# Improving Neural Networks by Preventing Co-adaptation of Feature Detectors

G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever and R. R. Salakhutdinov (2012)

# Problem in a nutshell

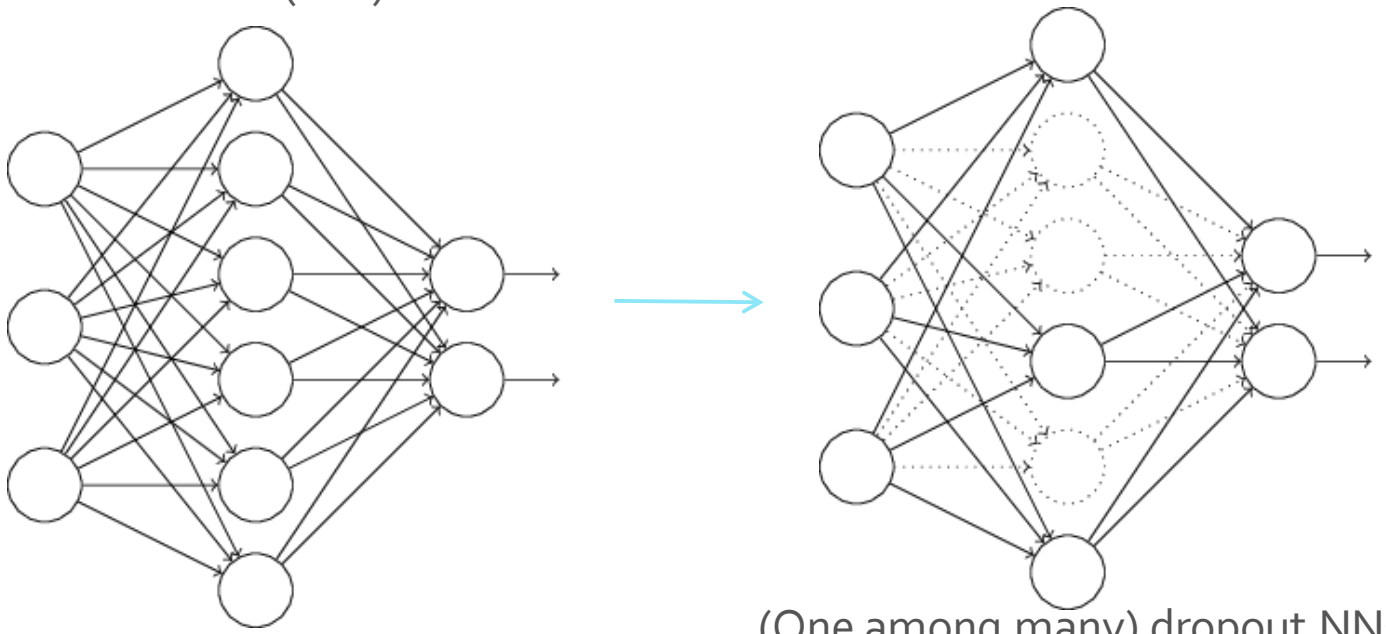
- Training a large feed-forward neural network using a small dataset may result in an "overfit" model

# (One) Existing solution

- Model Averaging
  - Instead of using one(large) NN, use many different(large) NNs to classify or predict
  - Test data on each model
  - To obtain the final output, combine individual model outputs(predictions) via averaging or voting scheme
  - But, training and testing on multiple large NNs becomes computationally expensive

# Proposed solution

- Model Averaging using "dropout"
  - Instead of using one large NN, use many dropout (smaller) NNs to classify or predict
  - Test data on (one) "mean network"



(One among many) dropout NNs

# Proposed solution in detail

## Training

- For each training instance (mini-batch)
  - Create a dropout NN by omitting random 50% hidden and/or 20% input units
  - Renormalize weight vector(W) when it exceeds the upper bound on L2 norm of W
  - After each epoch, increment the learning rate by a factor of 0.998

