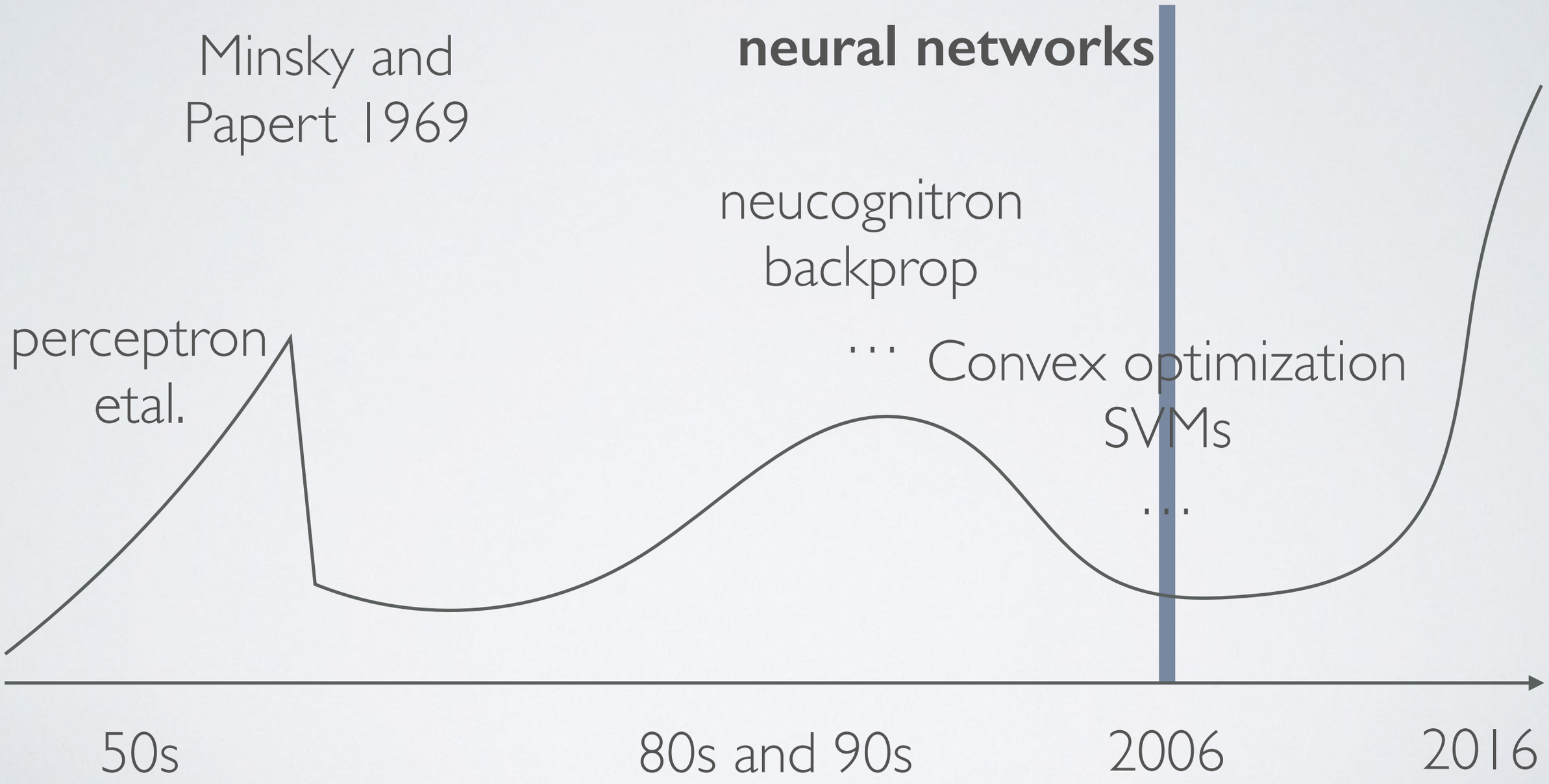
Deep learning

Yann LeCun[1,2], Yoshua Bengio[3] & Geoffrey Hinton[4,5]

