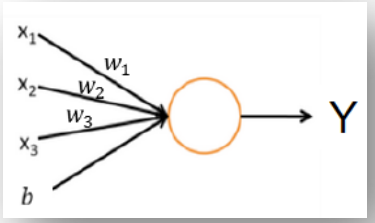
CS273B lecture 5: autoencoder

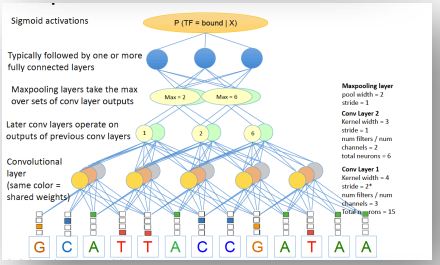
James Zou

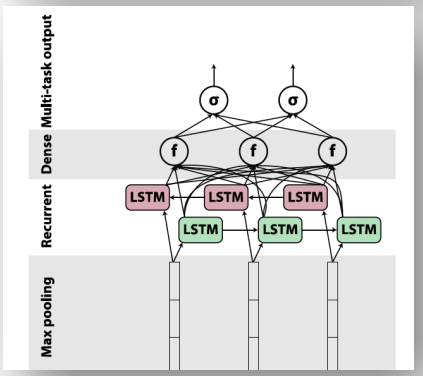
October 9, 2017

Deep supervised learning

- Feedforward
- Convnets
- RNN, LSTM







Learning a nonlinear mapping from inputs to outputs.

Predicting:
TF binding,
gene expression,
disease status from images,
risk from SNPs,
protein structure
…

Deep unsupervised learning

- Nonlinear dimensional reduction and patterns mining.
- In many settings, have more unlabeled examples than labeled.
- Learn useful representations from unlabeled data.
- Better representation may improve prediction accuracy.

Lecture outline

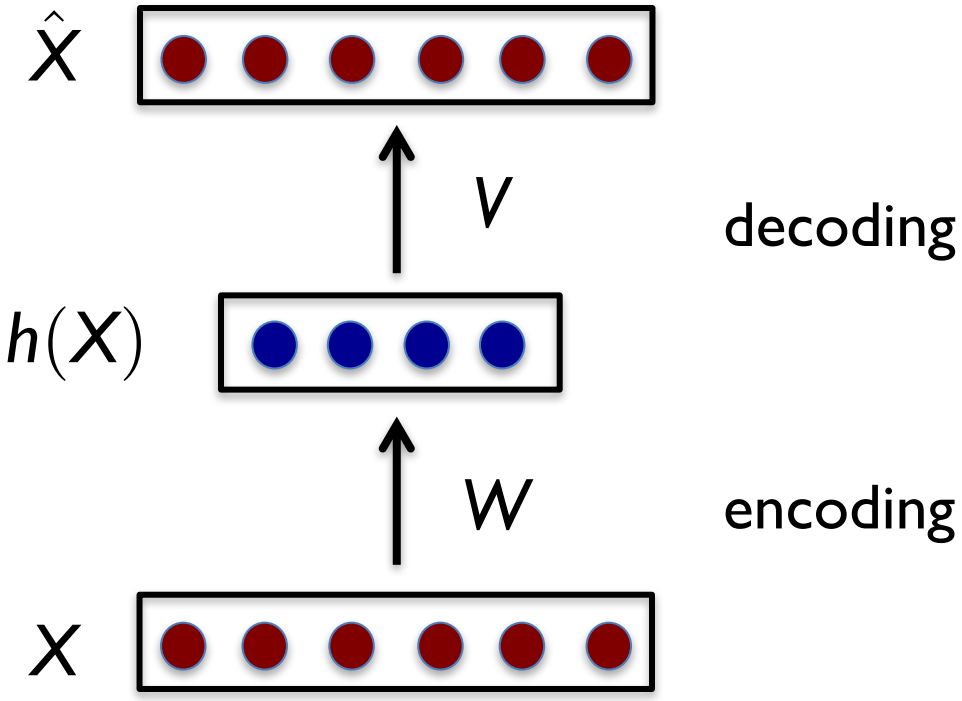
- Autoencoder.
- Denoising autoencoder.
- Application 1: breast cancer gene expression
- Application 2: medical records.

Low dimensional structure

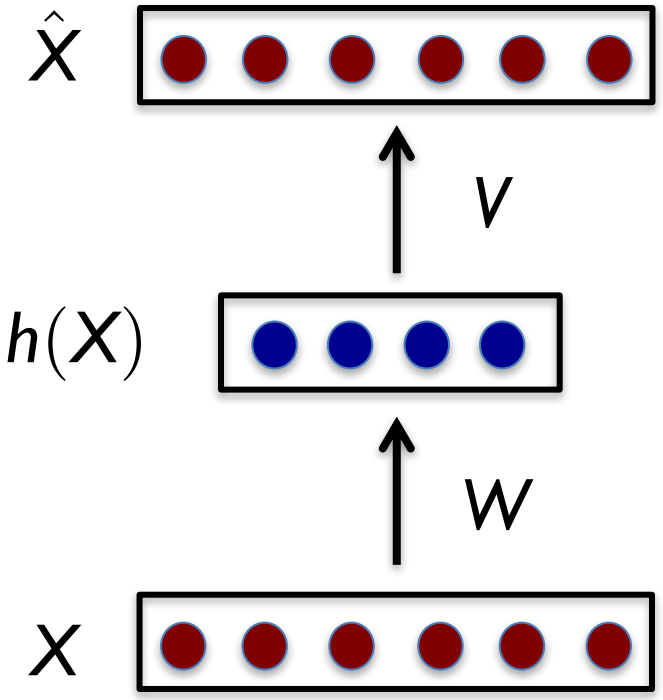
What is the latent dimensionality of each row of images?

Urtasun and Zemel.

Autoencoder



Autoencoder



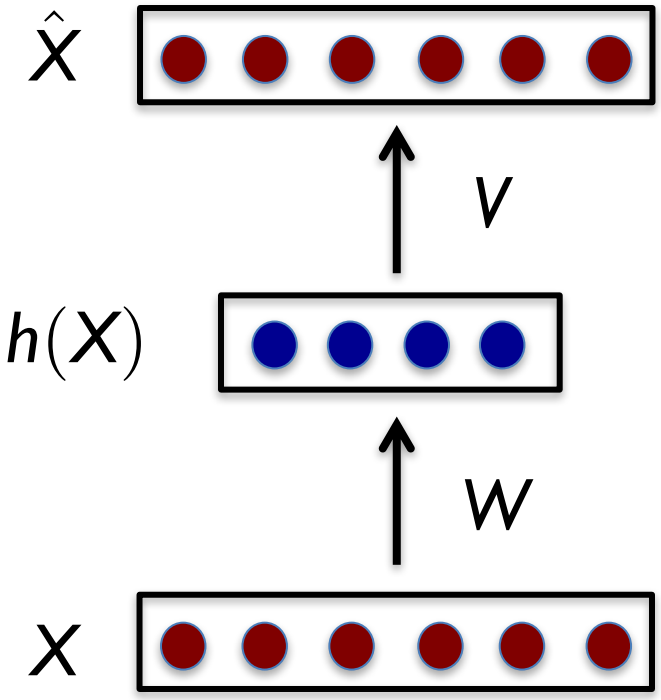
$$\hat{X} = g(V \cdot h + b_d)$$

$$h = f(W \cdot X + b_e)$$

$$W, V = \arg\min_{W,V} \sum_X ||X - \hat{X}||^2$$

Train with backprop as before.