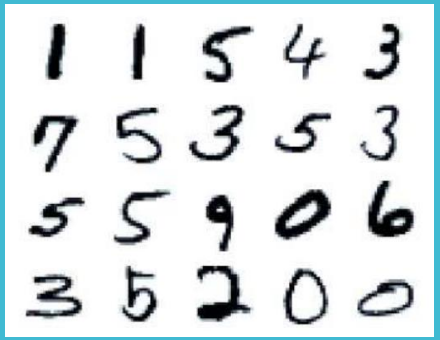
## Testing

- Create a "mean network"
  - In the initial NN(with all hidden units), reduce all outgoing weights by a factor of 0.5
  - Log likelihood estimate is guaranteed to be higher than each individual network*

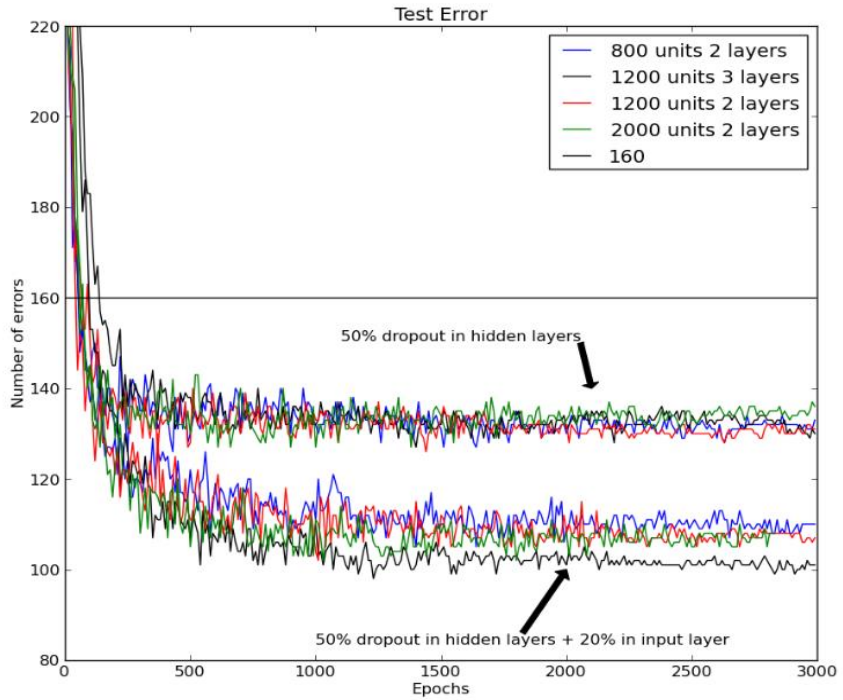
# Experiments and Results I

MNIST dataset

(60000 train, 20000 test)



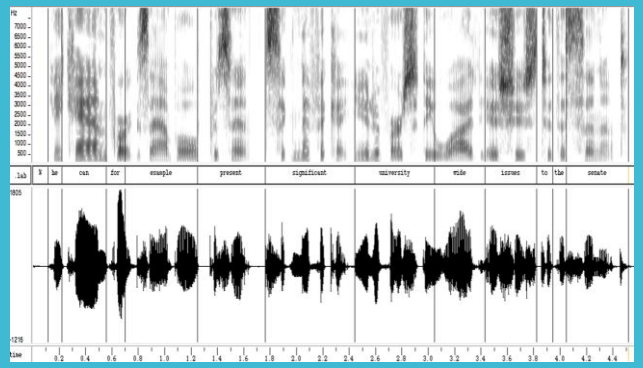
- "Dropout" training using different feed-forward NN architectures
- "Dropout" fine-tuning on deep belief nets and deep Boltzmann machines
  - #Units in each hidden layer – 500, 1000
  - #Errors reduced from 95 to 79 after 50% dropout in DBM; from 118 to 92 in DBN