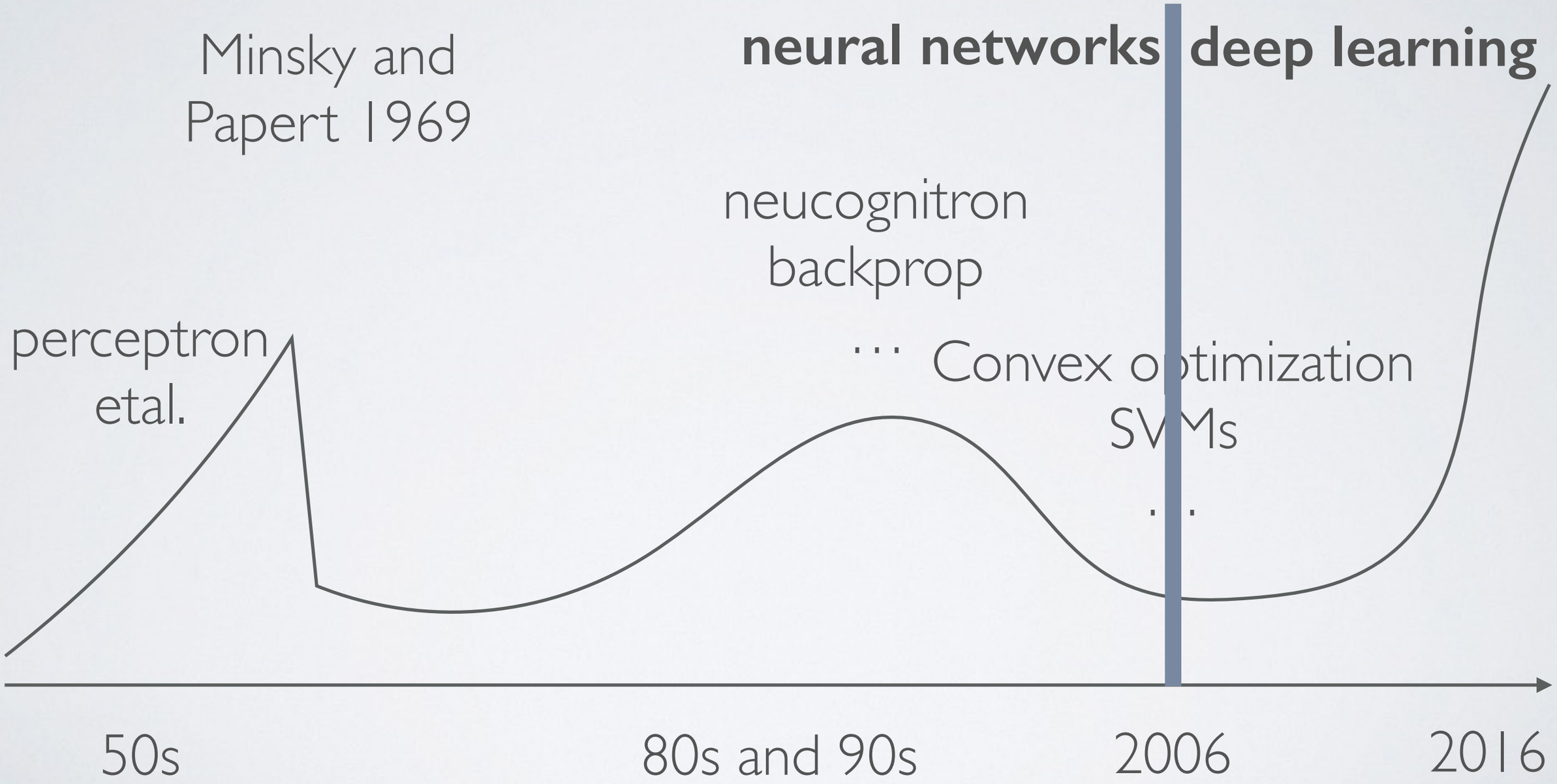
THE DEEP LEARNING CONSPIRACY



THE DEEP LEARNING CONSPIRACY



THE DEEP LEARNING CONSPIRACY



THE DEEP LEARNING CONSPIRACY

neural networks

deep learning

Convex optimization
SVMs
…

GPUs

Dropout

ReLU

2006

2016

THE DEEP LEARNING CONSPIRACY

30 days of = x days of

THE DEEP LEARNING CONSPIRACY

30 c ars of

THE DEEP LEARNING CONSPIRACY

30 days of = x years of

THE DEEP LEARNING CONSPIRACY

30 days of = 26 years of

WHY DEEP LEARNING?

traditional models

deep models

learned

(linear) classification

learned jointly

hand designed

Feature extraction

learned jointly

DESIGNING FEATURES IS HARD

SHALLOW OR LINEAR CLASSIFIERS NOT ENOUGH

A linear classifier, or any other 'shallow' classifier operating on raw pixels could not possibly distinguish the latter two, while putting the former two in the same category.

SHALLOW OR LINEAR CLASSIFIERS NOT ENOUGH

```
from sklearn import svm
c = svm.LinearSVC()

N,D = 1000, 2000
for it in range(10):
    # Generate some random features
    f = np.random.normal(size=(2*N,D))
    l = np.arange(2*N) < N
    # How many are linearly seperable
    print( c.fit(f,l).score(f,l) )
```

SHALLOW OR LINEAR CLASSIFIERS NOT ENOUGH

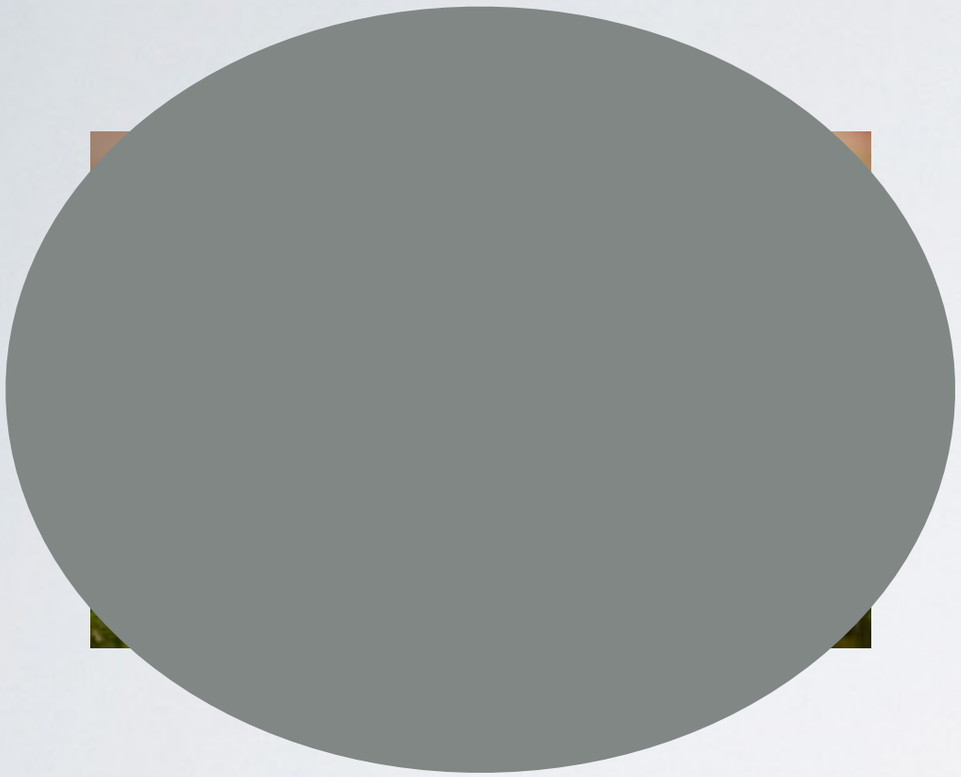
```
1.0
1.0
1.0
1.0
1.0
1.0
1.0
1.0
1.0
1.0
```

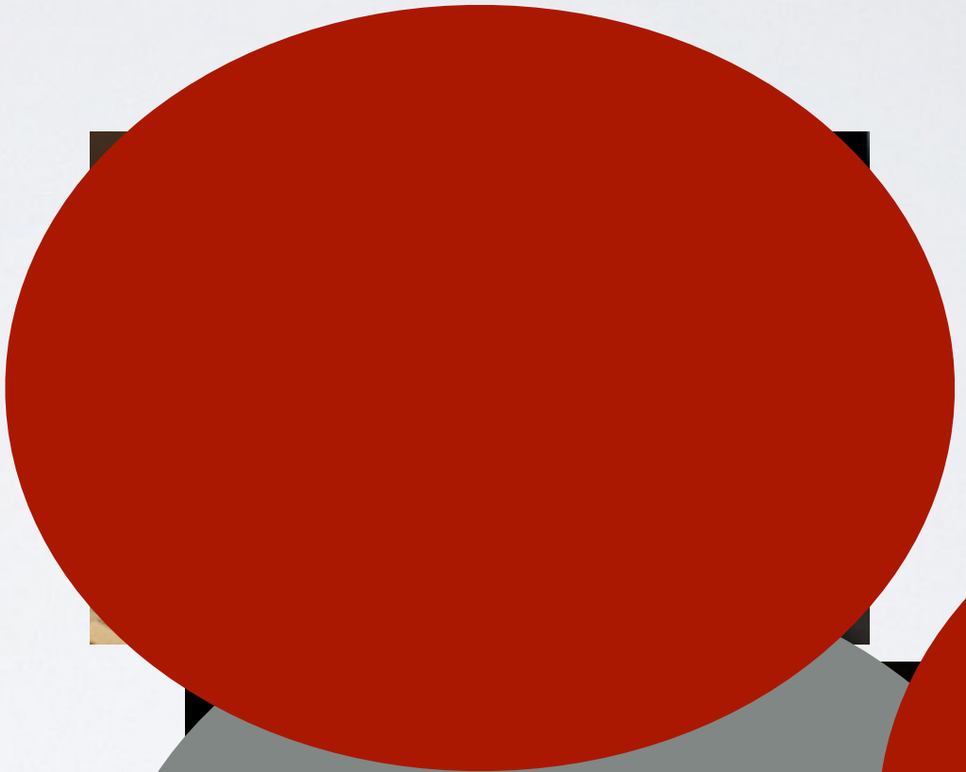
SHALLOW OR LINEAR CLASSIFIERS NOT ENOUGH

might not hold in high dimensional spaces such as images

A linear classifier, or any other 'shallow' classifier operating on raw pixels could not possibly distinguish the latter two, while putting the former two in the same category.

