Topological Data Analysis and Deep Learning

University of Lisbon
September 30, 2020

Gunnar Carlsson

Stanford University

What is Deep Learning?

- Methodology based on neural networks
- Has produced outstanding classification results for complex data
- Images
- Text
- Molecules (Guowei Wei)

Problems

- Adversarial examples
- General lack of transparency
- Limits usefulness in many key domains, financial regulation, health care
- Would like to be able to learn more complex models

Neural Networks

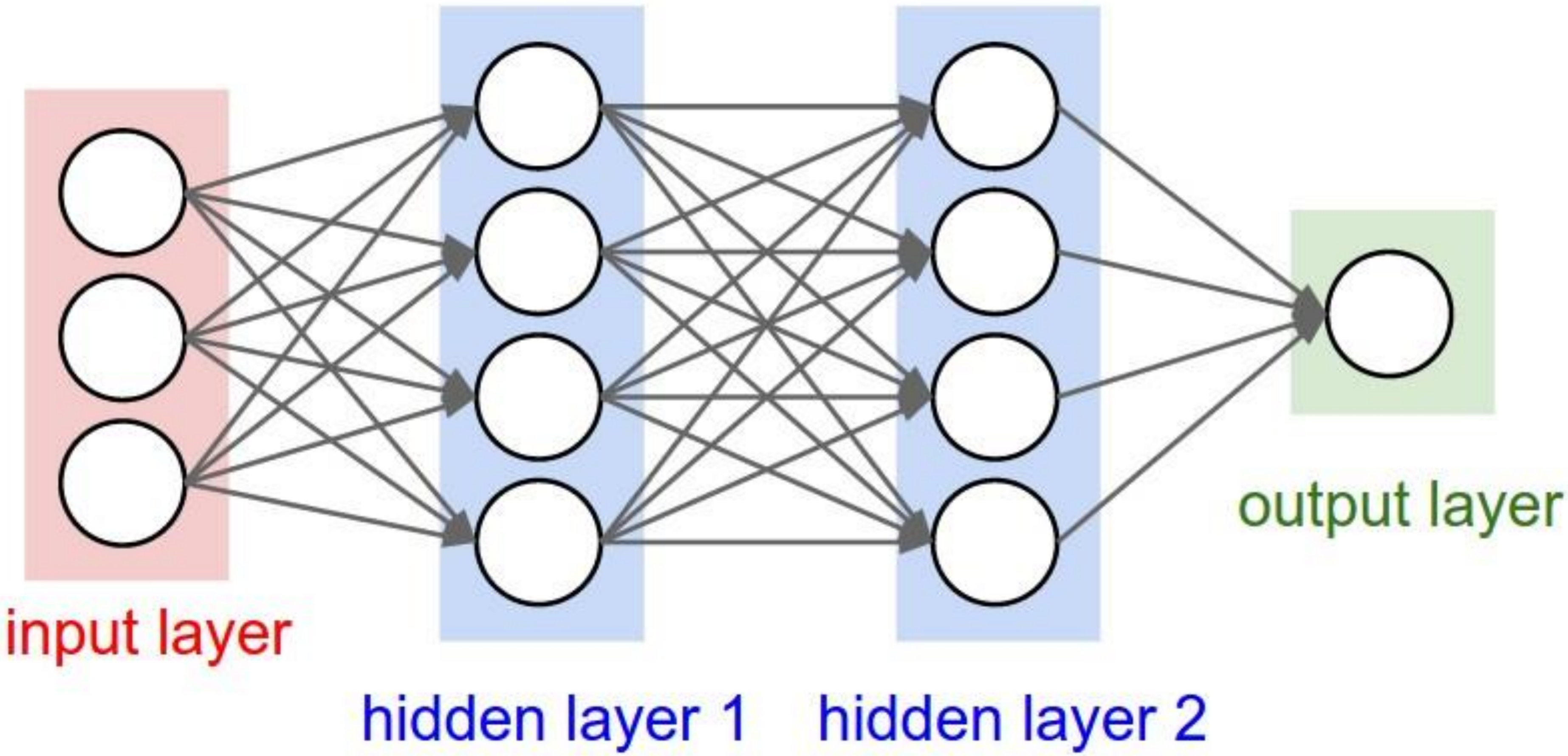
Formal Description of Neural Networks

- Given a directed graph Γ
- A *state* σ of the directed graph is an assignment of a value (real or Boolean) to each vertex
- A *coefficient system* χ for Γ is an assignment of a value (real or Boolean) to each edge
- For any vertex v, we let $\wp(v)$ denote the collection of all pairs (w, e), where w is a vertex of Γ, $e = (w, v)$ is a directed edge of Γ
- Fixing a coefficient system χ and a state σ of Γ, we define the *update* of σ by χ at a vertex v to be a *fixed* symmetric formula of the collection of pairs (σ_w, χ_e) for all $(w, e) \in \wp(v)$

Formal Description of Neural Networks

- Can be regarded as an automaton, with the state set being vectors with coordinates the vertices of Γ, and with the actions being given by the choice of a coefficient system.
- Can also be regarded as a computational formula, with the input variables being the values of the state, and the "coefficients" begin the values of the coefficient system.
- Many different kinds of neural networks.
- The complete directed graph on a set can be used, but is difficult to handle computationally.
- A large useful class is *feed forward networks*

Neural Networks



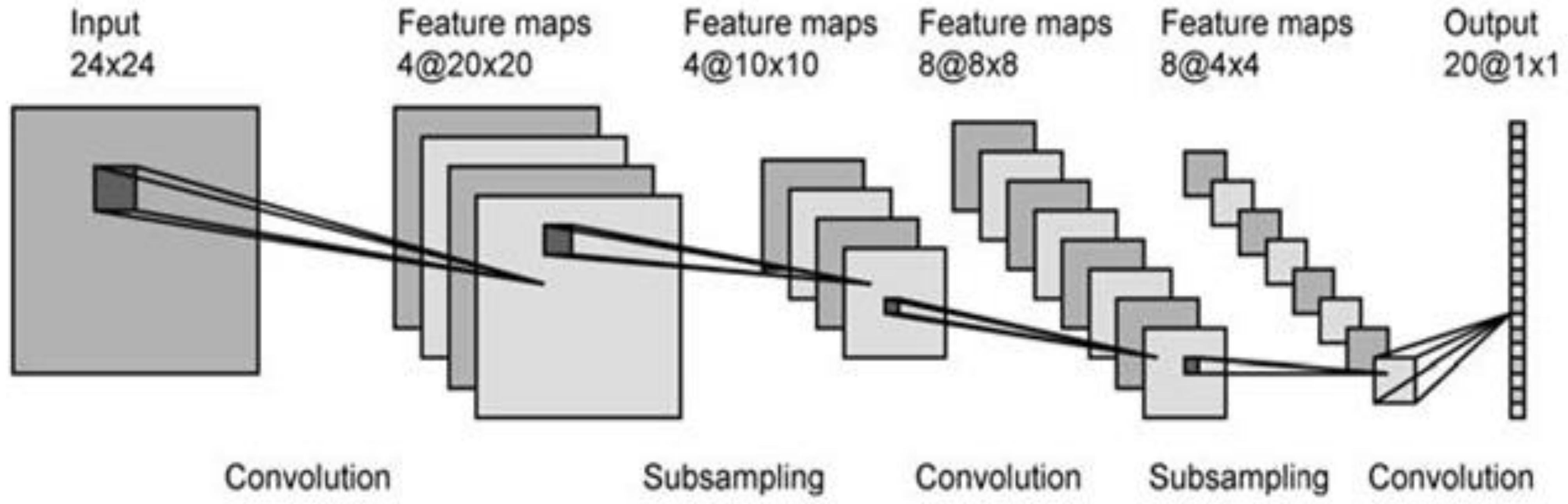
Neural Networks

- Given a data set and an output function, perhaps Boolean valued
- Weights are assigned to the directed edges of the network
- Activation at a node is computed using uniform function of activations of nodes connected to it
- Network is ”trained” by optimization algorithms acting on the set of weights
- Final output is a formula (very large) determined by the final set of weights

Convolutional Neural Networks

- Structure of network adapted to specific cases
- Images (2D rectangular arrays)
- Text (1D arrays)
- Time series (1D arrays)

Convolutional Neural Networks



Does Learning by CNN’s Behave Like Human Learning?