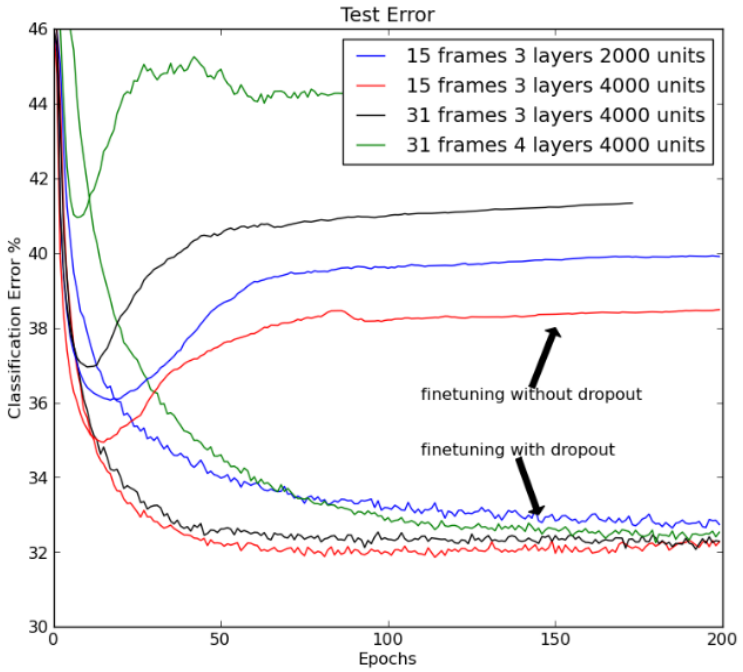
# Experiments and Results II

TIMIT dataset

(630 speakers - 10 word sentence)



- "Dropout" training using NN architectures
  - #Units in input layer – 15,31; output layer – 10
- "Dropout" fine-tuning vs Standard backpropagation fine-tuning



# Experiments and Results III

Reuters Corpus Volume I

```
<REUTERS TOPICS=''YES'' LEWISSPLIT=''TRAIN''
CGISPLIT=''TRAINING-SET'' OLDID=''12981'' NEWID=''798''>
<DATE> 2-MAR-1987 16:51:43.42</DATE>
<TOPICS><D>livestock</D><D>hog</D></TOPICS>
<TITLE>AMERICAN PORK CONGRESS KICKS OFF TOMORROW</TITLE>
<DATELINE> CHICAGO, March 2 - </DATELINE><BODY>The American Pork
Congress kicks off tomorrow, March 3, in Indianapolis with 160
of the nations pork producers from 44 member states determining
industry positions on a number of issues, according to the
National Pork Producers Council, NPPC.
Delegates to the three day Congress will be considering 26
resolutions concerning various issues, including the future
direction of farm policy and the tax law as it applies to the
agriculture sector. The delegates will also debate whether to
endorse concepts of a national PRV (pseudorabies virus) control
and eradication program, the NPPC said. A large
trade show, in conjunction with the congress, will feature
the latest in technology in all areas of the industry, the NPPC
added. Reuter
\&\#3;</BODY></TEXT></REUTERS>
```

- 804,414 text documents - 103 different topics
- 50 classes with 402,738 documents
- Random split of data into two halves for training and testing
- Utilize 2000 most frequent words to represent documents
- Standard backpropagation NN error rate reduced from 31.7% to 29.63% with 50% dropout

# Experiments and Results IV

CIFAR-10 dataset

- 50000 train – 10000 test (tiny) 32x32 images
- 10 output classes
- State-of-the-art -18.5% error rate
- CNN without dropout - 16.6% error rate
  - 3 convolutional layers
  - 3 max-pooling layers
- CNN with dropout - 15.6% error rate
  - 50% dropout on last hidden layer

# Experiments and Results V

IMAGENET dataset

- Millions of high-resolution images resized to 256x256 inputs
- 1000 classes with 1000 examples each
- State-of-the-art - 47.7% error rate
- CNN without dropout – 48.6% error rate
  - 5 convolutional layers
  - 3 max-pooling layers after 1st, 2nd and 5th layers
- CNN with 50% dropout – 42.4% error rate
  - 2 more hidden layers before the output layer

# Summary

- Proposed "dropout" technique solved the problem of overfitting
- Implementation to test generalization performance
- Improved accuracy in popular speech recognition, natural language processing and object recognition datasets compared to standard backpropagation techniques without dropout.