KERNELS





JOINTLY LEARNING THE FEATURE SPACE

WHY DEEP LEARNING?

traditional models

deep models

learned

(linear) classification

learned jointly

hand designed

Feature extraction

learned jointly

TRAINING

parameters

(linear) classification

elementwise nonlinearity

linear transformation

elementwise nonlinearity

linear transformation

Feature extraction

TRAINING

loss $l(y)$

(linear) classification $y = A_3 z_4 + b_3$

elementwise nonlinearity $z_4 = \rho(z_3)$

linear transformation $z_3 = A_2 z_2 + b_2$

elementwise nonlinearity $z_2 = \rho(z_1)$

linear transformation $z_1 = A_1 x + b_1$

Feature extraction

$$\partial l / \partial b_3 = \partial l / \partial y \cdot \partial y / \partial b_3$$

loss $l(y)$ $\partial l / \partial y$

(linear) classification

$y = A_3 z_4 + b_3$ $\partial l / \partial z_4$

elementwise nonlinearity

$z_4 = \rho(z_3)$

linear transformation

$z_3 = A_2 z_2 + b_2$

elementwise nonlinearity

$z_2 = \rho(z_1)$

linear transformation

$z_1 = A_1 x + b_1$

Feature extraction

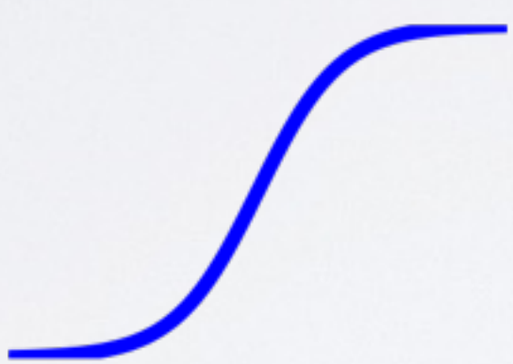
RELU

$$z_4 = \rho(z_3)$$

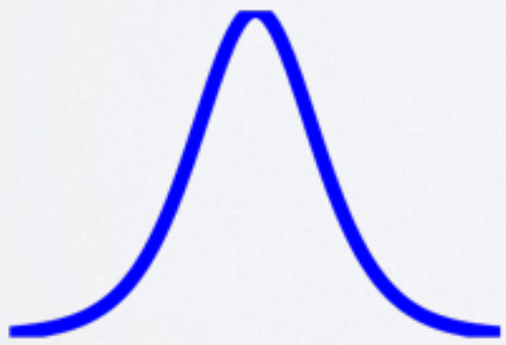
$$\partial l / \partial z_3 = \partial l / \partial z_4 \cdot \partial \rho / \partial z_3$$