Joint work with Rickard Brüel Gabrielsson

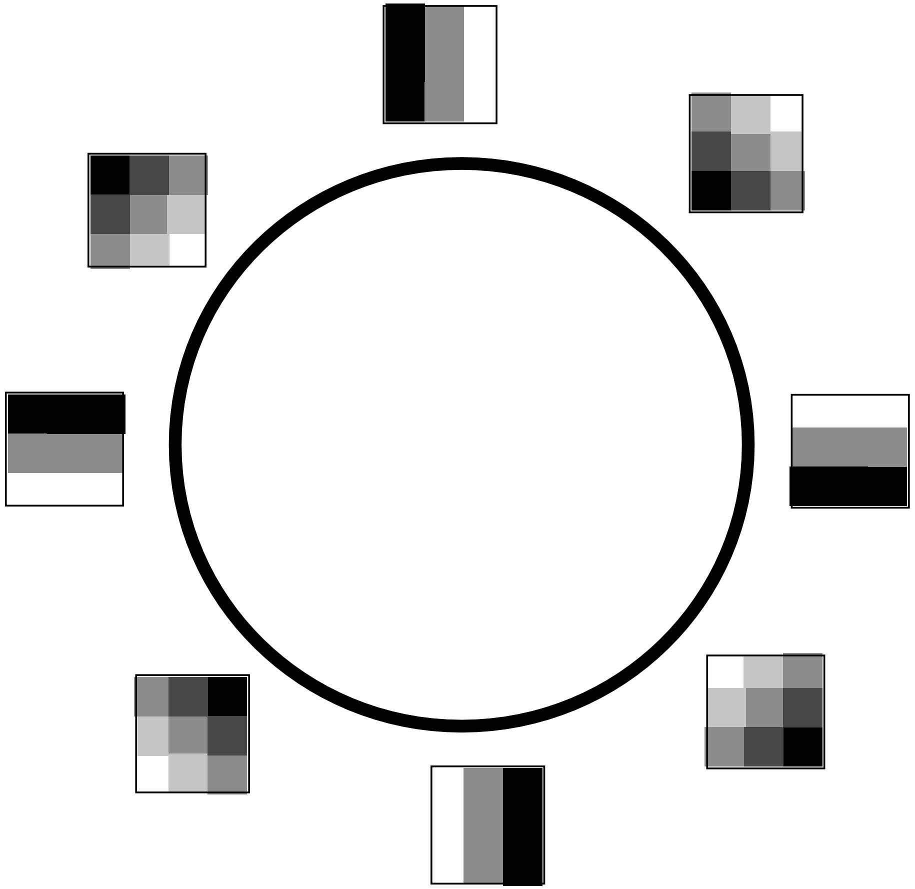
What Do We Want to Know?

- Can we see similarities to what we have found in image patch data?
- What happens as the network learns?
- What are the “responsibilities” of the various layers?

Mumford Data Set (De Silva, Ishkhanov, Zomorodian, C.)

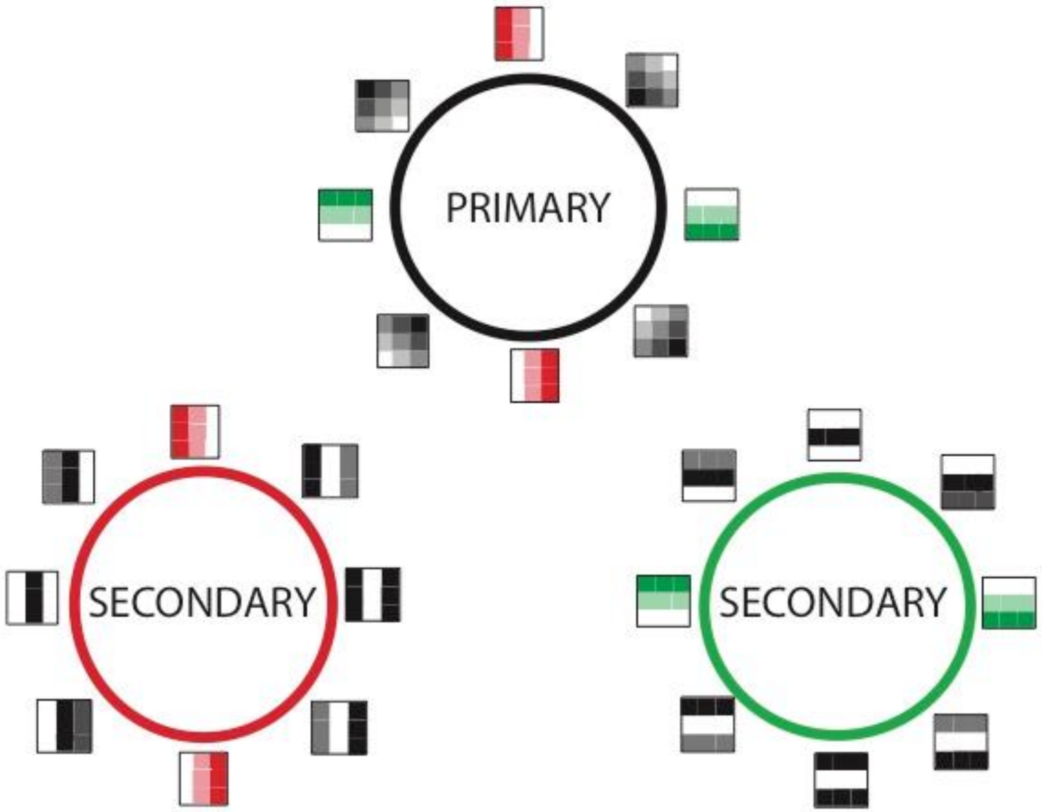
- Analysis of a data set of 3 x 3 patches in natural images
- Studied only ”high variance” patches
- Studies only densest such patches (frequently occurring motifs) – density proxies of varying locality
- Motivated by goal of understanding how tuning of neurons in visual cortex is affected by statistics of natural images

