Neural networks

Deep learning - example

DEEP LEARNING

Topics: pseudocode

- for l=1 to L
  - build unsupervised training set (with $\mathbf{h}^{(0)}(\mathbf{x}) = \mathbf{x}$):

$$\mathcal{D} = \left\{ \mathbf{h}^{(l-1)}(\mathbf{x}^{(t)}) \right\}_{t=1}^{T}$$

  - train "greedy module" (RBM, autoencoder) on $\mathcal{D}$
  - use hidden layer weights and biases of greedy module to initialize the deep network parameters $\mathbf{W}^{(l)}$, $\mathbf{b}^{(l)}$
- Initialize $\mathbf{W}^{(L+1)}$, $\mathbf{b}^{(L+1)}$ randomly (as usual)
- Train the whole neural network using (supervised) stochastic gradient descent (with backprop)

DEEP LEARNING

Topics: pseudocode

- for l=1 to L
  - build unsupervised training set (with $\mathbf{h}^{(0)}(\mathbf{x}) = \mathbf{x}$):

$$\mathcal{D} = \left\{\mathbf{h}^{(l-1)}(\mathbf{x}^{(t)})\right\}_{t=1}^{T}$$

  - train "greedy module" (RBM, autoencoder) on $\mathcal{D}$
  - use hidden layer weights and biases of greedy module to initialize the deep network parameters $\mathbf{W}^{(l)}$, $\mathbf{b}^{(l)}$

pre-training

- Initialize $\mathbf{W}^{(L+1)}$, $\mathbf{b}^{(L+1)}$ randomly (as usual)
- Train the whole neural network using (supervised) stochastic gradient descent (with backprop)

DEEP LEARNING

Topics: pseudocode

- for l=1 to L
  - build unsupervised training set (with $\mathbf{h}^{(0)}(\mathbf{x}) = \mathbf{x}$):

$$\mathcal{D} = \left\{\mathbf{h}^{(l-1)}(\mathbf{x}^{(t)})\right\}_{t=1}^{T}$$

  - train "greedy module" (RBM, autoencoder) on $\mathcal{D}$
  - use hidden layer weights and biases of greedy module to initialize the deep network parameters $\mathbf{W}^{(l)}$, $\mathbf{b}^{(l)}$

pre-training

- Initialize $\mathbf{W}^{(L+1)}$, $\mathbf{b}^{(L+1)}$ randomly (as usual)
- Train the whole neural network using (supervised) stochastic gradient descent (with backprop)