sigmoid ReLU

$\partial \rho$



$\partial \rho / \partial z_3$



CONVNETS

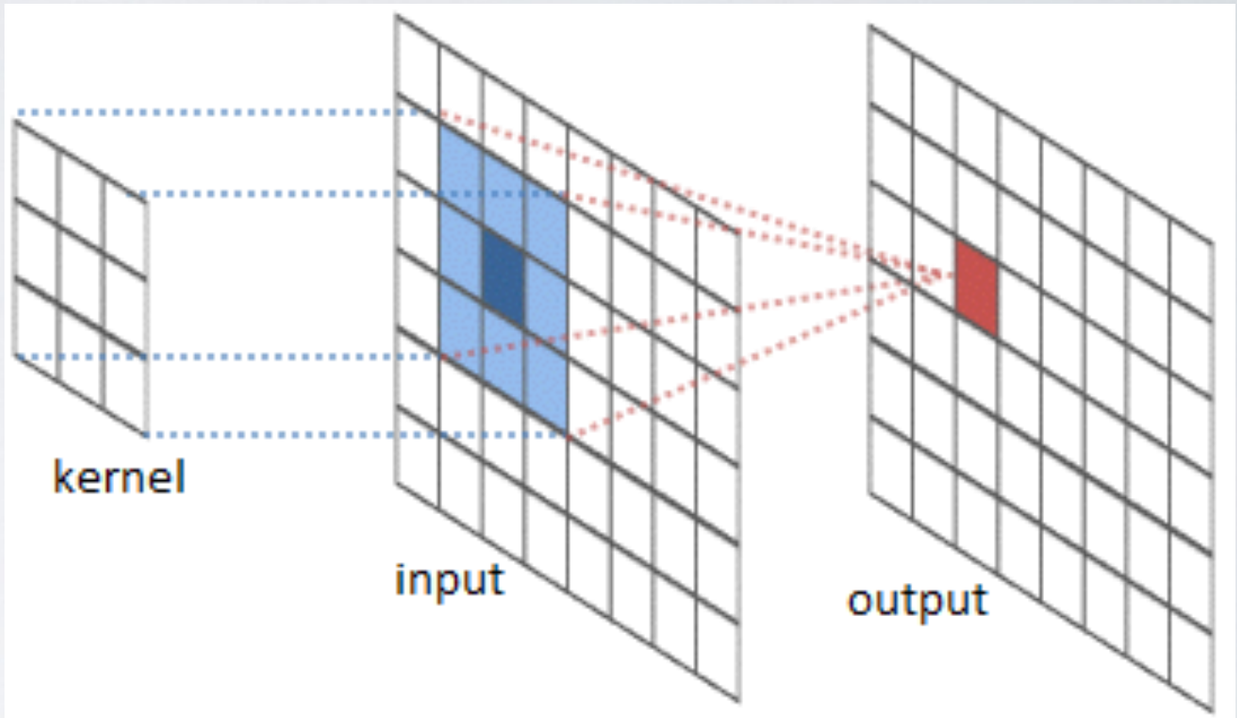
(linear) classification

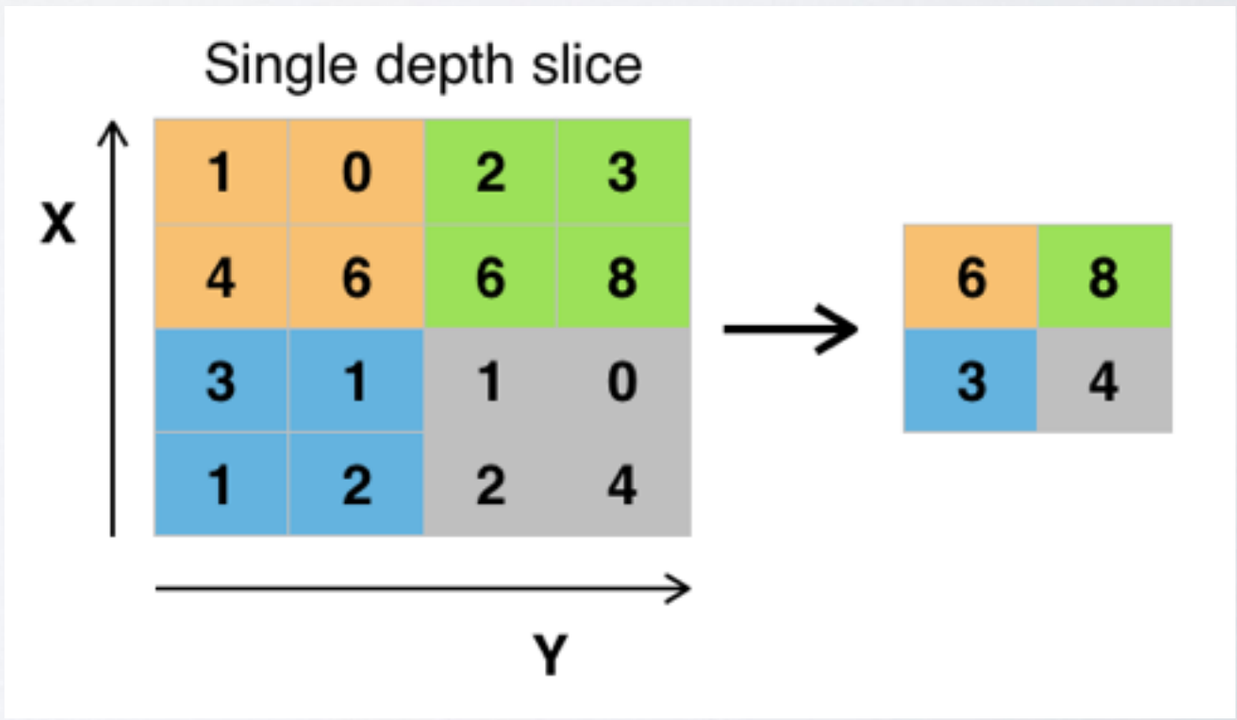
elementwise nonlinearity

linear transformation

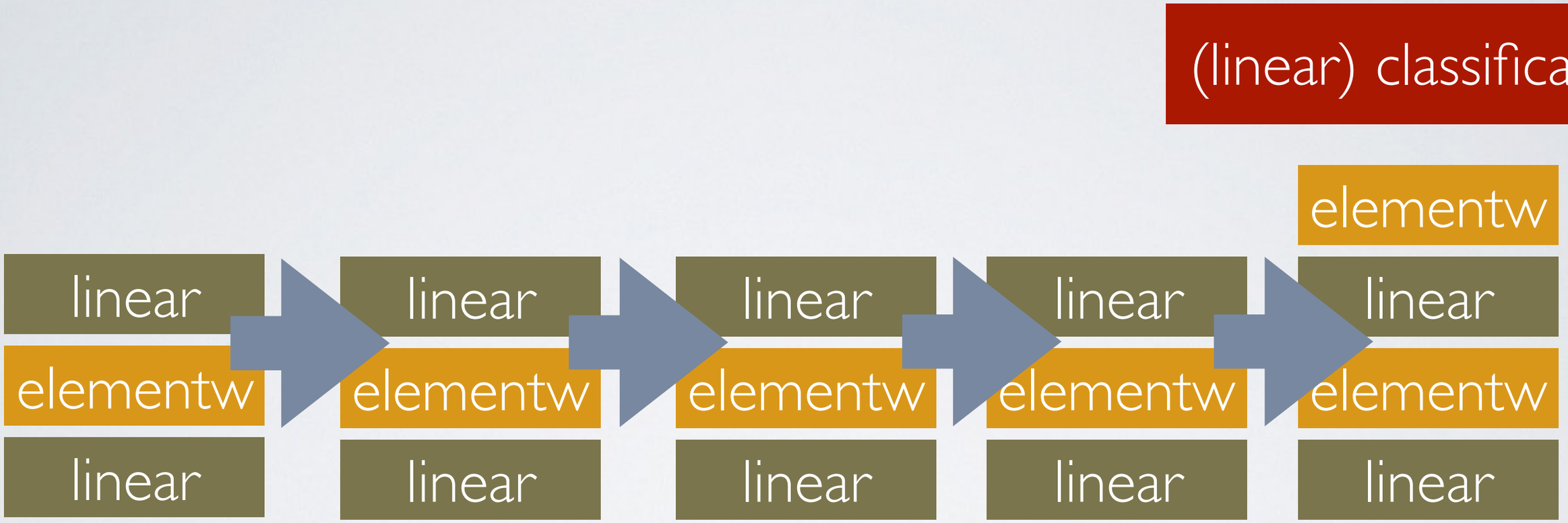
elementwise nonlinearity

linear transformation





RNNS

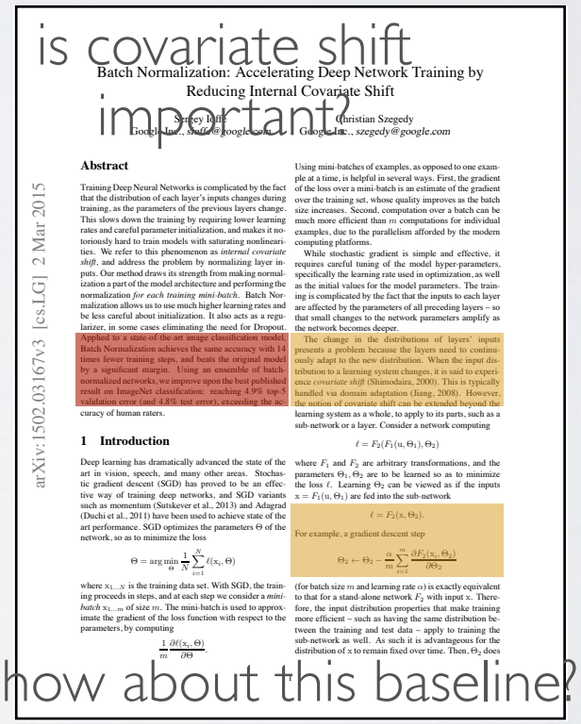

(linear) classification
elementw
linear
linear
linear
linear
linear
elementw
elementw
elementw
elementw
elementw
linear
linear
linear
linear
linear

H e l l o

DISCUSSION

HOW TO PRESENT A PAPER

- Email your slides to Philipp one week in advance
- Read and review the paper carefully



SHALLOW OR LINEAR CLASSIFIERS NOT ENOUGH

```
from sklearn import svm
c = svm.LinearSVC()

N,D = 1000, 2000
for it in range(10):
    # Generate some random features
    f = np.random.normal(size=(2*N,D))
    l = np.arange(2*N) < N
    # How many are linearly seperable
    print( c.fit(f,l).score(f,l) )
```

HAVE A STORY

- Motivate well
- Provide context


REVIEW

doi:10.1038/nature14539

Deep learning

Yann LeCun[1,2], Yoshua Bengio[3] & Geoffrey Hinton[4,5]