Image Patch Analysis: Primary Circle



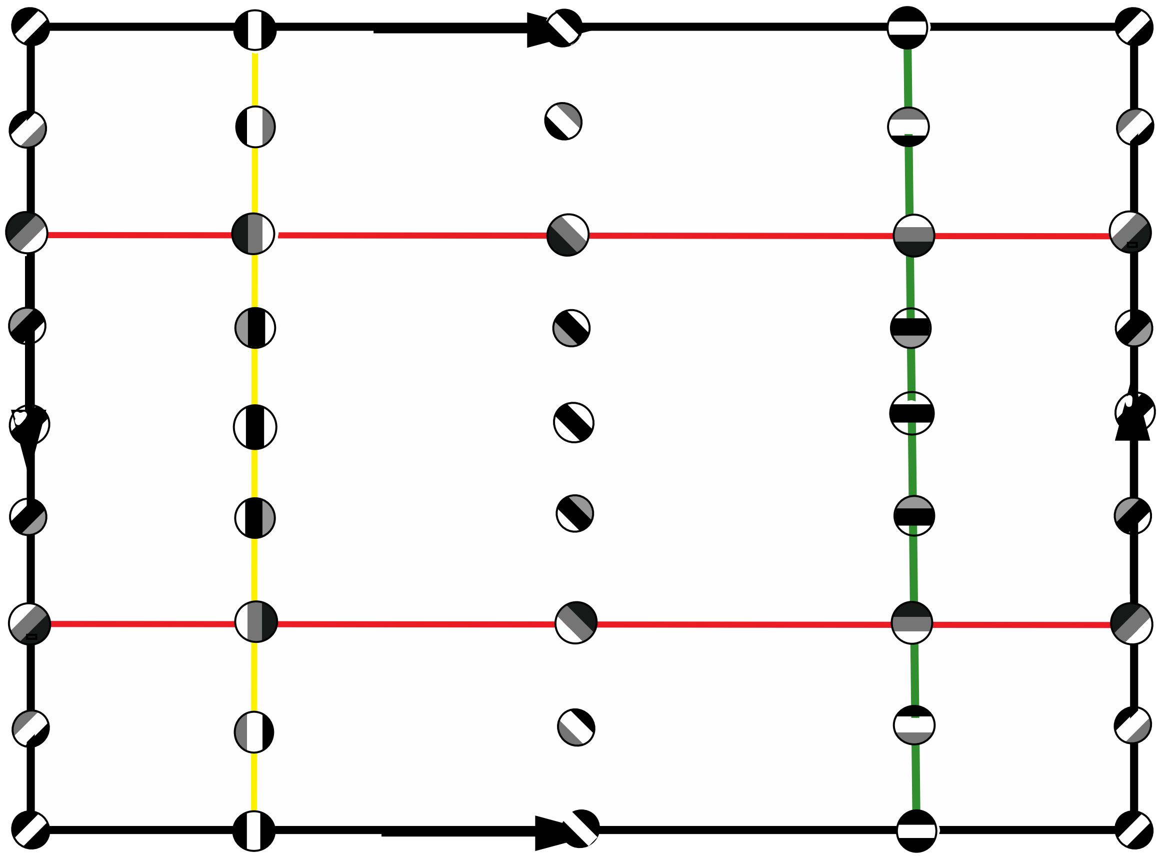
Highest density high variance patches – non-local density measure

Image Patch Analysis: Three Circle Model



More local density measure

Image Patch Analysis: Klein Bottle

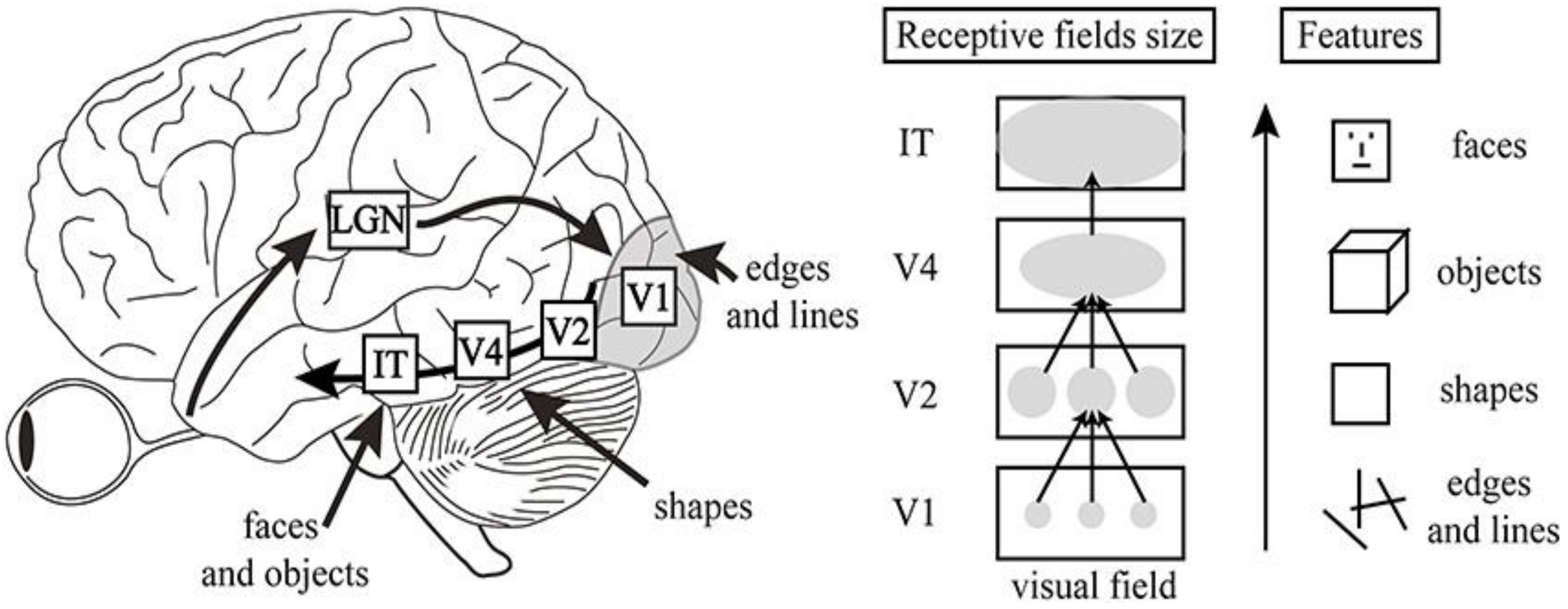


Still weaker threshold

Primary Visual Cortex

- Primary visual cortex (V1) lowest level processing beyond retina
- Higher levels (V2,V4, LGN, etc.) perform more abstract tasks
- Hubel-Wiesel show that individual neurons detect edges and lines
- Consistent with idea of compression of frequent signals