Image credits: lifehacker.com

# Reducing the Dimensionality of Data with Neural Networks

G. E. Hinton and R. R. Salaktudinov(2006)

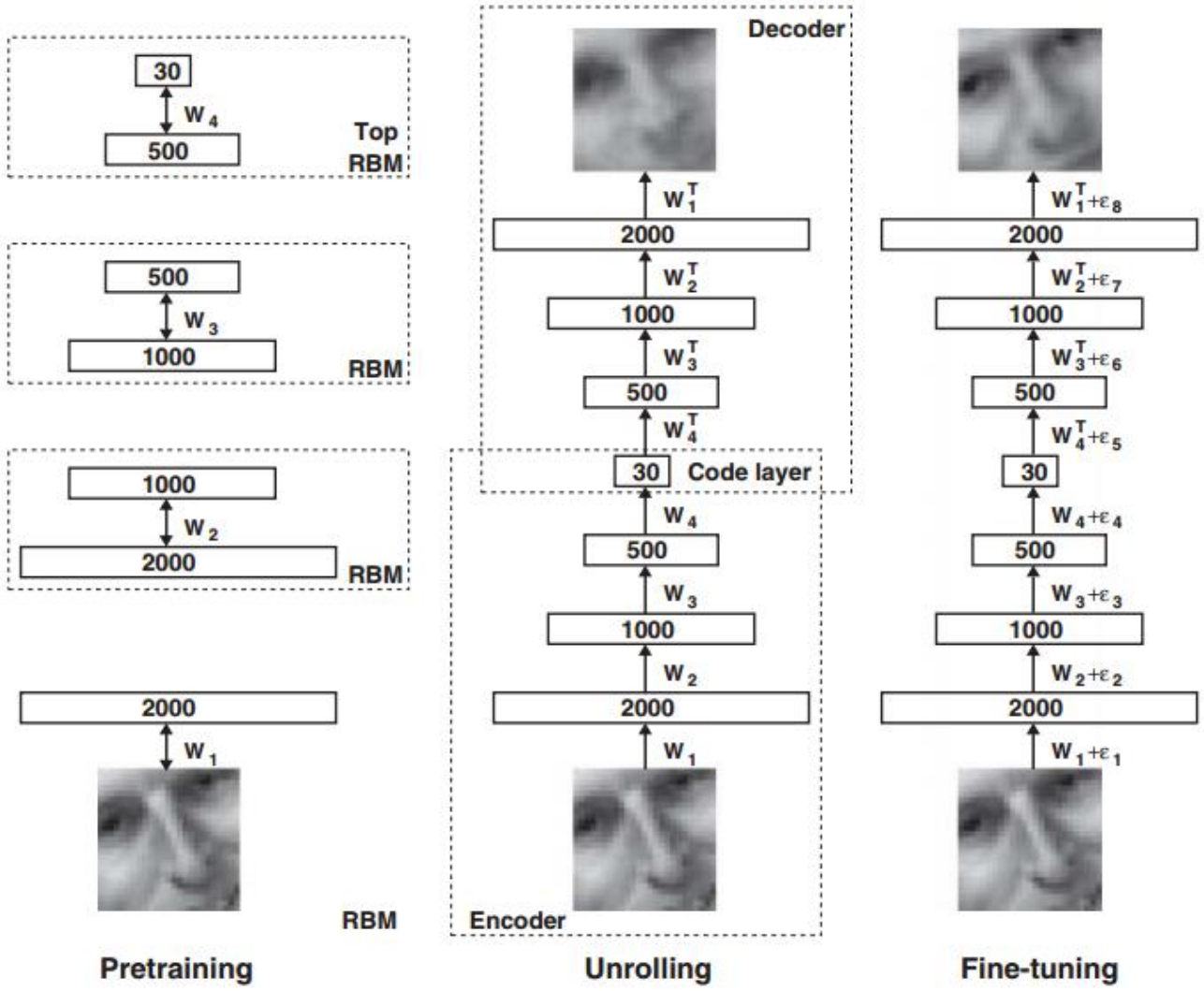
# Problem in a nutshell

- High dimensional data and large NNs
- Dimensionality reduction usually done using
  - Principal Component Analysis(PCA) for images
  - Latent Semantic Analysis(LSA) for documents(words)
- Autoencoders can be used
- But, how to find out "good" initial weights

# Proposed solution

- Pretraining
  - Learn a stack of restricted Boltzmann machines (RBMs)