BE VISUAL

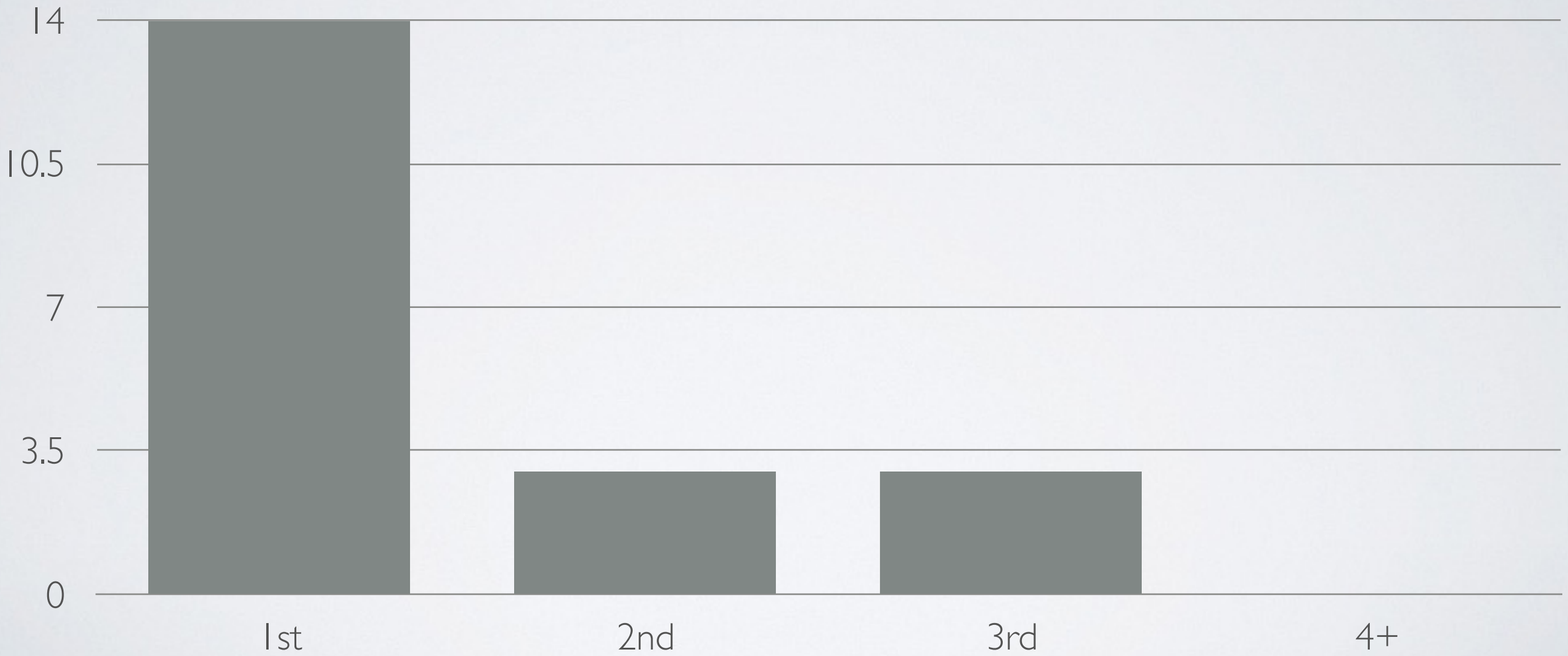
every slide should have a picture

A picture is worth a thousand words.