Autoencoder

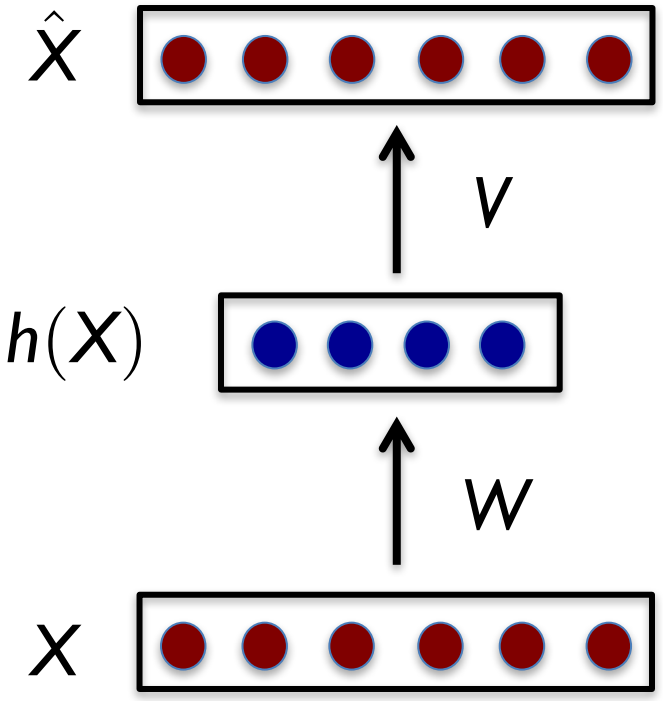


If encoding and decoding are linear then

$$W, V = \arg\min_{W,V} \sum_{x} ||X - VWX||^2$$

What does this remind you of?

Autoencoder



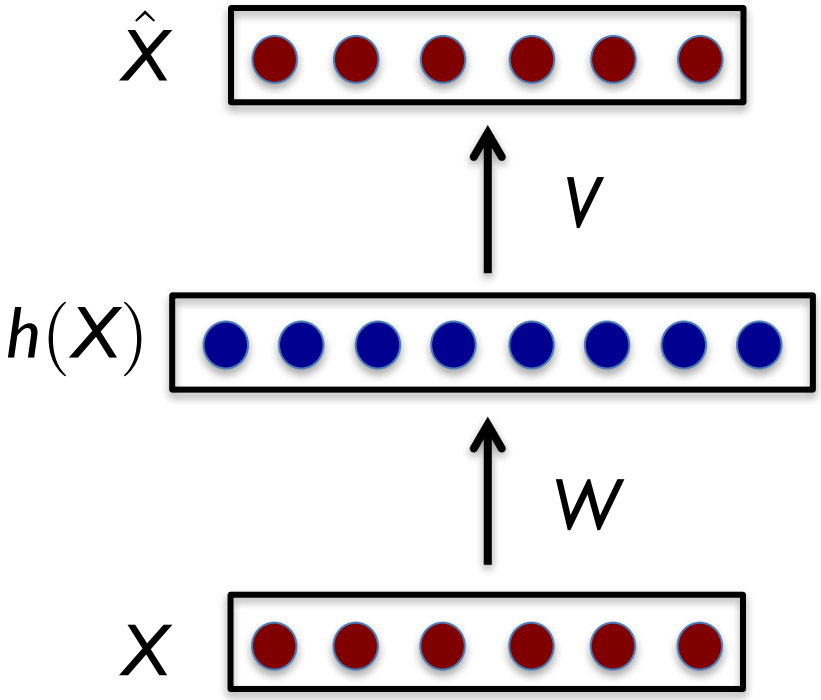
If encoding and decoding are linear then

$$W, V = \arg\min_{W,V} \sum_{x} ||X - VWX||^2$$

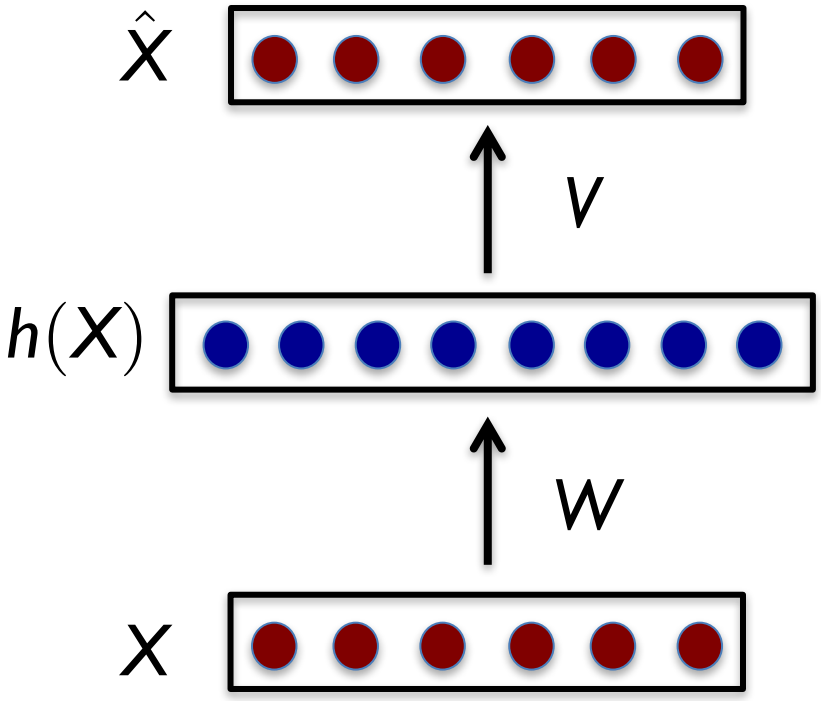
Linear autoencoder is basically just PCA!

General *f* and *g* corresponds to nonlinear dimensional reduction.

What is wrong with this picture?