fine-tuning

DEEP LEARNING

Topics: datasets

- Datasets generated with varying number of factors of variations

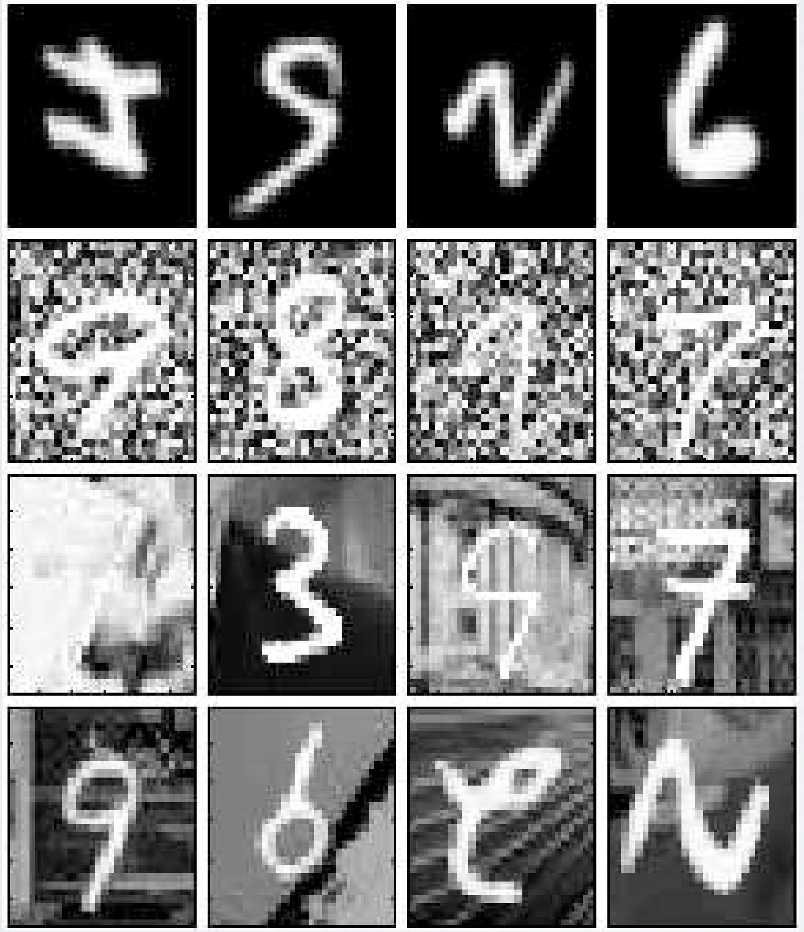
Variations on MNIST

MNIST-rotation

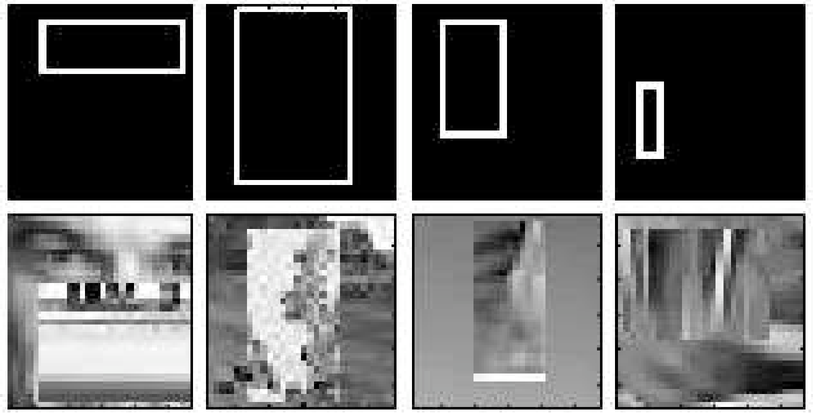
MNIST-random-background

MNIST-image-background

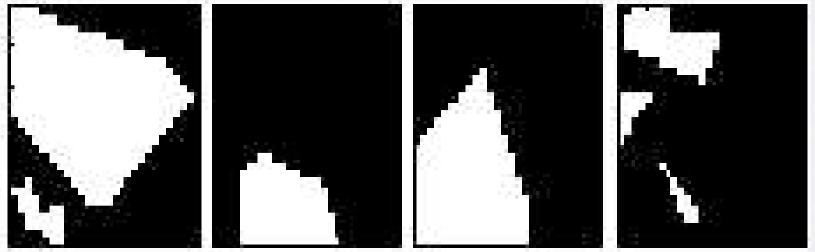
MNIST-background-rotation



Tall or wide?



Convex shape or not?



An Empirical Evaluation of Deep Architectures on Problems with Many Factors of Variation
Larochelle, Erhan, Courville, Bergstra and Bengio, 2007