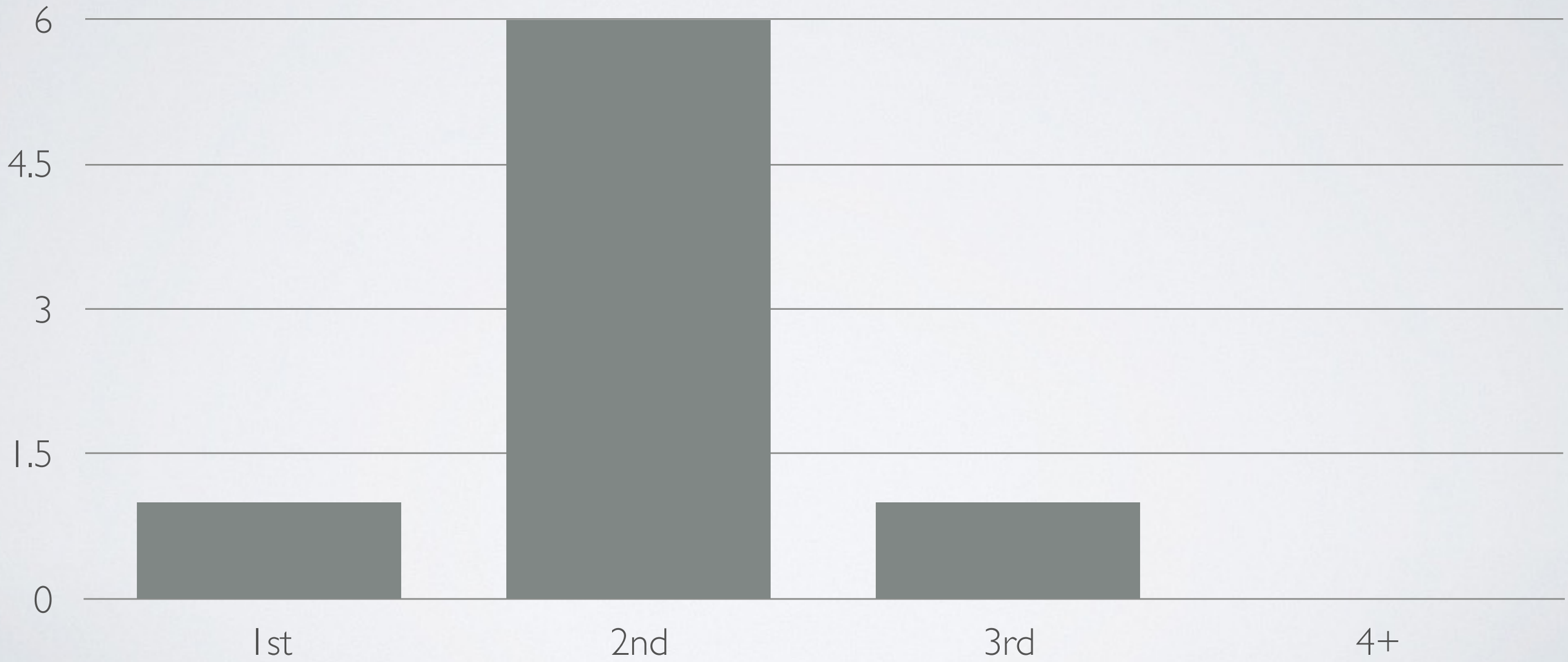
DESIGNING FEATURES IS HARD

BE INTERACTIVE

PAPER ASSIGNMENT



PAPER ASSIGNMENT



PAPER ASSIGNMENT

Posted online later today

SADDLE POINTS

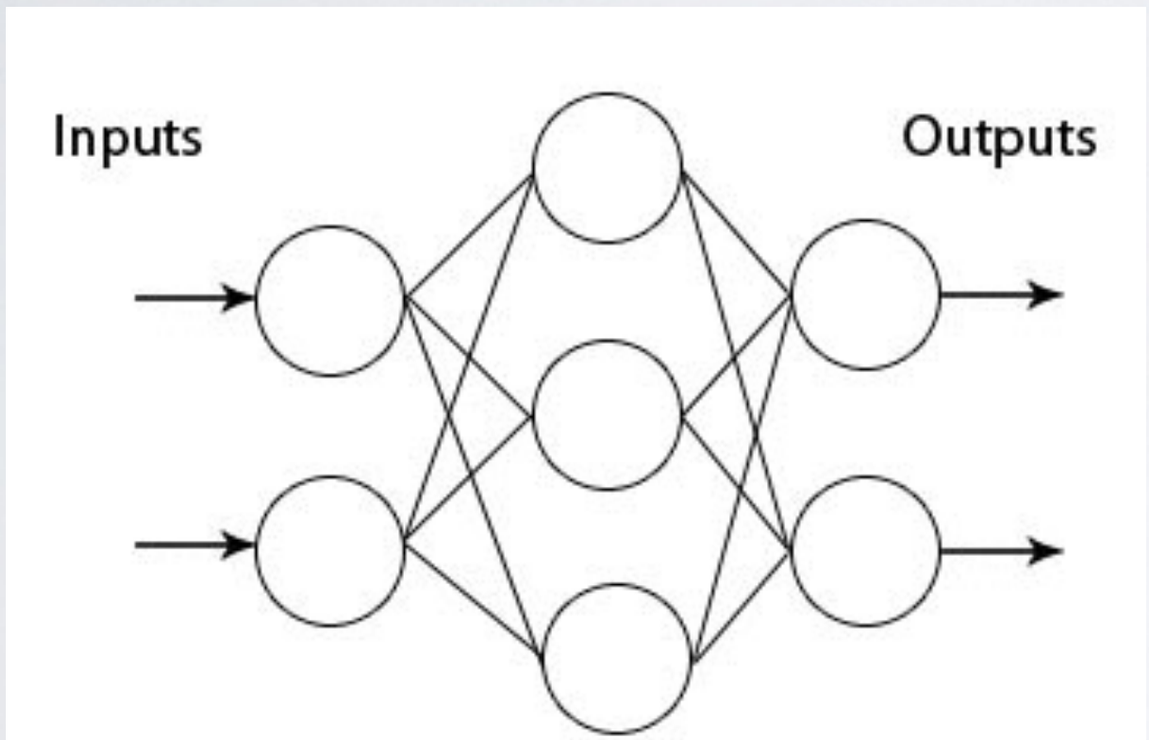
Identifying and attacking the saddle point problem in high-dimensional non-convex optimization

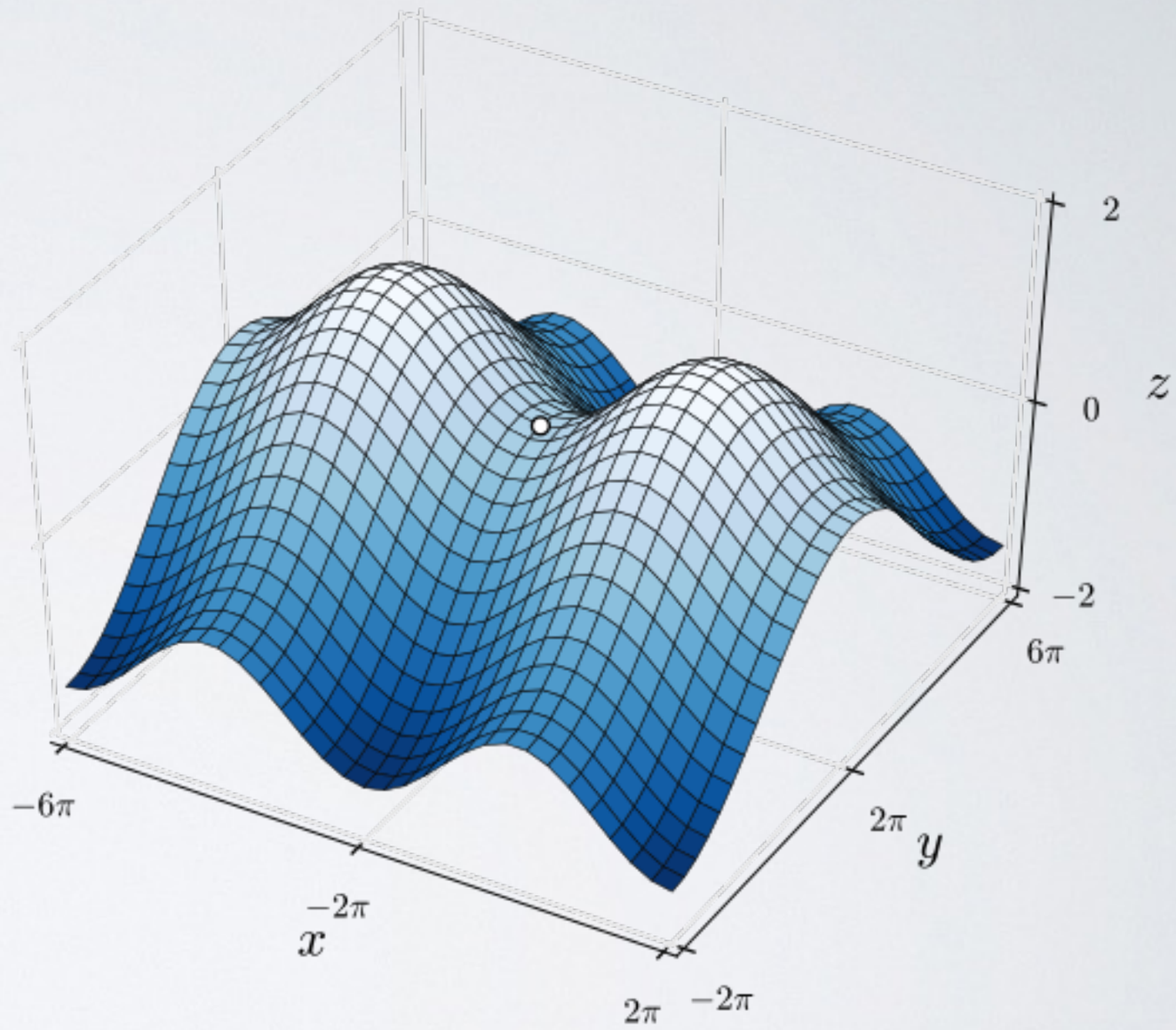
Yann N. Dauphin **Razvan Pascanu** **Caglar Gulcehre** **Kyunghyun Cho**
Université de Montréal
dauphiya@iro.umontreal.ca, r.pascanu@gmail.com,
gulcehrc@iro.umontreal.ca, kyunghyun.cho@umontreal.ca