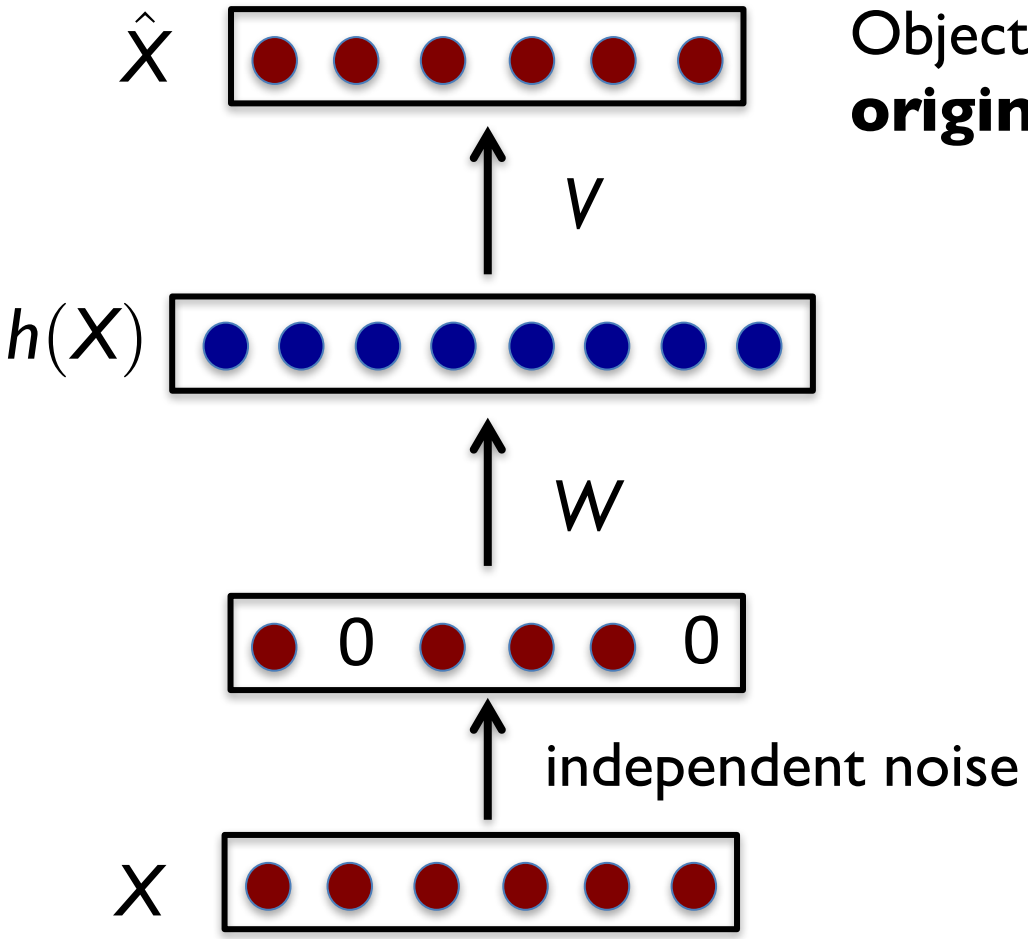
What is wrong with this picture?



h(X) can just copy X exactly!

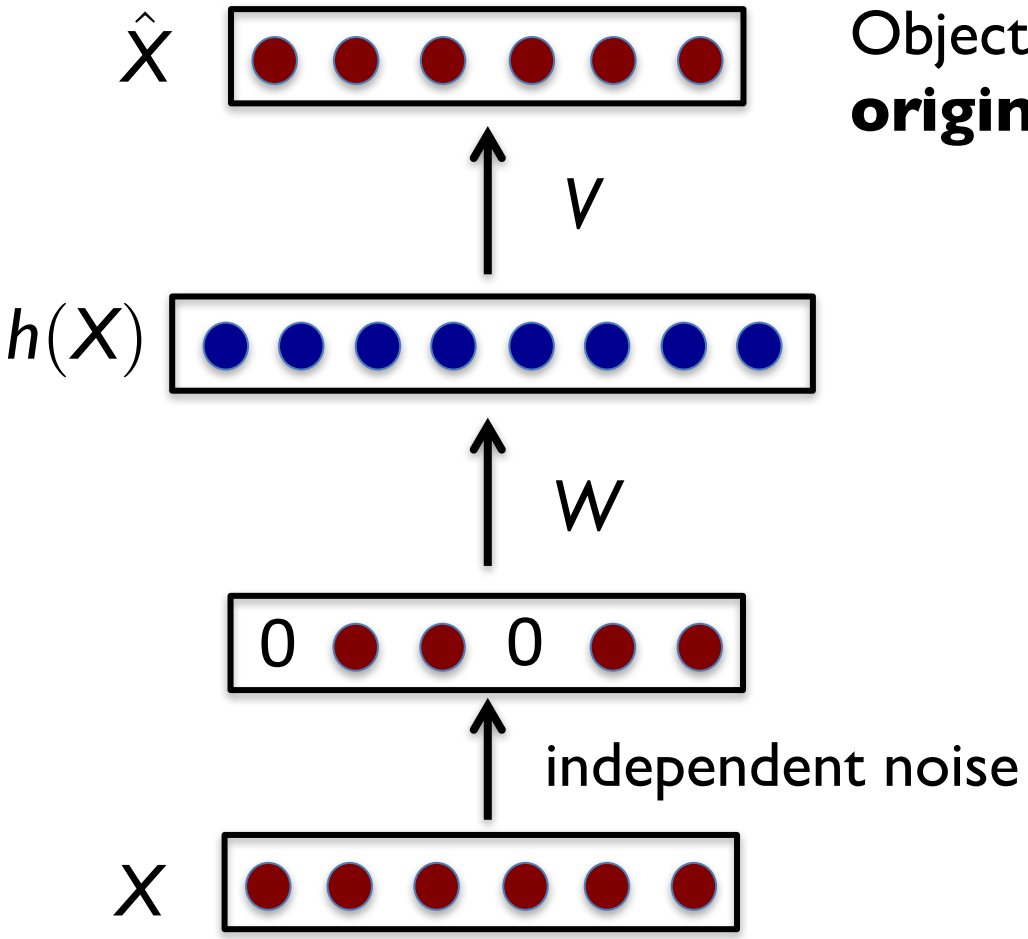
Overcomplete. Need to impose sparsity on *h*.

Denoising autoencoder



Objective: reconstruct the **original** clean input.

Denoising autoencoder



Objective: reconstruct the **original** clean input.