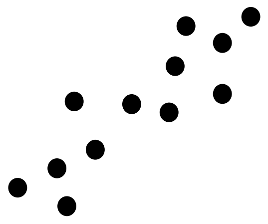
Visual Pathway



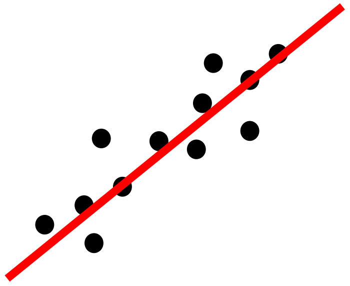
Topological Data Analysis

Data has shape and the shape matterw

The Shape of Data

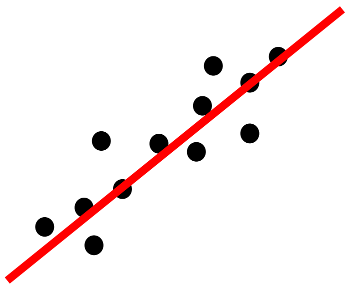


The Shape of Data

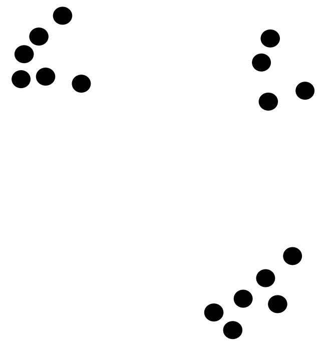


Regression

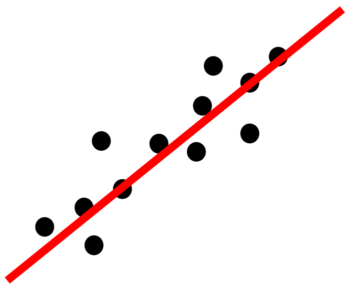
The Shape of Data



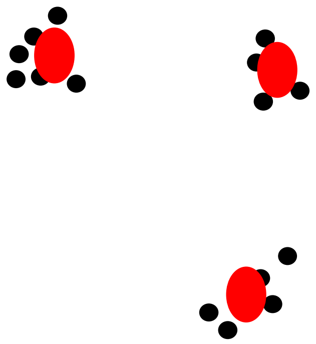
Regression



The Shape of Data

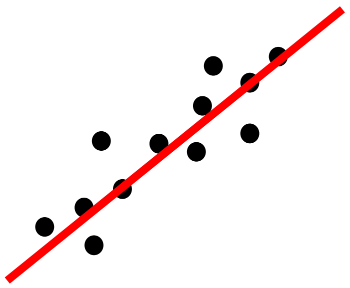


Regression

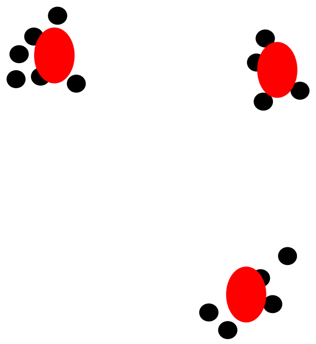


Clustering

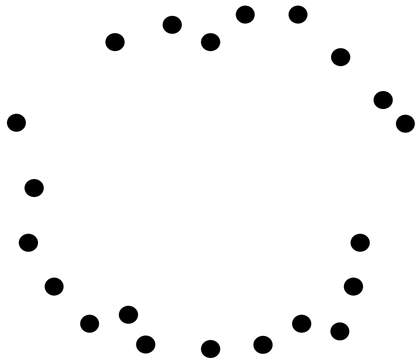
The Shape of Data



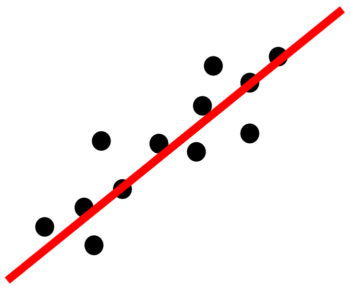
Regression



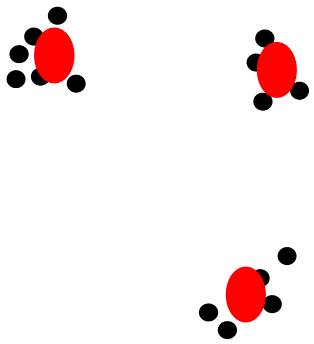
Cluster



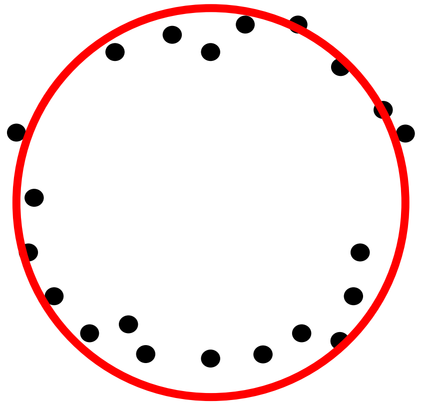
The Shape of Data



Regression

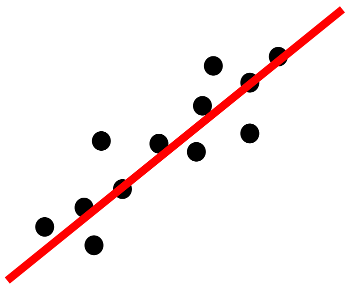


Clustering

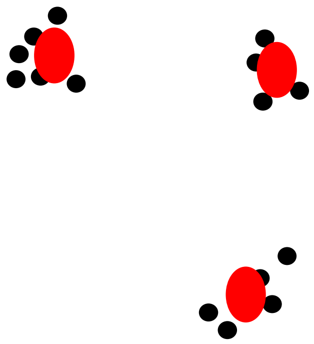


Loop