Surya Ganguli
Stanford University
sganguli@standford.edu

Yoshua Bengio
Université de Montréal, CIFAR Fellow
yoshua.bengio@umontreal.ca

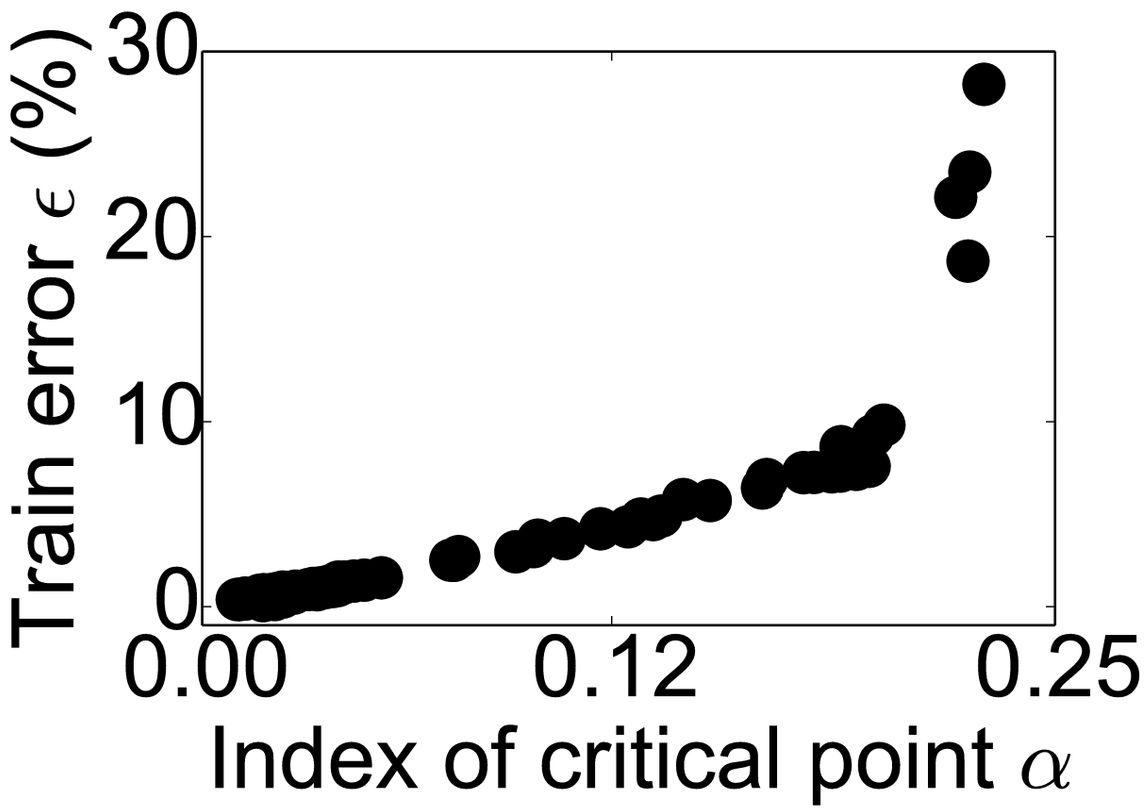
SADDLE POINTS



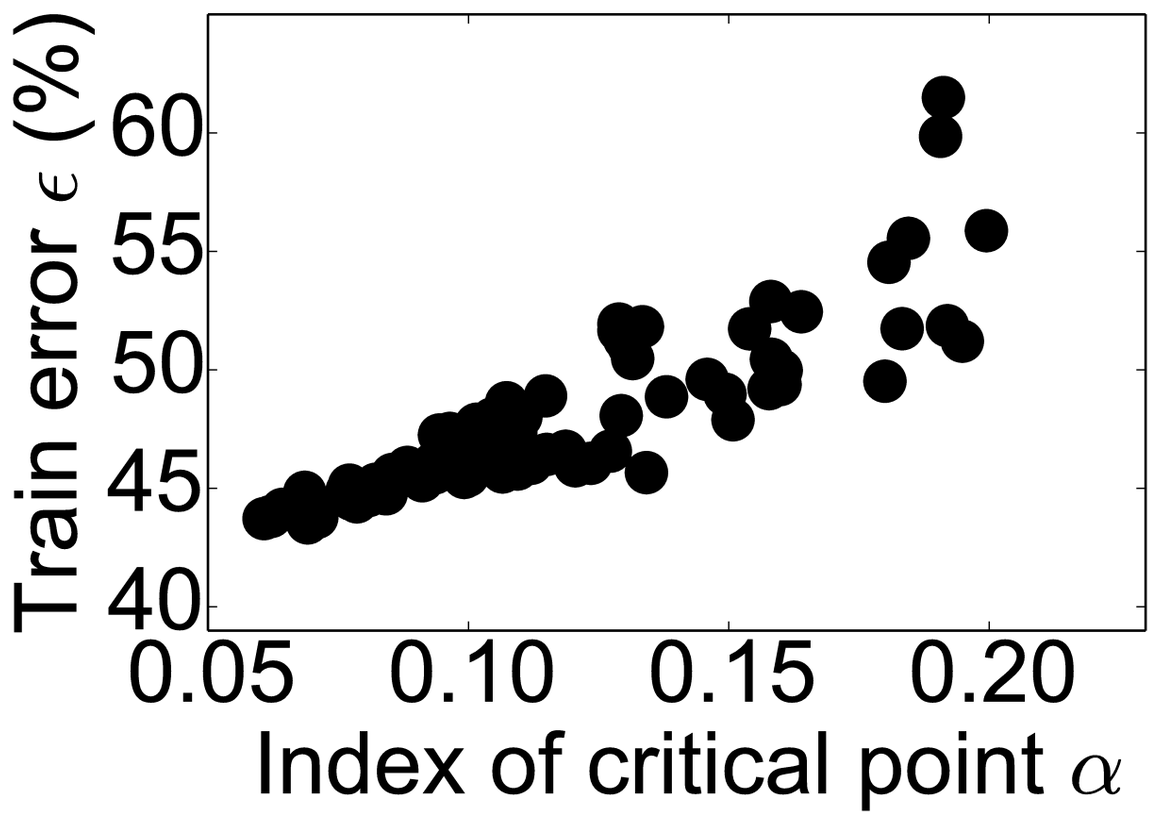


SADDLE POINTS

mnist