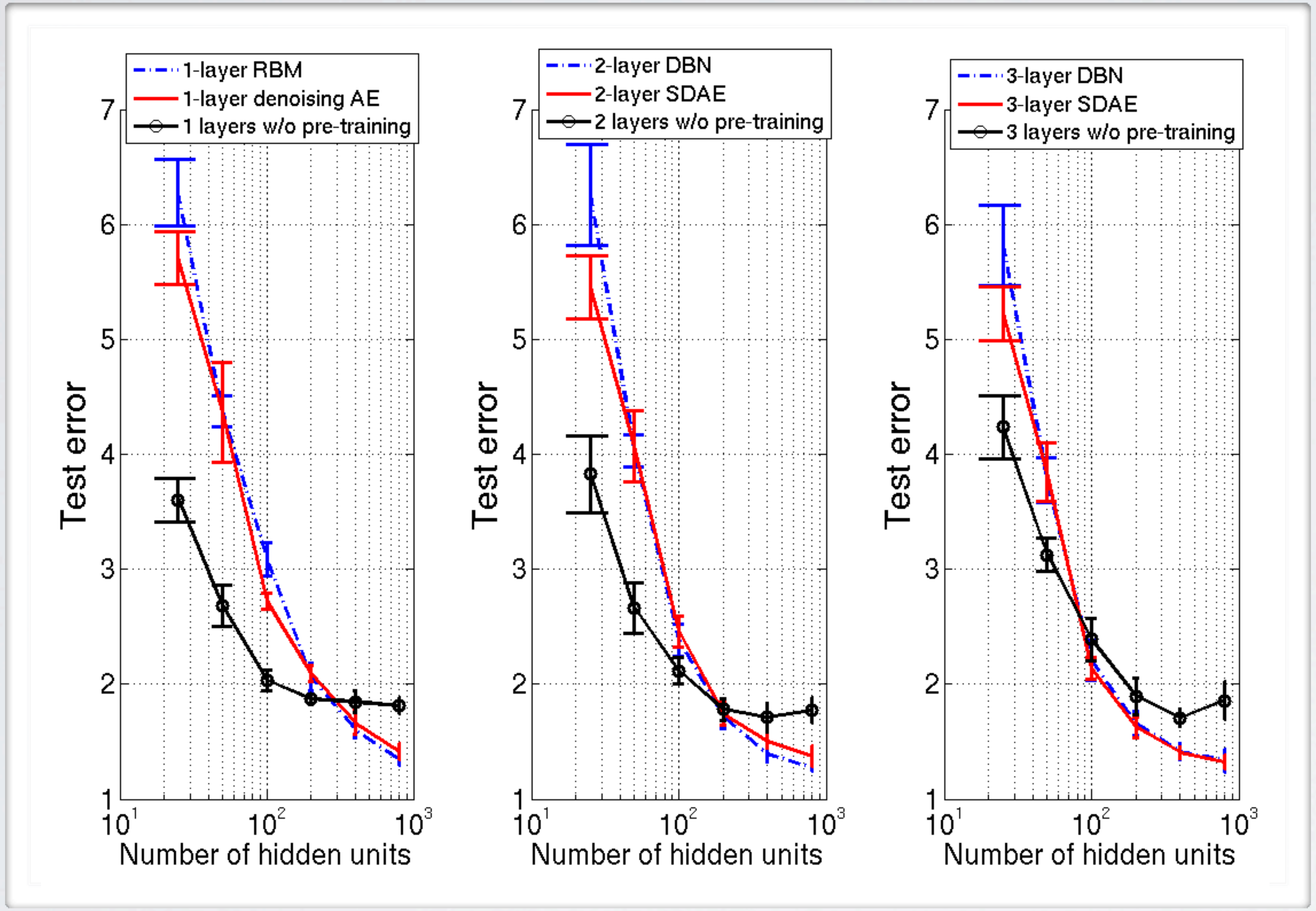
DEEP LEARNING

Topics: impact of initialization

Network		MNIST-small classif. test error	MNIST-rotation classif. test error
Type	Depth		
Deep net	1	**4.14** % ± 0.17	15.22 % ± 0.31
	2	**4.03** % ± 0.17	**10.63** % ± 0.27
	3	**4.24** % ± 0.18	11.98 % ± 0.28
	4	4.47 % ± 0.18	11.73 % ± 0.29
Deep net + autoencoder	1	3.87 % ± 0.17	11.43% ± 0.28
	2	**3.38** % ± 0.16	9.88 % ± 0.26
	3	**3.37** % ± 0.16	**9.22** % ± 0.25
	4	**3.39** % ± 0.16	**9.20** % ± 0.25
Deep net + RBM	1	3.17 % ± 0.15	10.47 % ± 0.27
	2	**2.74** % ± 0.14	9.54 % ± 0.26
	3	**2.71** % ± 0.14	**8.80** % ± 0.25
	4	**2.72** % ± 0.14	**8.83** % ± 0.24