  - Each RBM has only one layer of feature detectors
  - Each RBM's learned feature activations used as "data" for training the next RBM in the stack
- Unrolling
  - Create a deep autoencoder by "un-rolling" the RBMs
- Fine-tuning (weights)
  - Identify weights via backpropagation of error derivatives

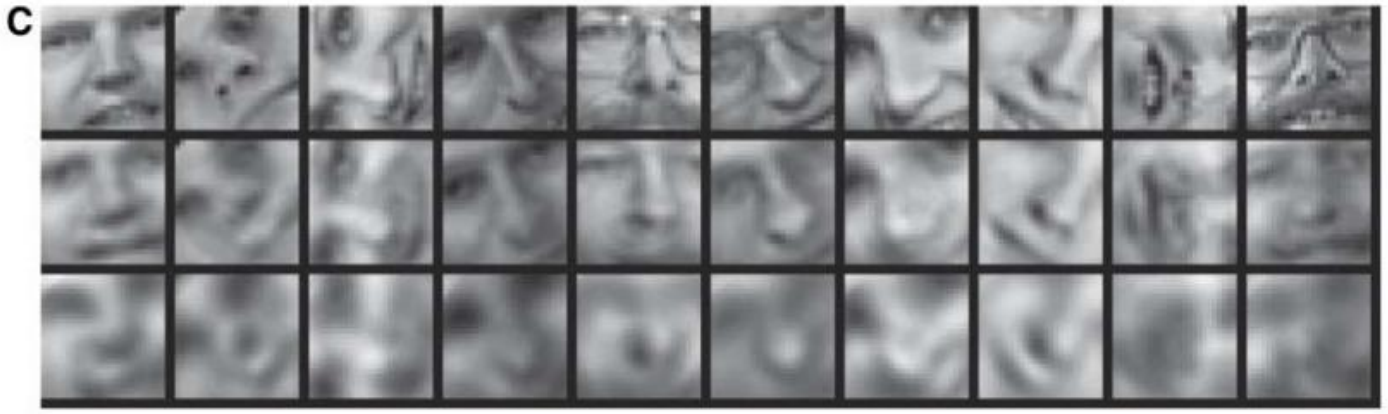
# Proposed solution(visual)



# Experiments and Results I

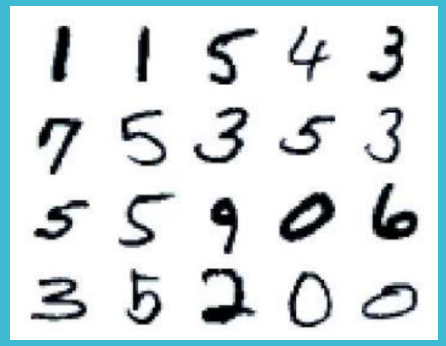
OLIVETTI dataset

- 30-dimensional 625-2000-1000-500-30 autoencoder(2nd row)
- 30- dimensional PCA(3rd row)
- Reconstructed images