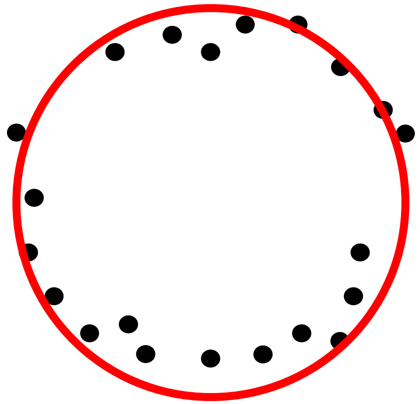
The Shape of Data



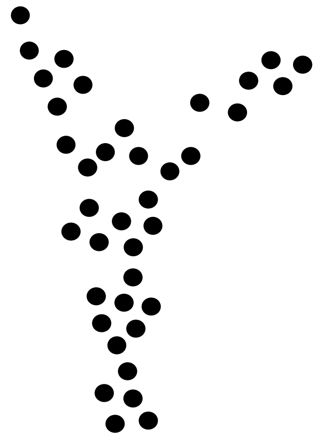
Regression



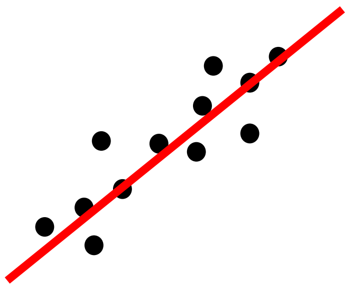
Cluster



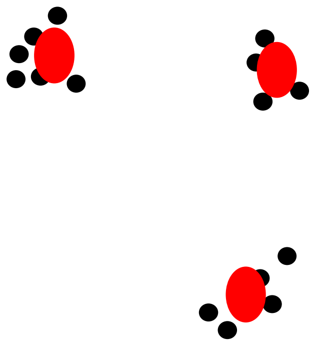
Loop



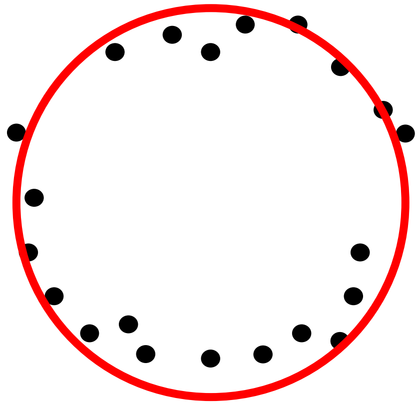
The Shape of Data



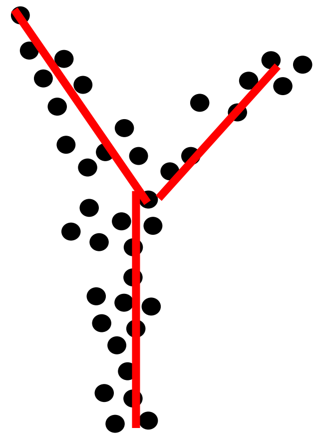
Regression



Cluster

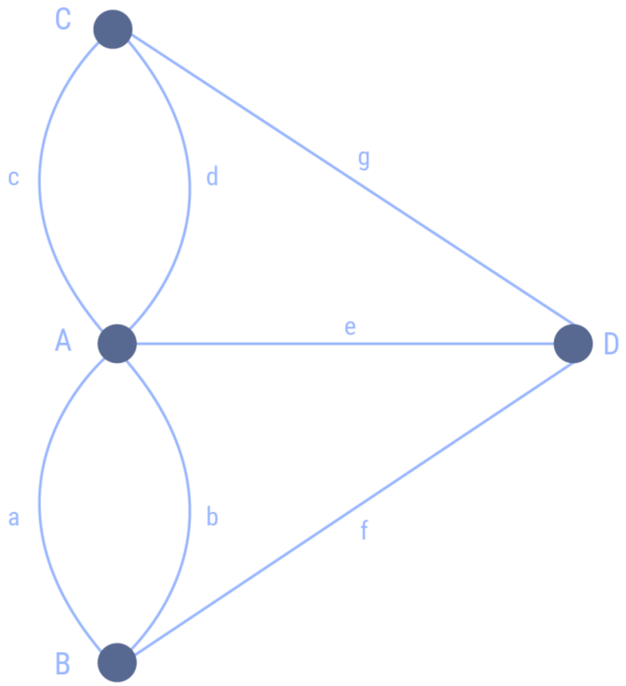


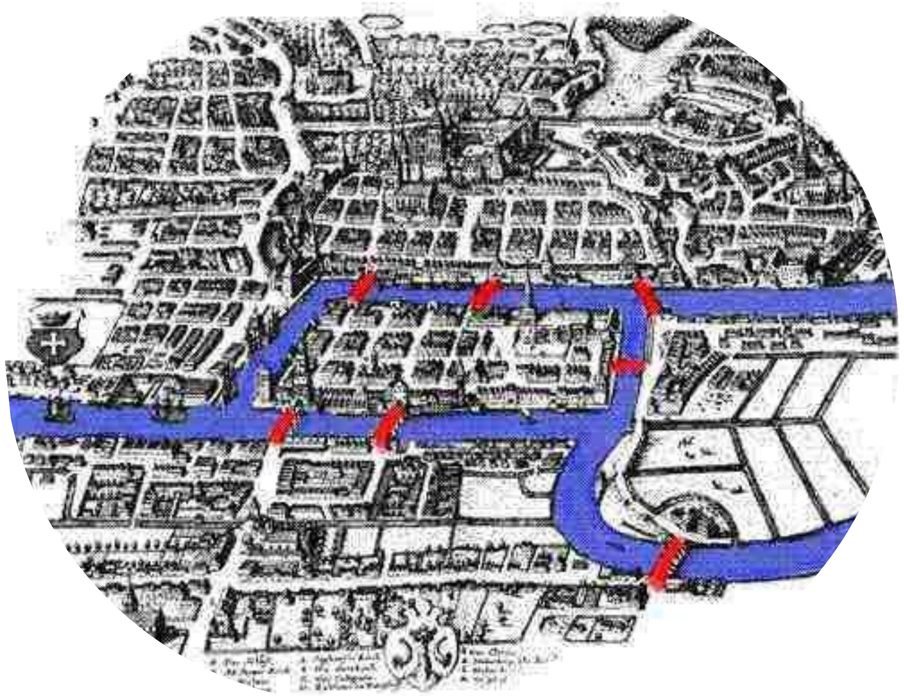
Loop



Flared

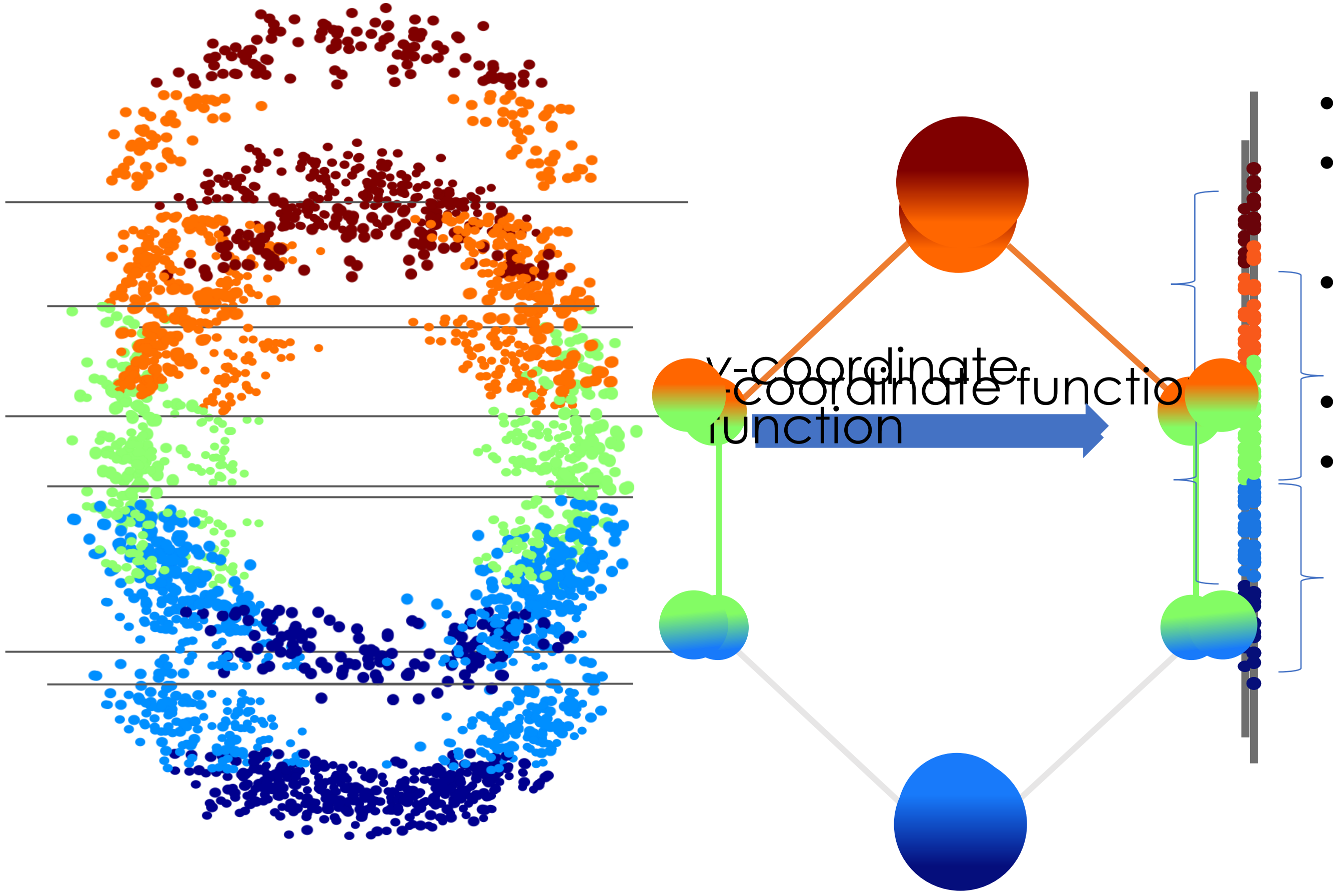
Topology





Königsberg Bridges

How to Build Networks – Mapper Construction

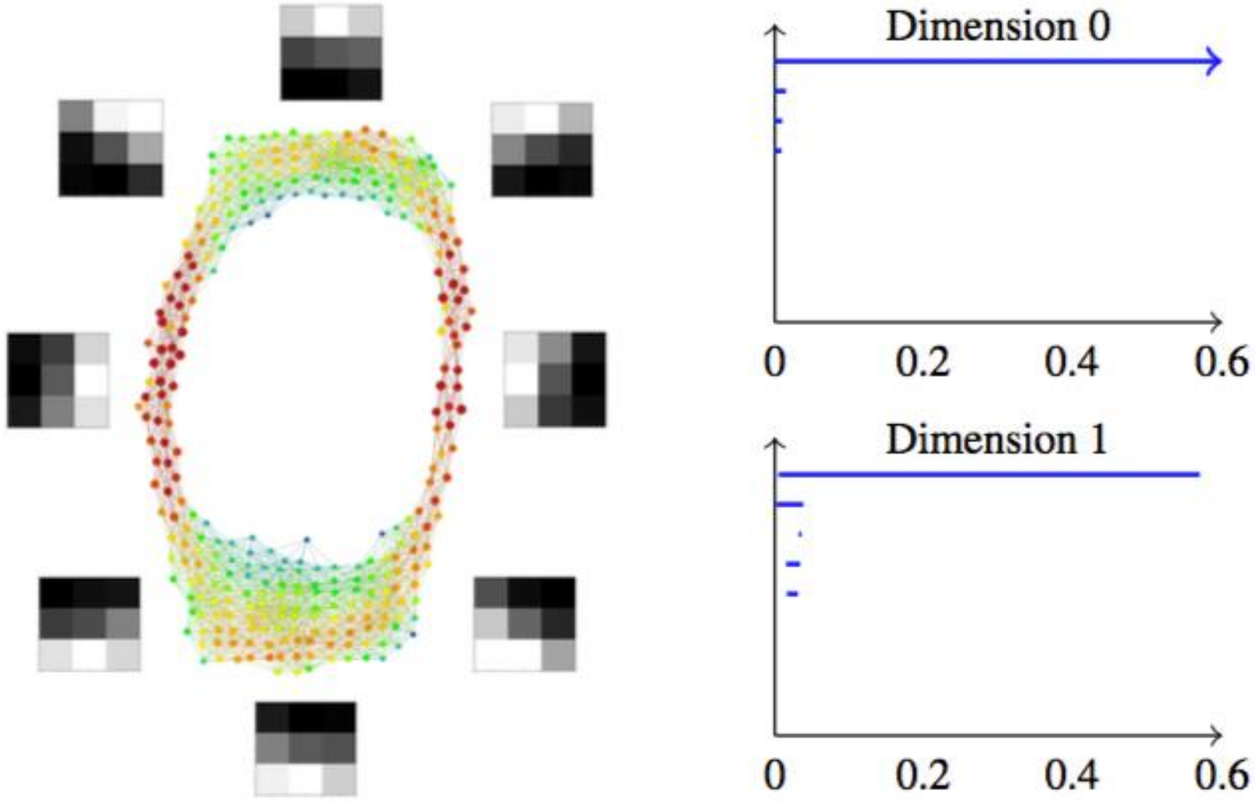


- Apply a projection to the data set
- Use the projection to bin the data into overlapping bins
- Cluster each bin using a fixed clustering method (requires data equipped with metric)
- Create a node for each “partial cluster”
- Create an edge between any two nodes whose corresponding clusters overlap