DEEP LEARNING

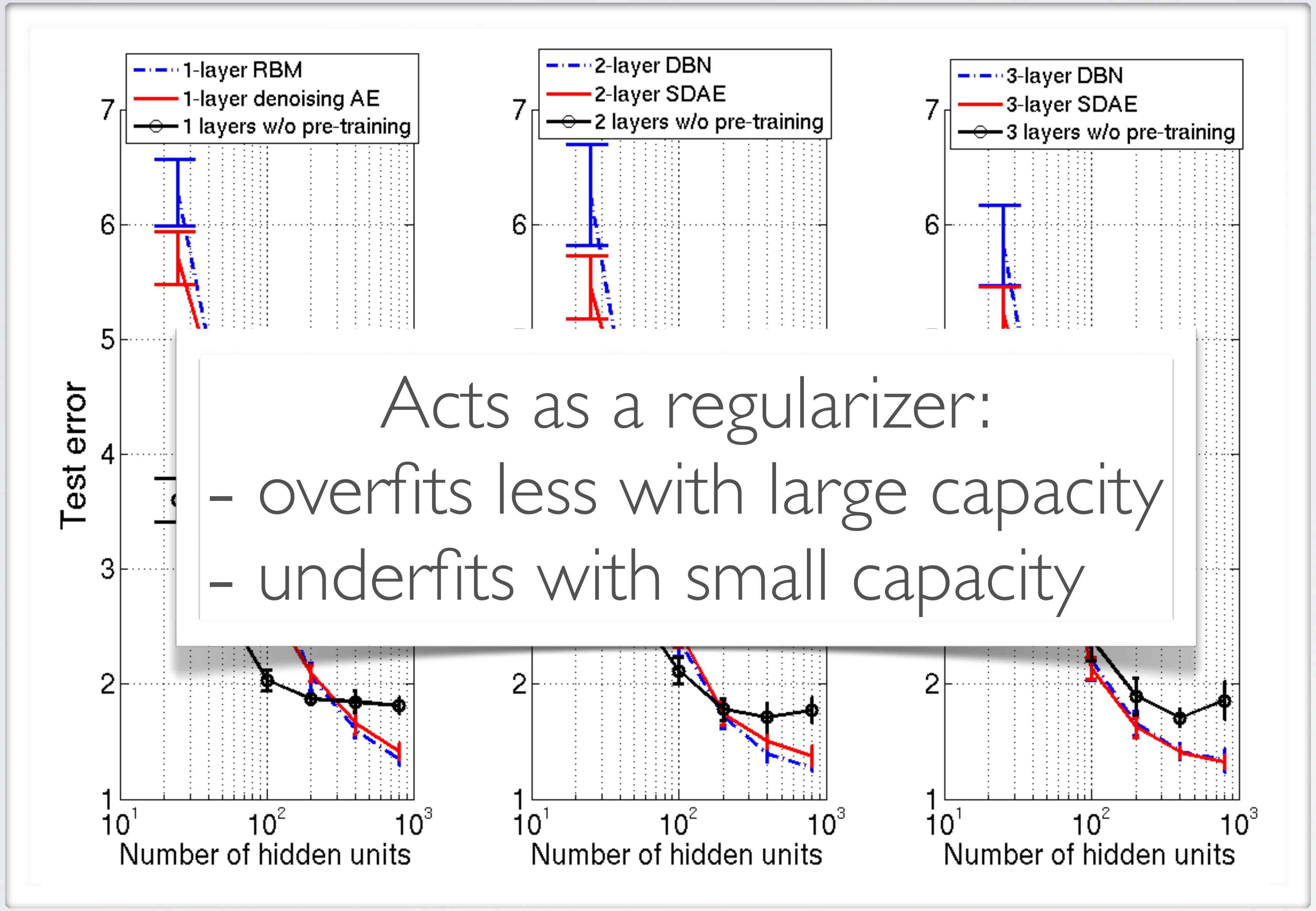
Topics: impact of initialization



Why Does Unsupervised Pre-training Help Deep Learning?
Erhan, Bengio, Courville, Manzagol, Vincent and Bengio, 2011