# Experiments and Results II

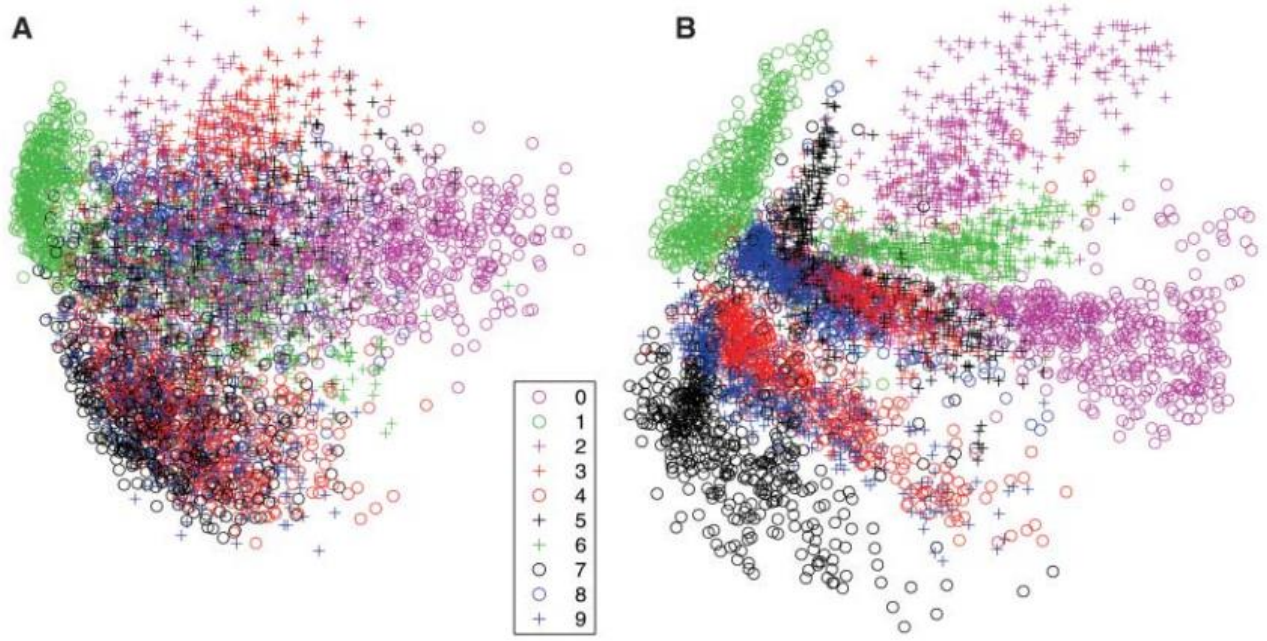
MNIST dataset



- 2-dimensional codes for 500 digits of each of the 10 classes

A) PCA

B) 784-1000-500-250-30 autoencoder



# Experiments and Results III

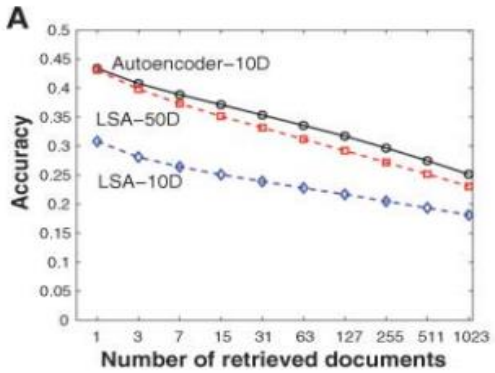
Reuters Corpus Volume I

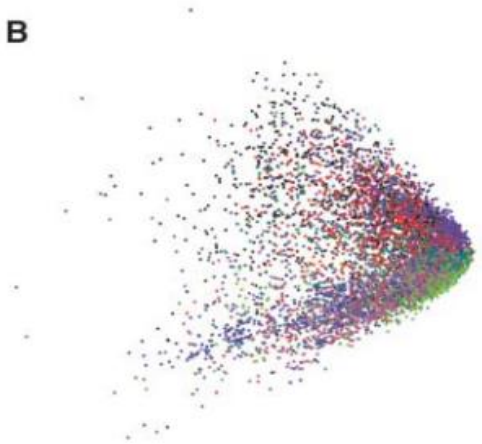
```
<REUTERS TOPICS=''YES'' LEWISSPLIT=''TRAIN''
CGISPLIT=''TRAINING-SET'' OLDID=''12981'' NEWID=''798''>
<DATE> 2-MAR-1987 16:51:43.42</DATE>
<TOPICS><D>livestock</D><D>hog</D></TOPICS>
<TITLE>AMERICAN PORK CONGRESS KICKS OFF TOMORROW</TITLE>
<DATELINE> CHICAGO, March 2 - </DATELINE><BODY>The American Pork
Congress kicks off tomorrow, March 3, in Indianapolis with 160
of the nations pork producers from 44 member states determining
industry positions on a number of issues, according to the
National Pork Producers Council, NPPC.
Delegates to the three day Congress will be considering 26
resolutions concerning various issues, including the future
direction of farm policy and the tax law as it applies to the
agriculture sector. The delegates will also debate whether to