Illustration of denoising autoencoder

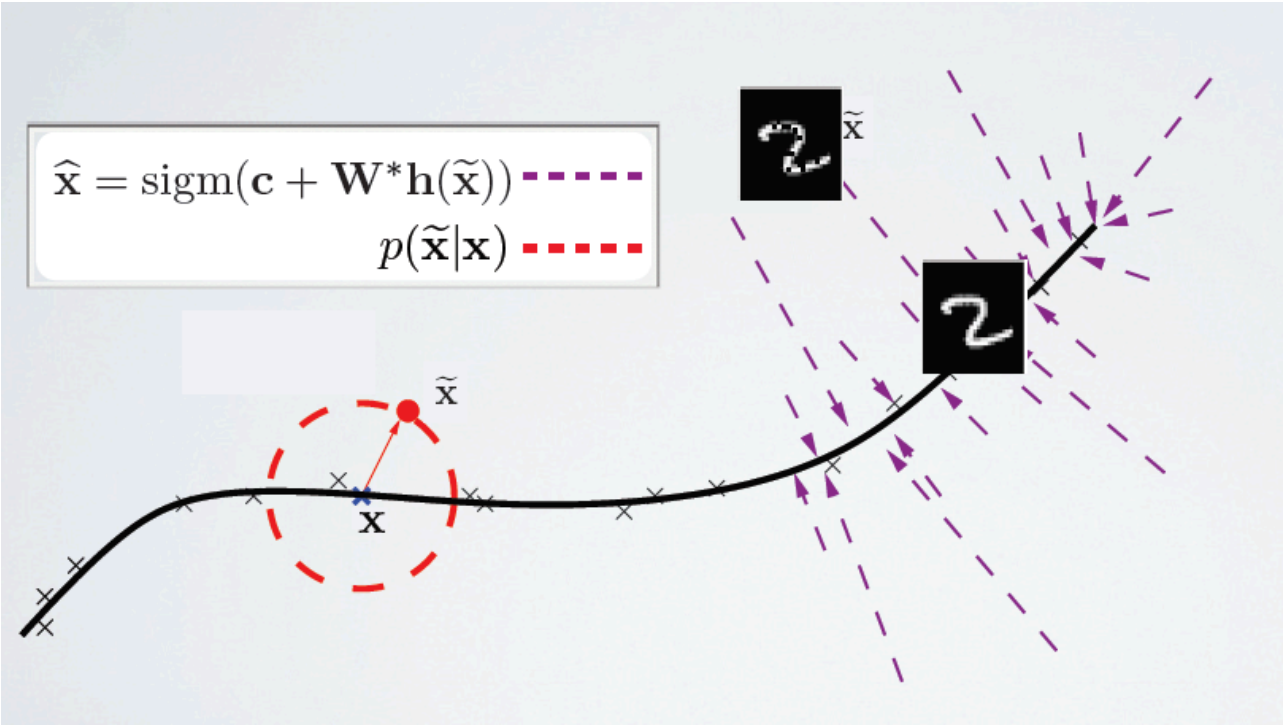
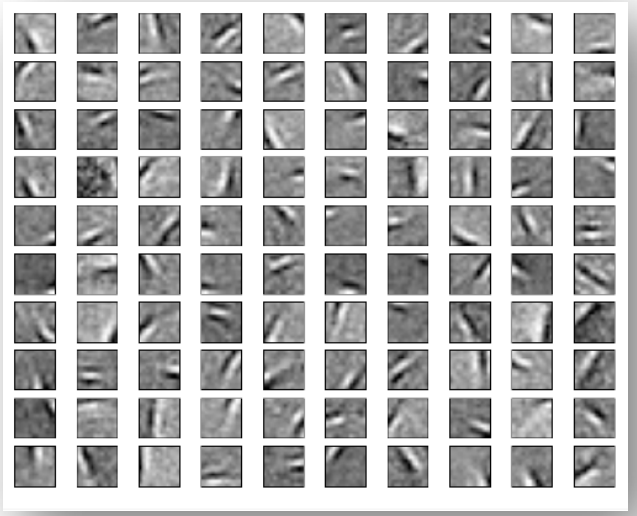