Topological Modeling

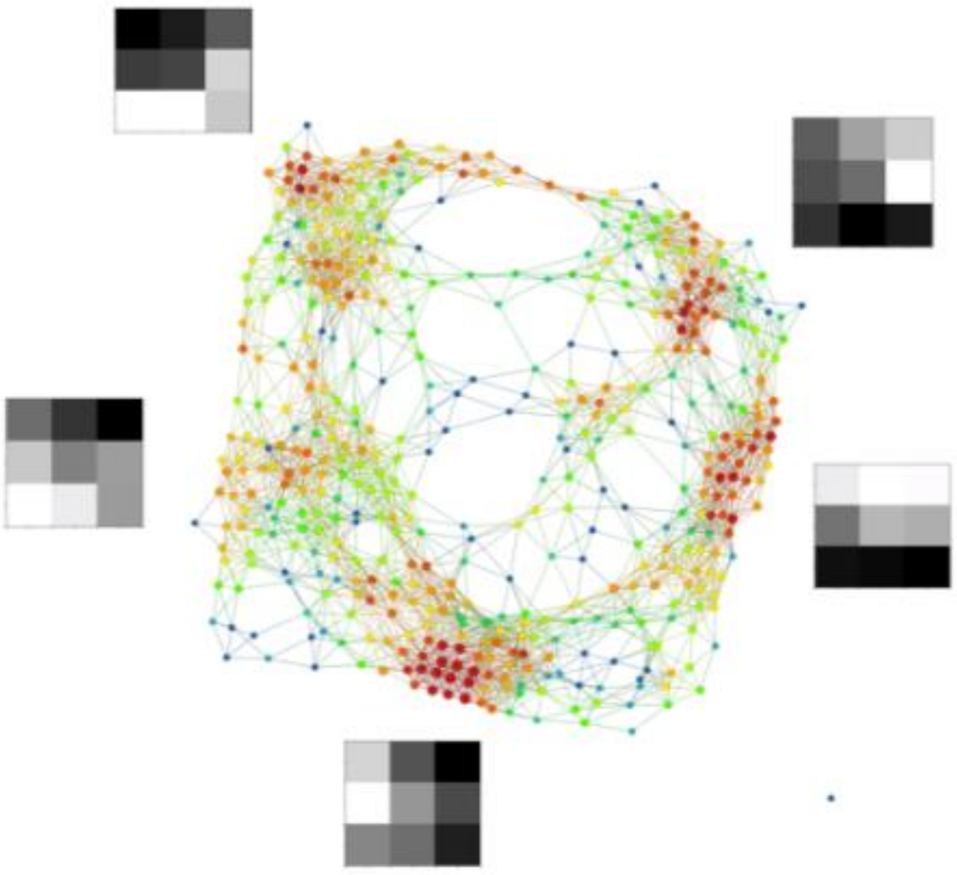
- Model output is a network, not equations
- Not just visualization, many capabilities
- Selection
- Segmentation
- Coloring, hot spot analysis
- Explain, justify
- Topological feature creation, selection
- Model assessment and improvement

Topological Analysis of Weight Spaces (MNIST)



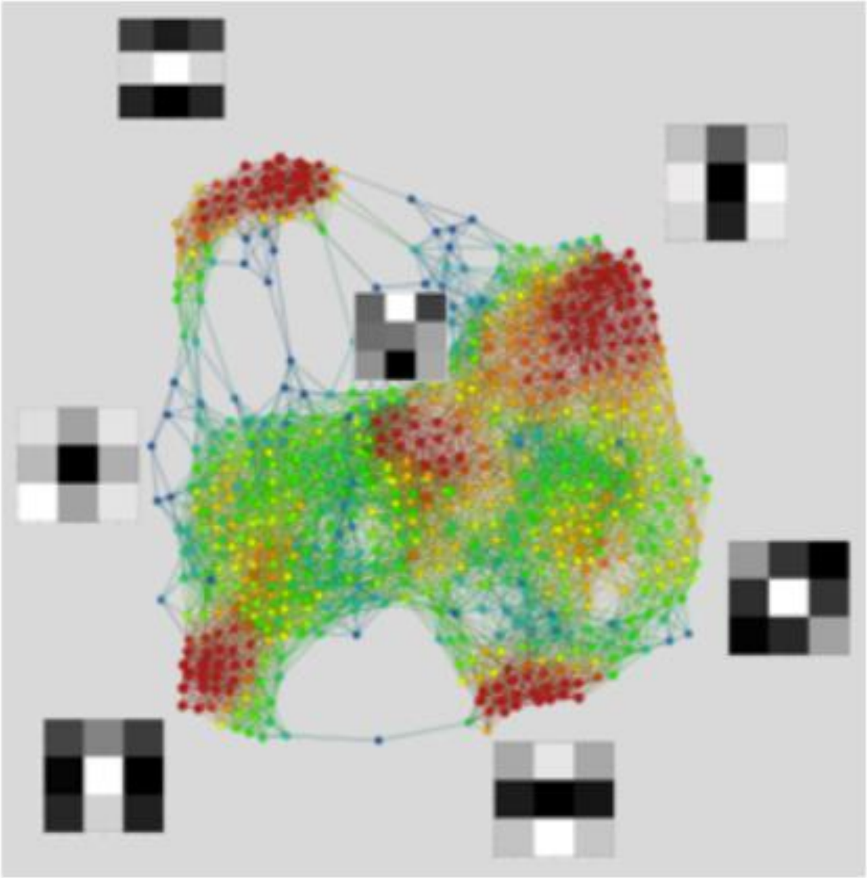
Non-local density thresholding for layer 1 of depth 2 net

Topological Analysis of Weight Spaces (MNIST)