endorse concepts of a national PRV (pseudorabies virus) control
and eradication program, the NPPC said. A large
trade show, in conjunction with the congress, will feature
the latest in technology in all areas of the industry, the NPPC
added. Reuter
\&\#3;</BODY></TEXT></REUTERS>
```

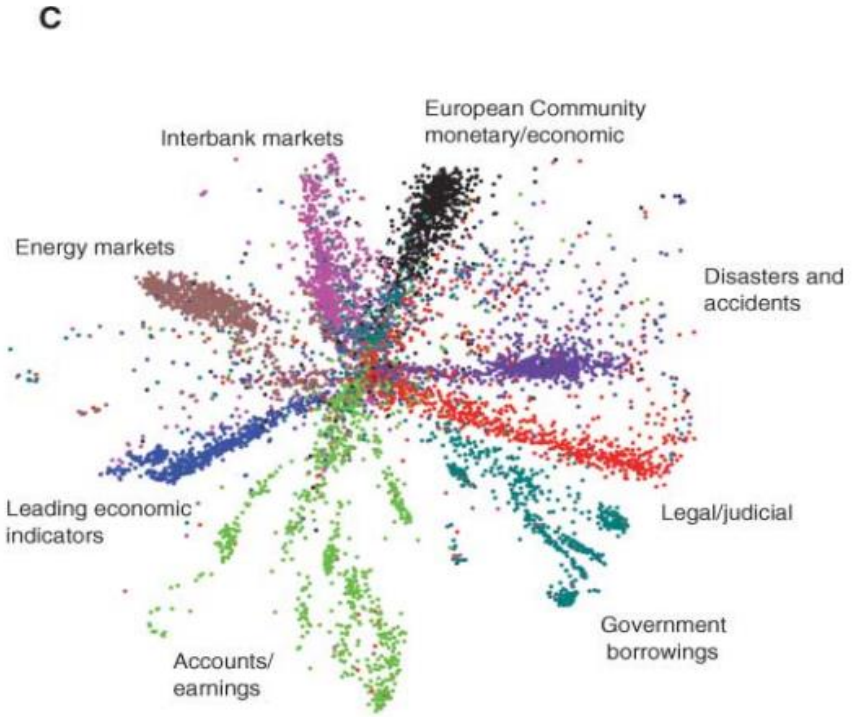
- 2-dimensional codes for 804,414 newswire stories

B) Latent Semantic Analysis (LSA)

C) 2000-500-250-125-10 autoencoder



# Summary

- Proposed pre-training methodology for autoencoders that removes unimportant variations in input data
- Implemented the same and achieved optimal results
- Outperformed standard dimensionality reduction techniques(PCA, LSA) on popular image recognition and natural language processing datasets



Image credits: lifehacker.com

# References


- Hinton, G.E. and Salakhutdinov, R.R., 2006. Reducing the dimensionality of data with neural networks. *science*, *313*(5786), pp.504-507.
- Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.R., 2012. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.