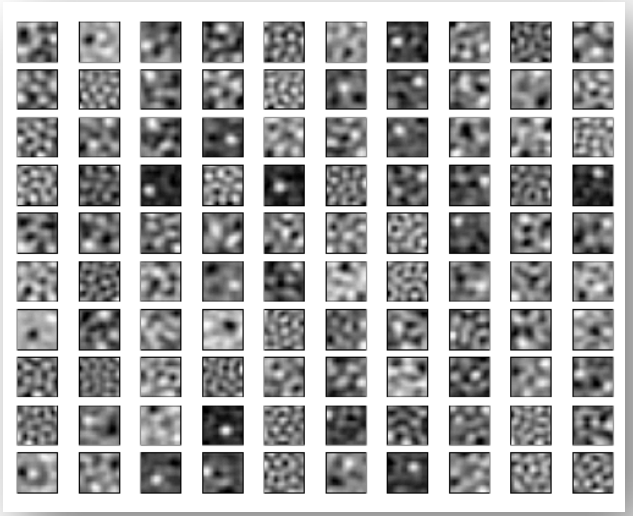
Figure from Hugo Larochelle

Filters from denoising autoencoder

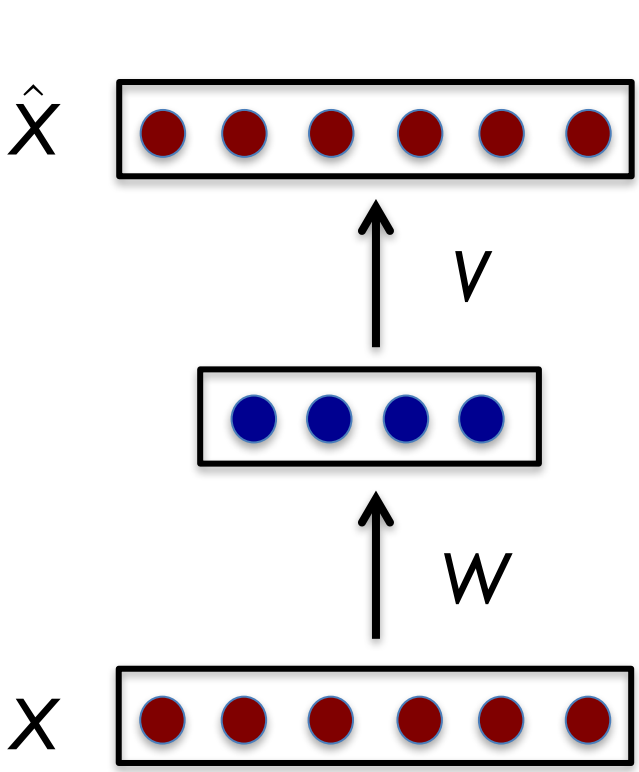
Basis learned by denoising autoencoder

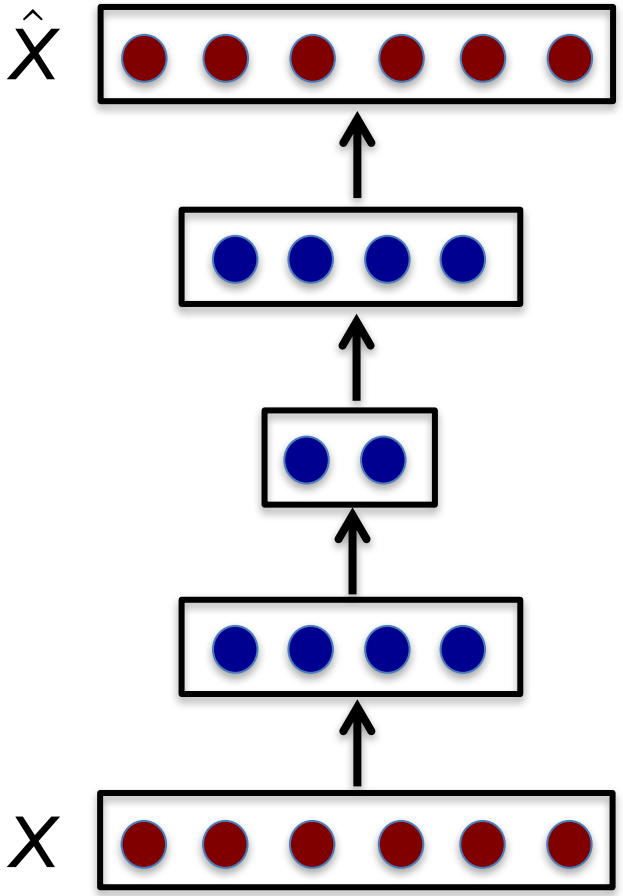


Basis learned by weight-decay autoencoder

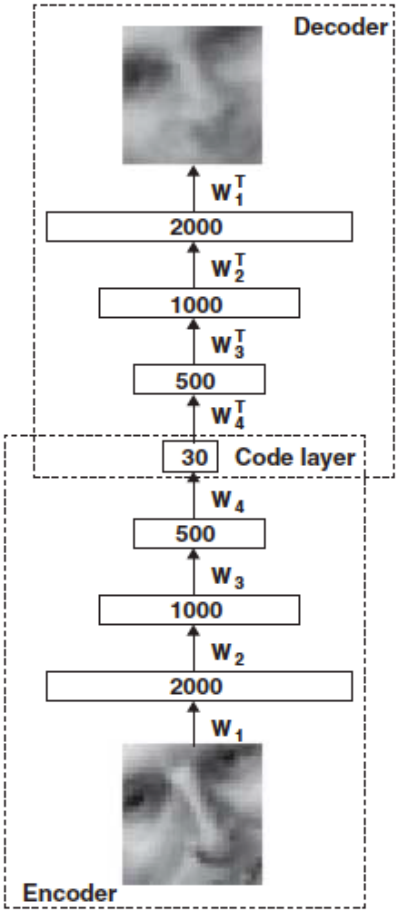


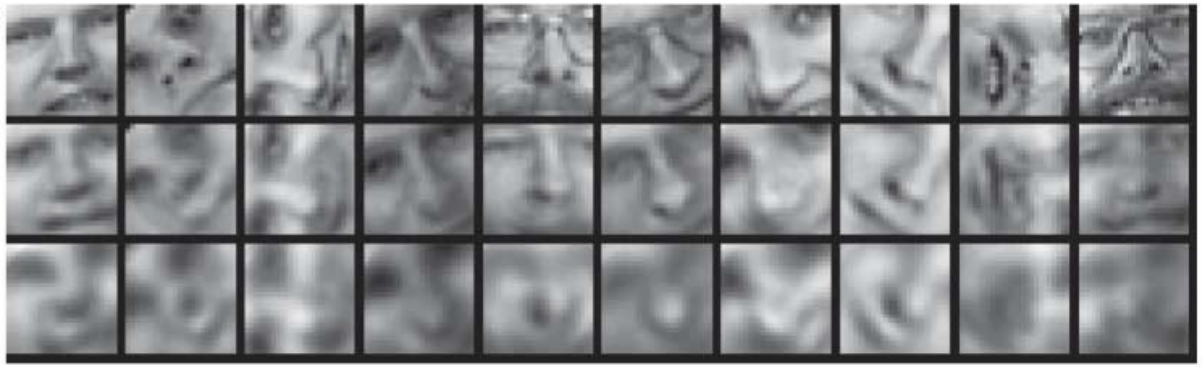
Deep autoencoder





Deep autoencoder example





original

DAE

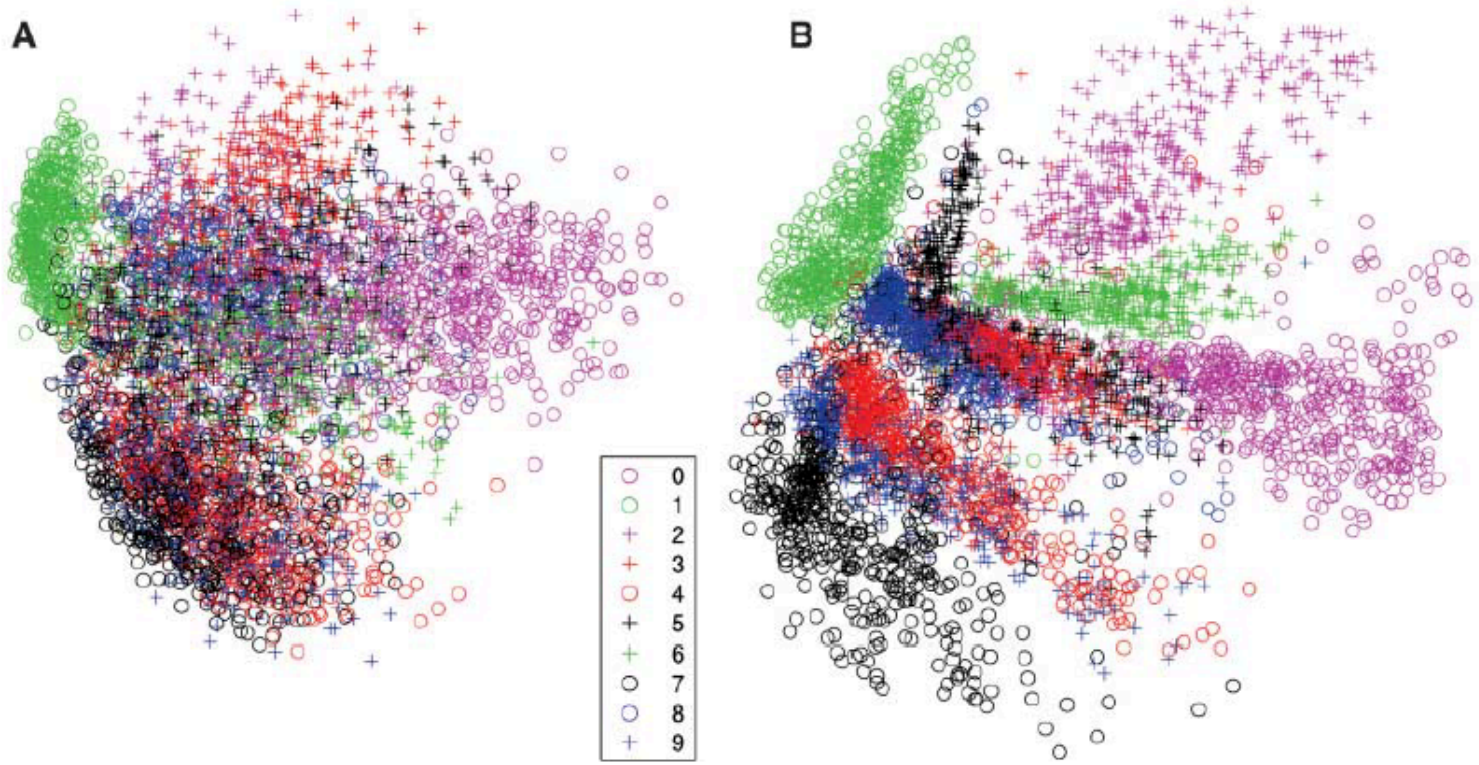
PCA

Hinton and Salakhutdinov. Science. 2016