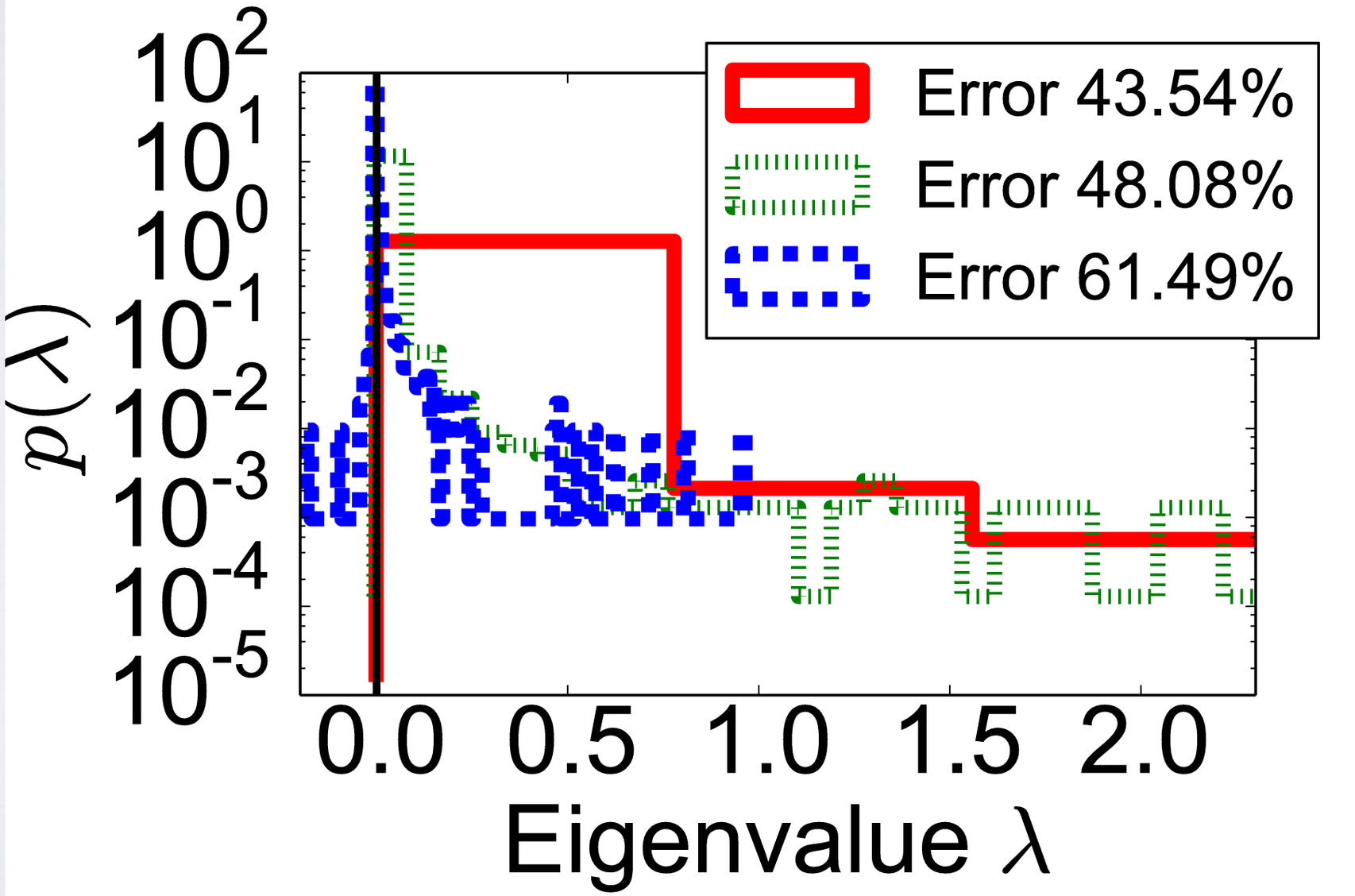


cifar

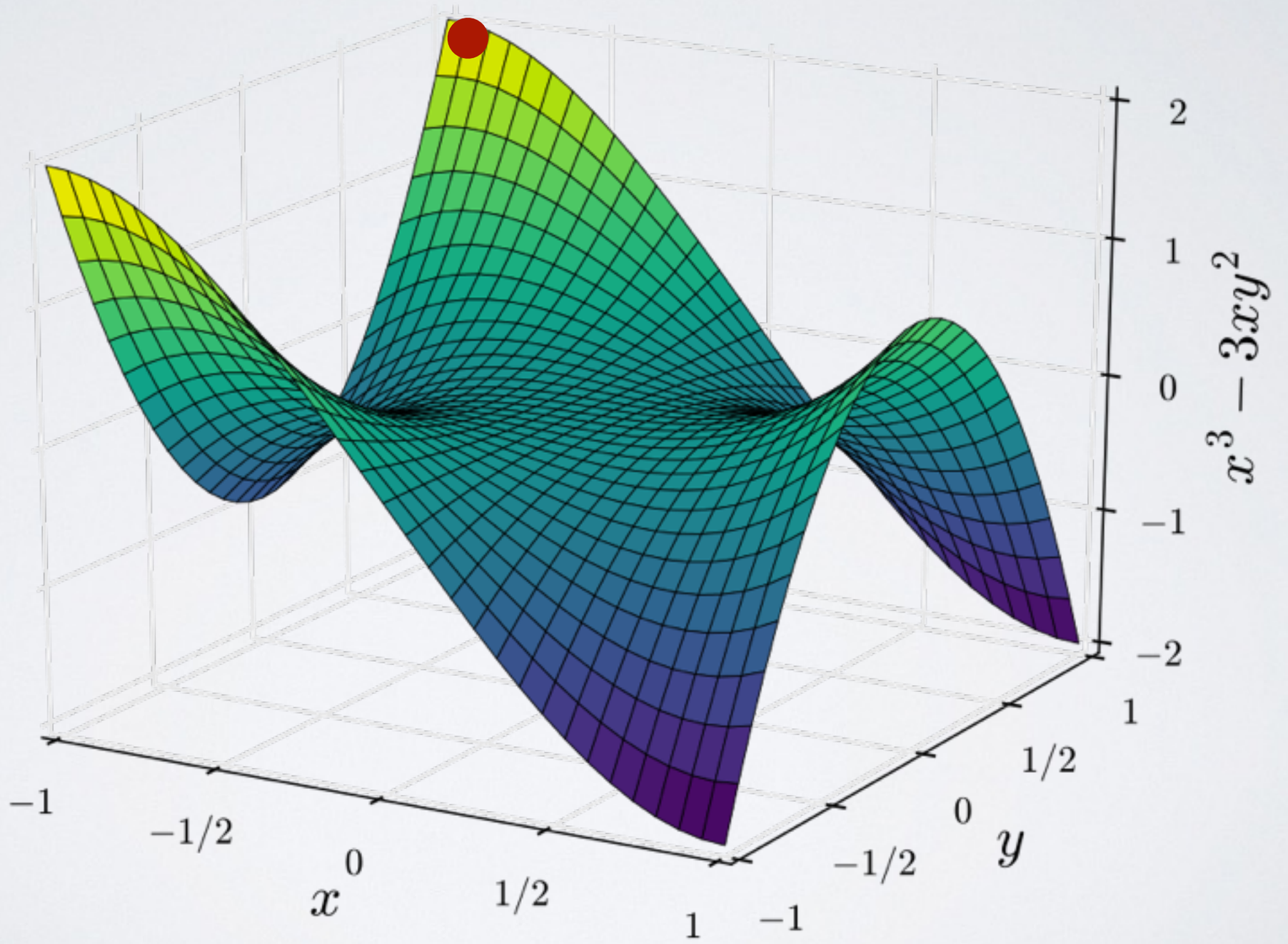


SADDLE POINTS

cifar



SADDLE POINTS



SADDLE POINTS

Escaping From Saddle Points – Online Stochastic Gradient for Tensor Decomposition

Rong Ge* Furong Huang † Chi Jin ‡ Yang Yuan §