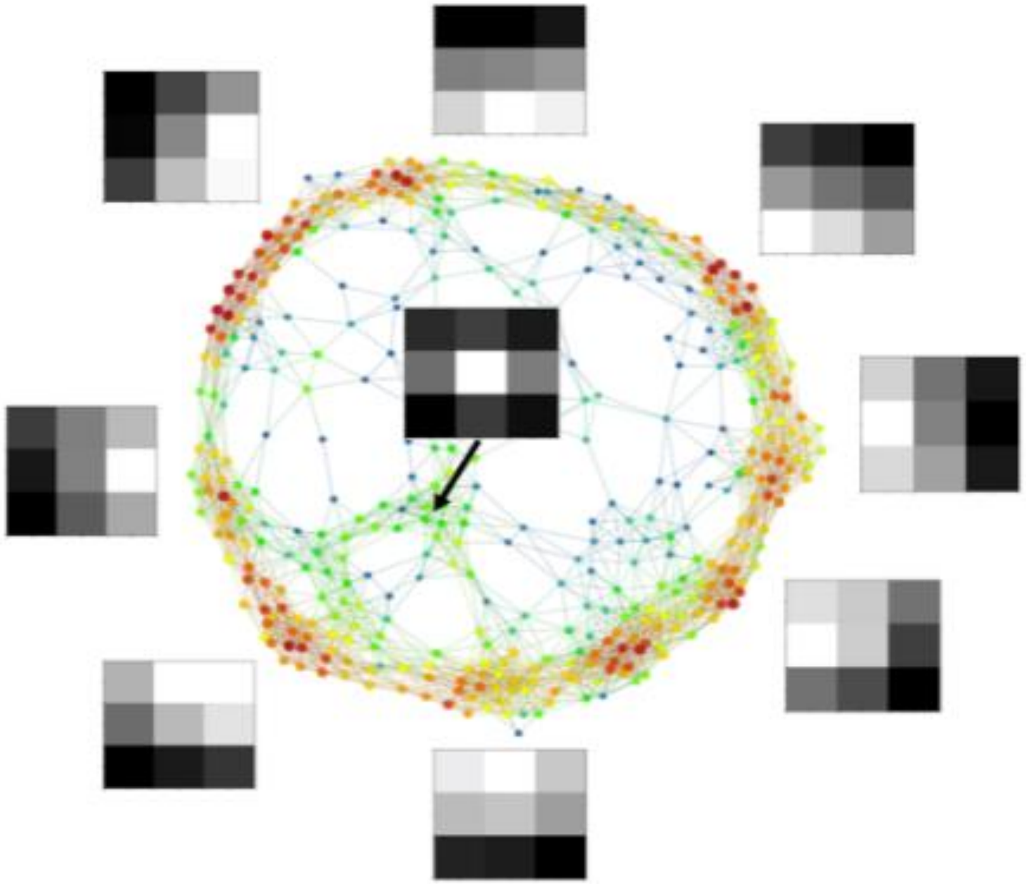
More localized density measurement for layer 1 of depth 2 net

Topological Analysis of Weight Spaces (Cifar10)



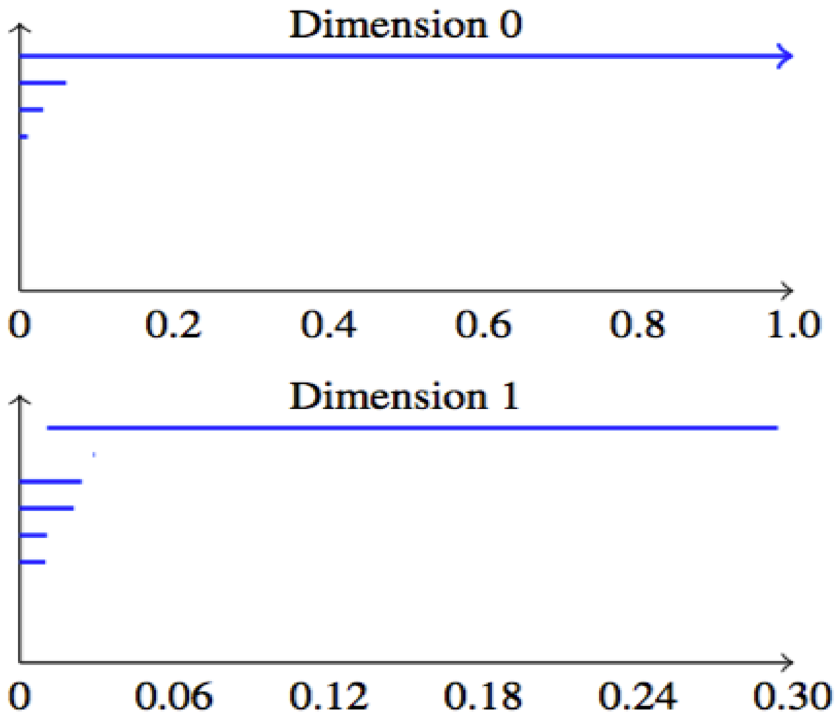
1st Layer of CNN for this data set, reduced to gray scale

Topological Analysis of Weight Spaces (Cifar10)



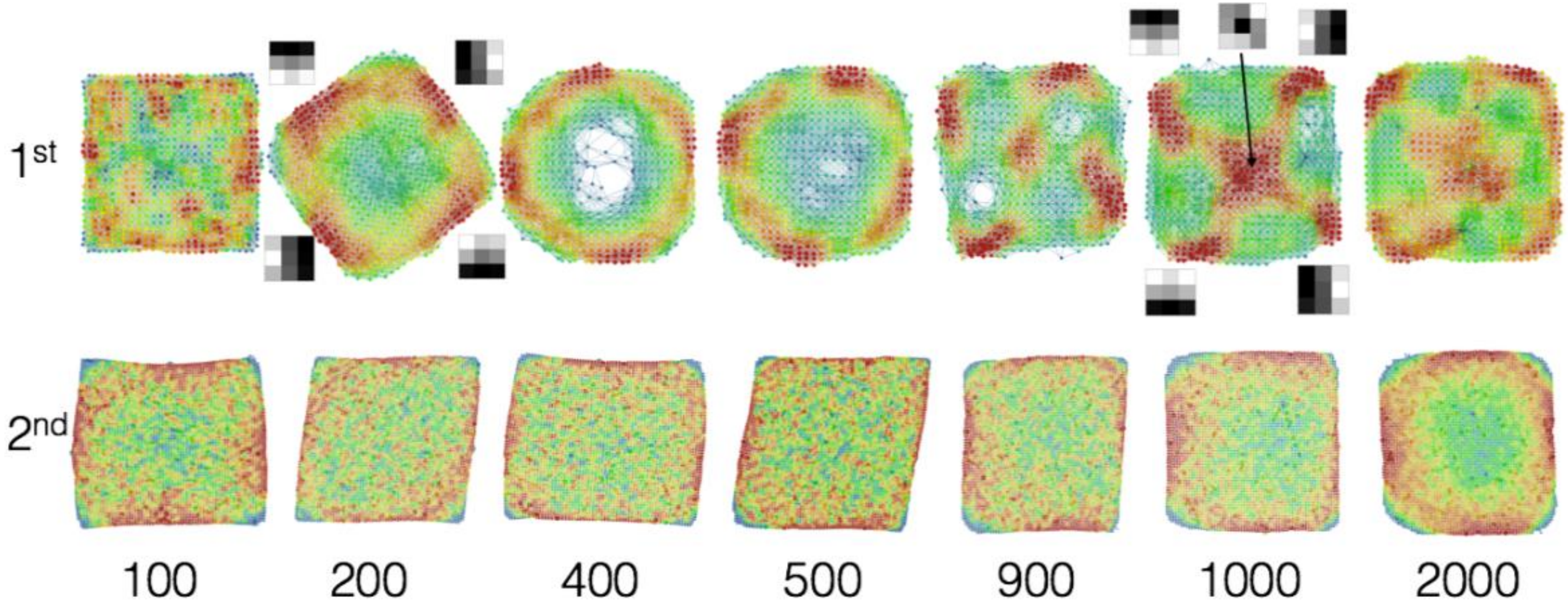
2nd Layer for single CNN, reduced to gray scale

Topological Analysis of Weight Spaces (Cifar10)