Deep autoencoder example

PCA

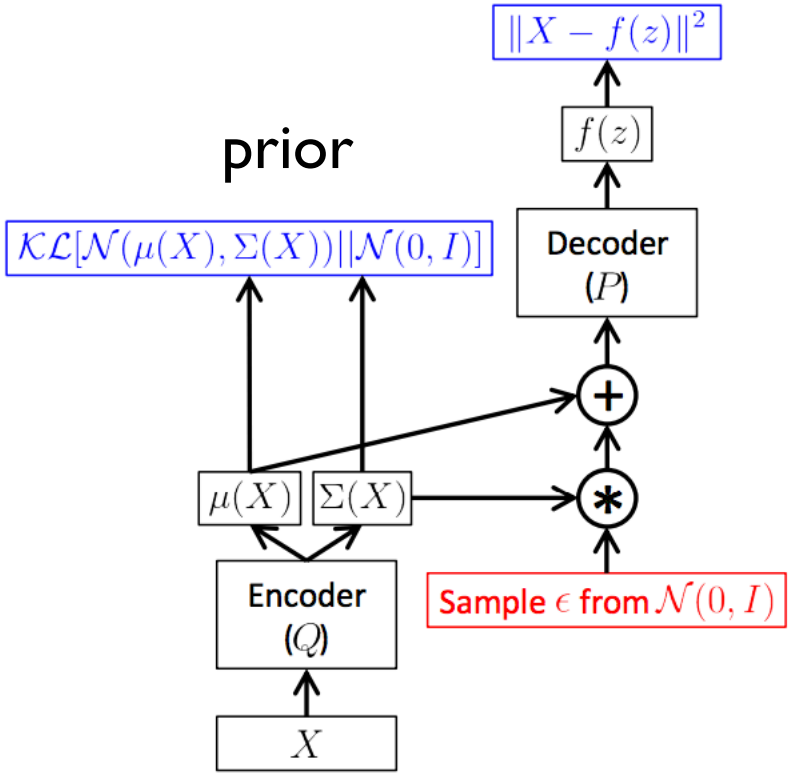
Deep autoencoder



Hinton and Salakhutdinov. Science. 2016

Variational autoencoder preview

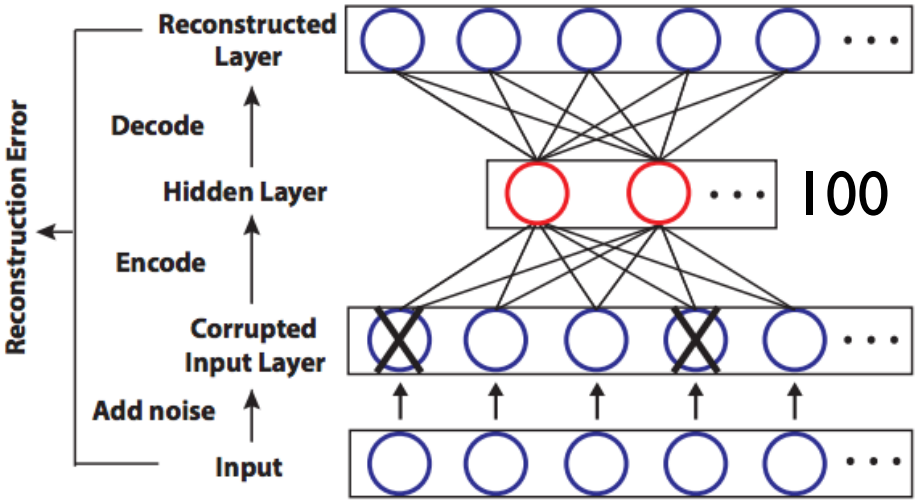
original reconstruction objective



probabilistic model

Doersch. 2016

Application: gene expression



$$y = sigmoid\left(Wx + b\right)$$
$$z = sigmoid\left(W'y + b'\right)$$
$$L_H\left(x, z\right) = -\sum_{k=1}^{d}[x_k \log z_k + (1 - x_k) \log\left(1 - z_k\right)]$$