DEEP LEARNING

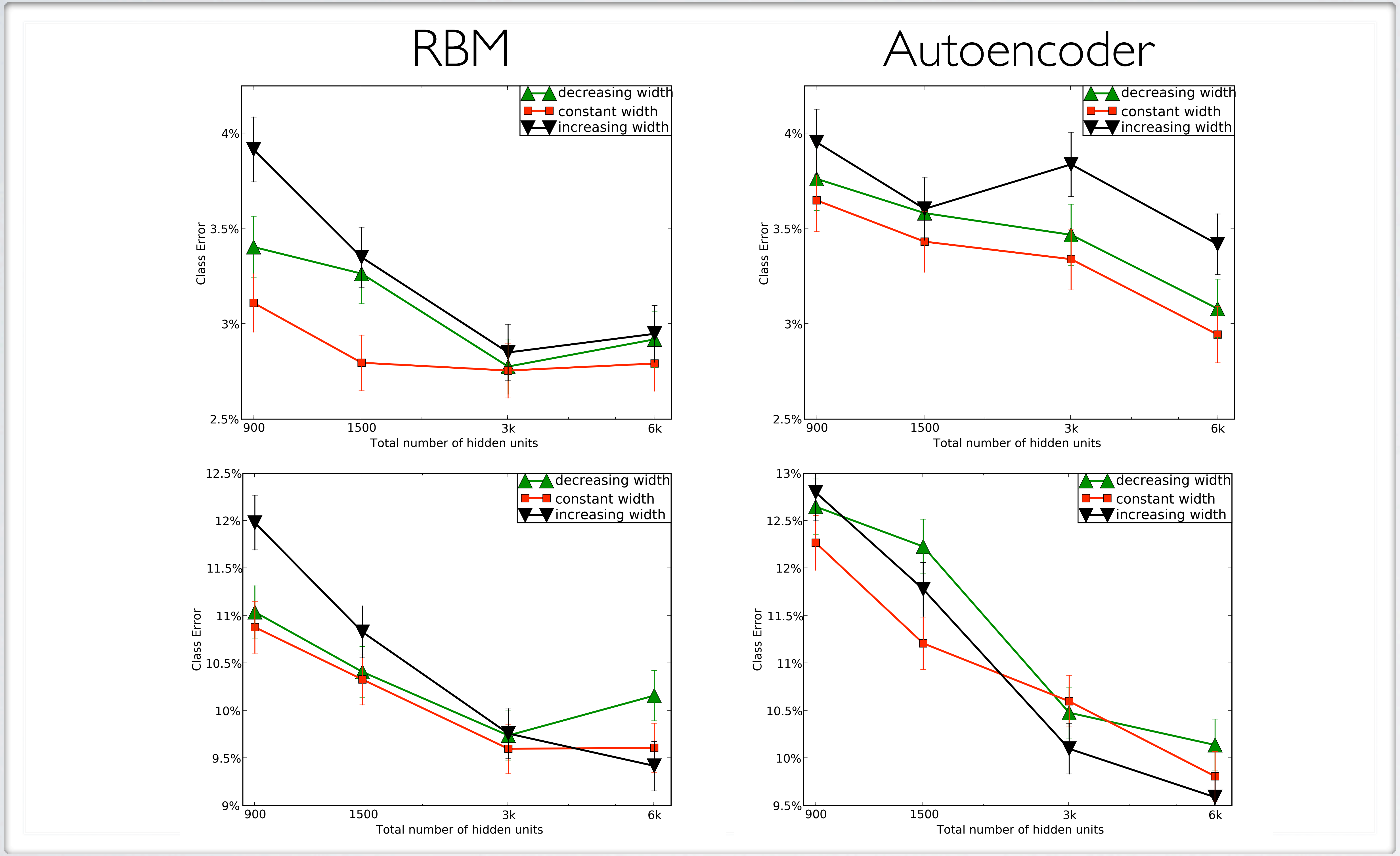
Topics: impact of initialization



Why Does Unsupervised Pre-training Help Deep Learning?
Erhan, Bengio, Courville, Manzagol, Vincent and Bengio, 2011

DEEP LEARNING

Topics: choice of hidden layer size



DEEP LEARNING

Topics: performance on different datasets

Dataset	SVM_{rbf} (Stacked Autoencoders →) SAA-3			

Dataset	SVM_{rbf}	SAA-3 (Stacked Autoencoders)	DBN-3 (Stacked RBMS)	SdA-3 (ν) (Stacked Denoising Autoencoders)
basic	**3.03±0.15**	3.46±0.16	3.11±0.15	**2.80±0.14** (10%)
rot	11.11±0.28	**10.30±0.27**	**10.30±0.27**	**10.29±0.27** (10%)
bg-rand	14.58±0.31	11.28±0.28	**6.73±0.22**	10.38±0.27 (40%)
bg-img	22.61±0.37	23.00±0.37	**16.31±0.32**	**16.68±0.33** (25%)
rot-bg-img	55.18±0.44	51.93±0.44	47.39±0.44	**44.49±0.44** (25%)
rect	**2.15±0.13**	2.41±0.13	2.60±0.14	**1.99±0.12** (10%)
rect-img	24.04±0.37	24.05±0.37	22.50±0.37	**21.59±0.36** (25%)
convex	19.13±0.34	**18.41±0.34**	**18.63±0.34**	**19.06±0.34** (10%)

Extracting and Composing Robust Features with Denoising Autoencoders,
Vincent, Larochelle, Bengio and Manzagol, 2008.