Sample 1: $gene_1$ $gene_2$ … $gene_{3000}$

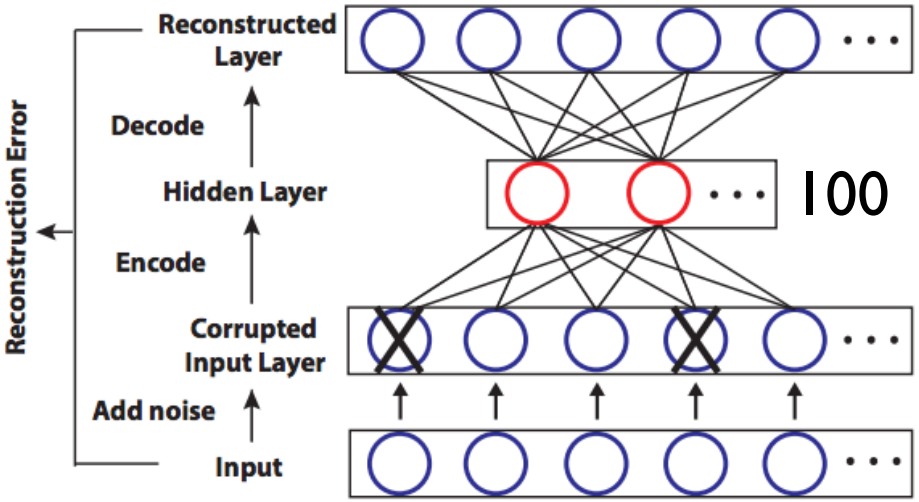
⋮

Sample n: $gene_1$ $gene_2$ … $gene_{3000}$

1992 breast cancer samples
+ 144 normal samples

Tan et al. 2015

Application: gene expression



$$y = sigmoid\left(Wx + b\right)$$

$$z = sigmoid\left(W'y + b'\right)$$

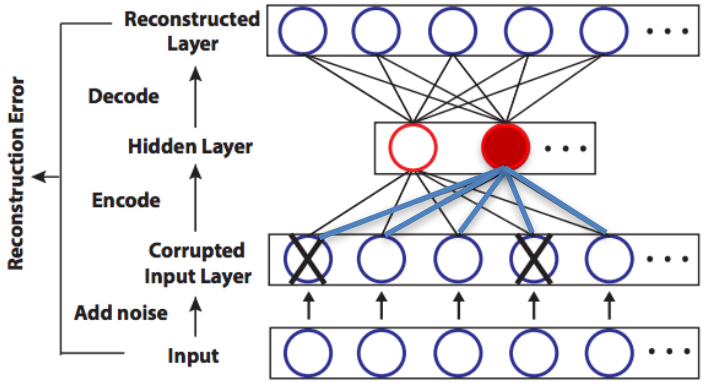
$$L_H\left(x, z\right) = -\sum_{k=1}^{d}[x_k \log z_k + (1 - x_k) \log\left(1 - z_k\right)]$$

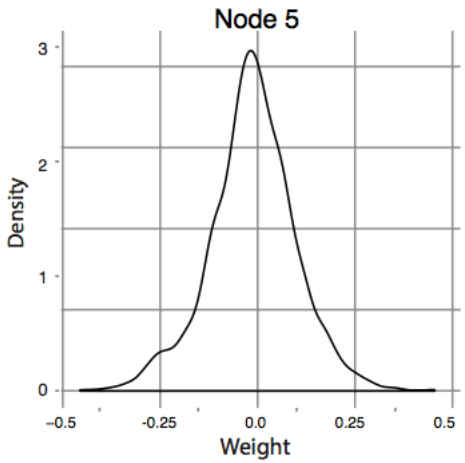
Batch size: 10
Corruption level: 0.1
Learning rate: 0.01

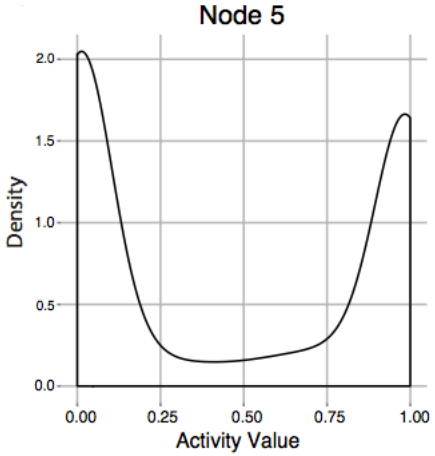
How would you interpret the hidden layer?

Tan et al. 2015

Interpreting the hidden layer

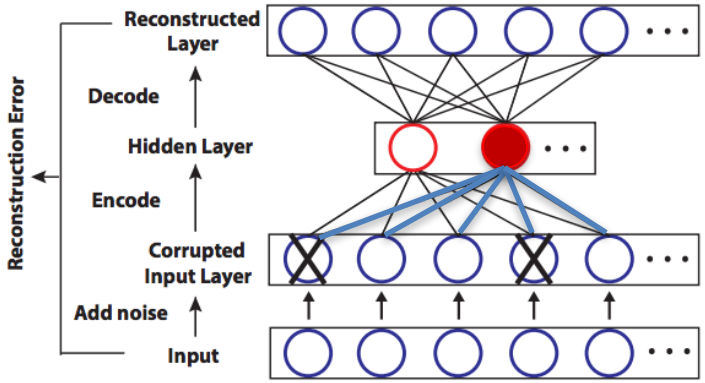


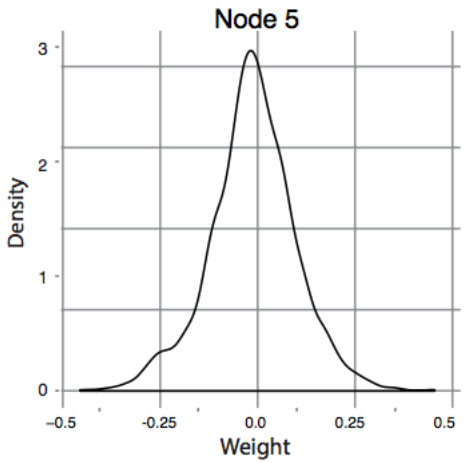


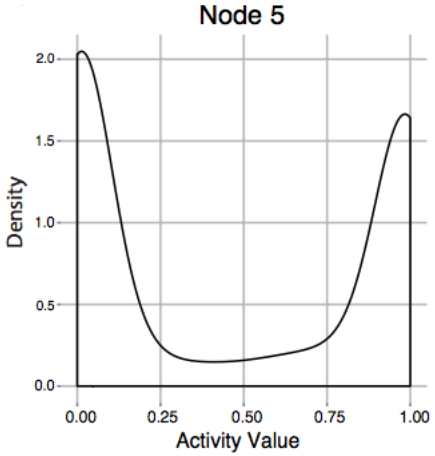


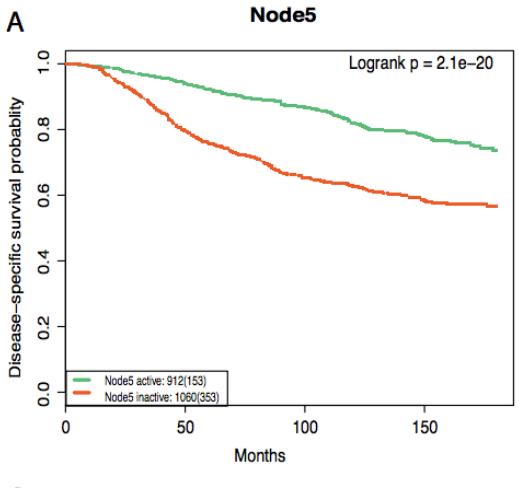
Tan et al. 2015

Interpreting the hidden layer



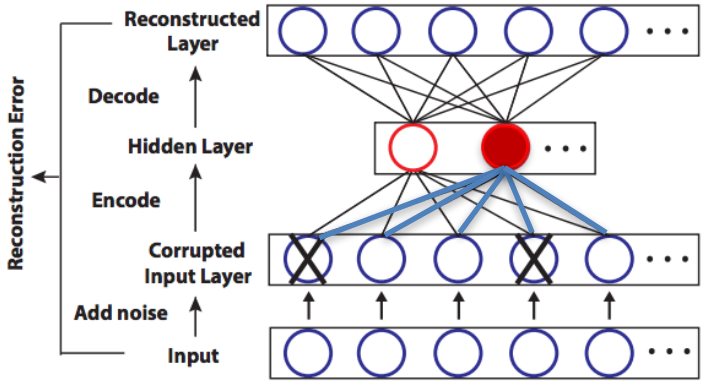


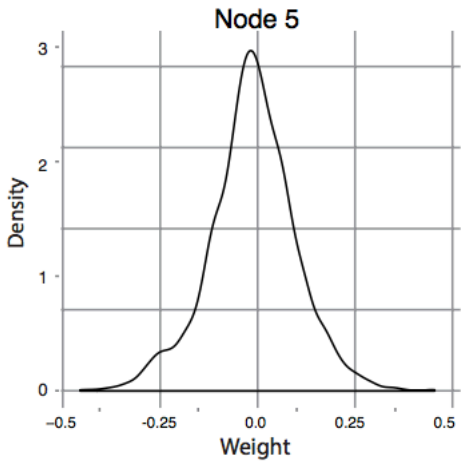


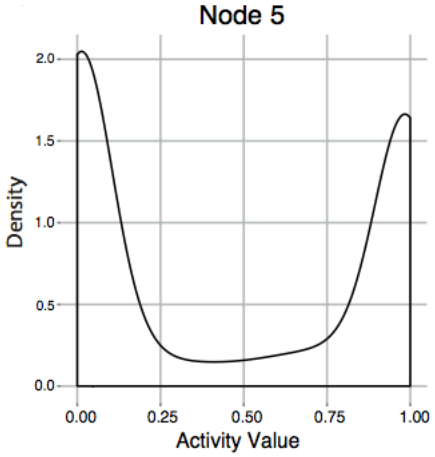


Tan et al. 2015

Interpreting the hidden layer







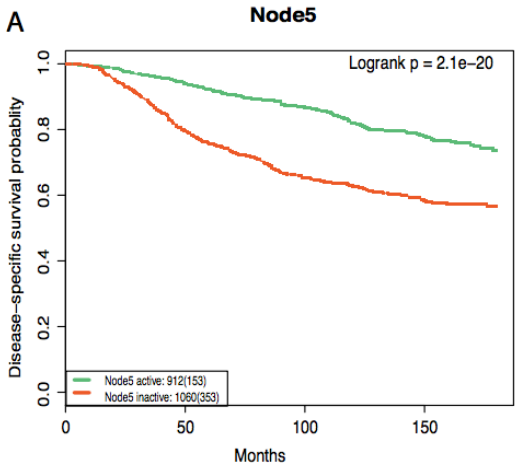


Table 4. PID pathways enriched in Node5.

Pathway	FDR q-value
FOXM1 transcription factor network	$< 1e^{-4}$
Aurora B signaling	$4.93e^{-4}$
Aurora A signaling	0.001
PLK1 signaling	0.003
Integrin-linked kinase signaling	0.068
C-MYB transcription factor network	0.074

Tan et al. 2015

Interpreting the hidden layer

Table 1. Performance of hidden nodes in classifying tumor from normal samples.

	METABRIC		TCGA
Node	Discovery	Test	Evaluation
64	0.970	0.968	0.996
99	0.957	0.959	0.998
38	0.879	0.887	0.911
43	0.873	0.873	0.750
69	0.871	0.872	0.906

Table 2. Performance of hidden nodes in classifying ER + from ER - samples.

	METABRIC		TCGA
Node	Discovery	Test	Evaluation
89	0.848	0.833	0.749
30	0.824	0.822	0.856
58	0.808	0.801	0.828
6	0.798	0.799	0.771
69	0.784	0.779	0.820

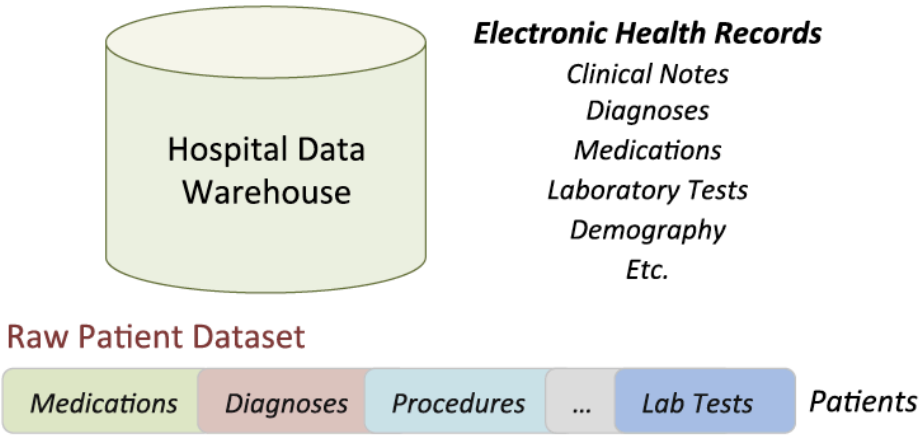
Tan et al. 2015

What are potential limitations?

- Comparison to baselines? PCA, regular autoencoder.
- Clustering and visualization?
- Imbalanced samples.
- Relations between neurons.

Tan et al. 2015

Application: medical records



Each patient = vector of 41k clinical descriptors

Stack of 3 denoising autoencoder
500 dim representation of each patient

Miotto et al. DeepPatient. 2016

Application: deep patient

700K patients up to 2013
500 dim representation of each patient

Random forest to predict *future* disease in 2014

Time Interval = 1 year (76,214 patients)			
		Classification Threshold = 0.6	
Patient Representation	AUC-ROC	Accuracy	F-Score
RawFeat	0.659	0.805	0.084
PCA	0.696	0.879	0.104
GMM	0.632	0.891	0.072
K-Means	0.672	0.887	0.093
ICA	0.695	0.882	0.101
DeepPatient	**0.773***	**0.929***	**0.181***

Miotto et al. DeepPatient. 2016

Application: deep patient

700K patients up to 2013
500 dim representation of each patient

Random forest to predict *future* disease in 2014

Time Interval = 1 year (76,214 patients)			
	Area under the ROC curve		
Disease	RawFeat	PCA	DeepPatient
Diabetes mellitus with complications	0.794	0.861	**0.907**
Cancer of rectum and anus	0.863	0.821	**0.887**
Cancer of liver and intrahepatic bile duct	0.830	0.867	**0.886**
Regional enteritis and ulcerative colitis	0.814	0.843	**0.870**
Congestive heart failure (non-hypertensive)	0.808	0.808	**0.865**
Attention-deficit and disruptive behavior disorders	0.730	0.797	**0.863**
Cancer of prostate	0.692	0.820	**0.859**
Schizophrenia	0.791	0.788	**0.853**
Multiple myeloma	0.783	0.739	**0.849**
Acute myocardial infarction	0.771	0.775	**0.847**

Miotto et al. DeepPatient. 2016