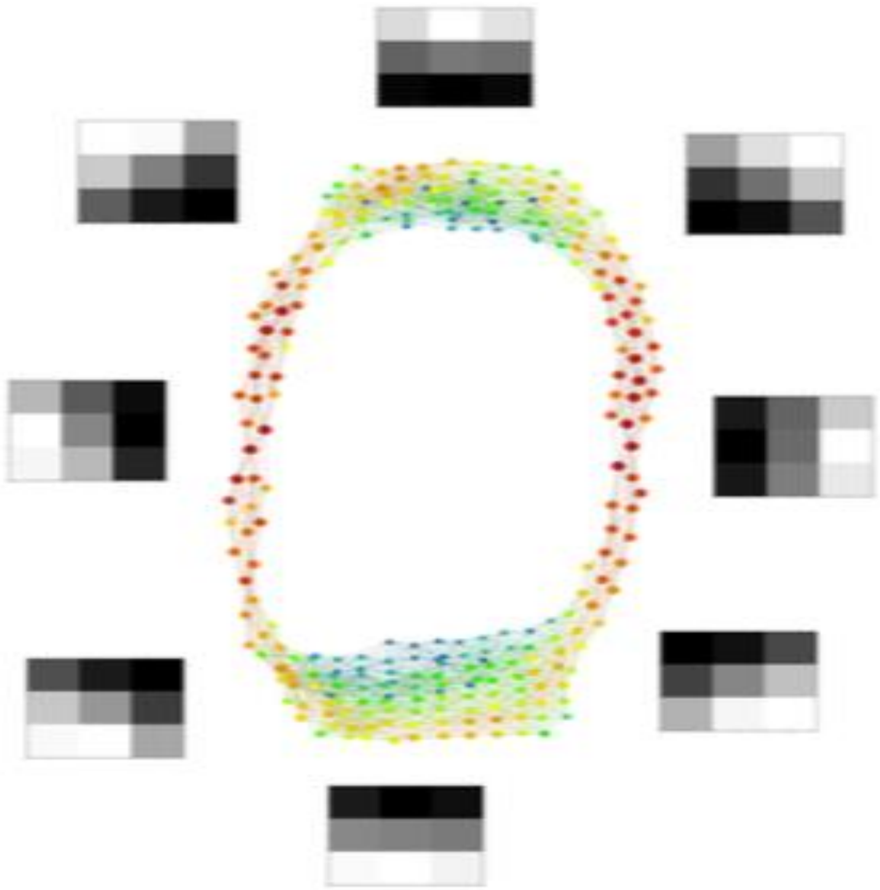
1D barcode for tightly thresholded data set of for 2nd layers , reduced to gray scale

Topological Analysis of Weight Spaces (Cifar10)



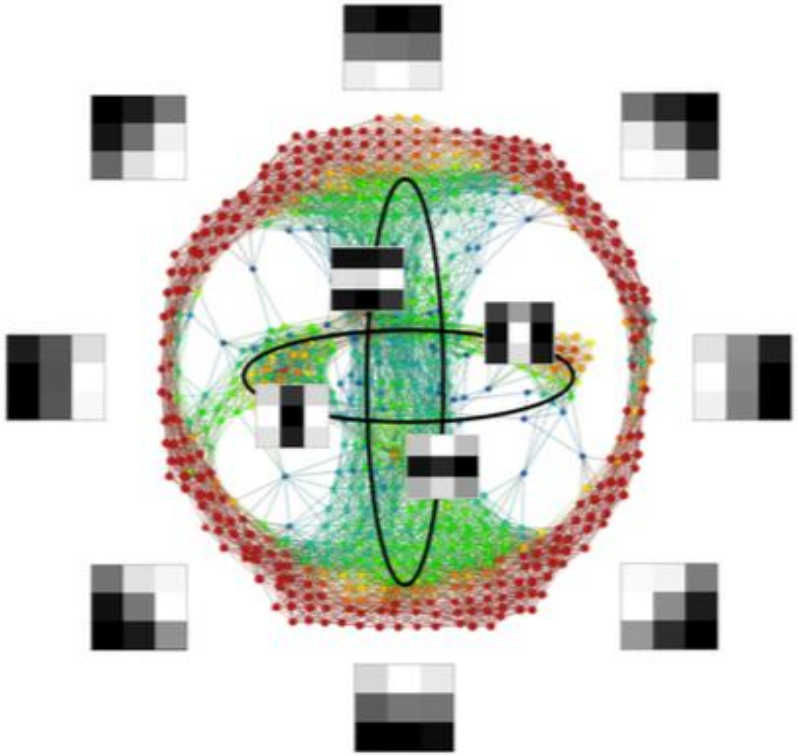
Mapper representations over the number of iterations, tightly density thresholded, gray scale reduced

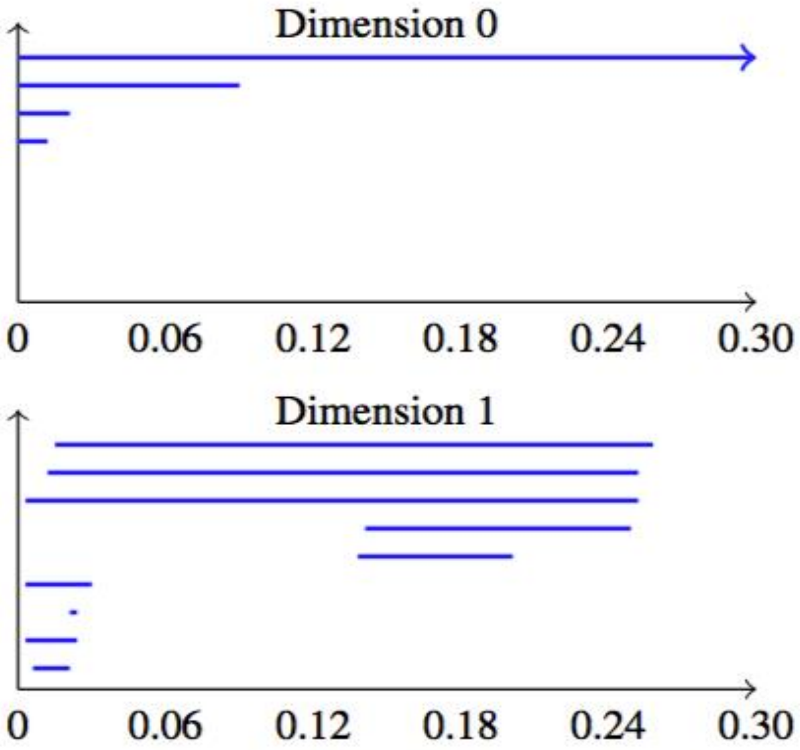
Topological Analysis of Weight Spaces (Cifar10)



1st layer, Coarse density thresholding, color retained

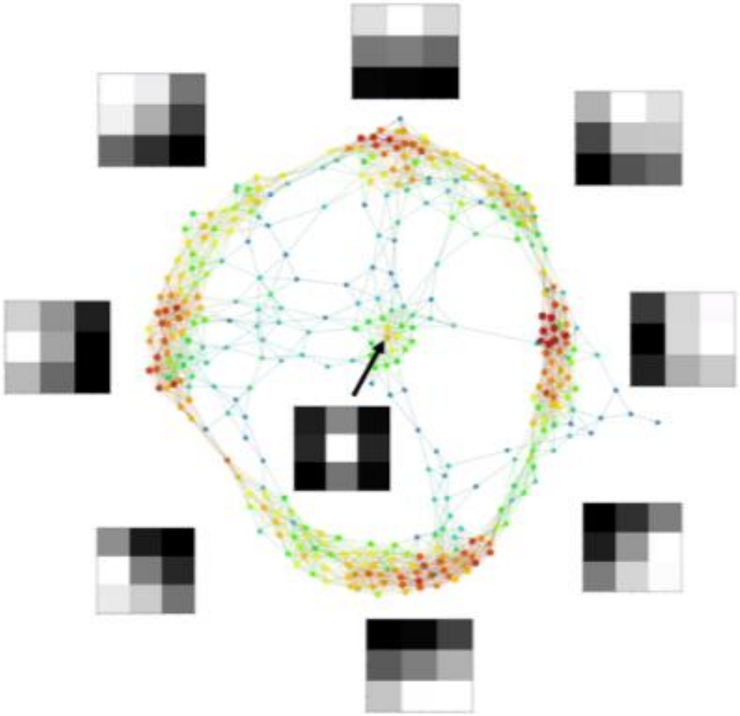
Topological Analysis of Weight Spaces (Cifar10)





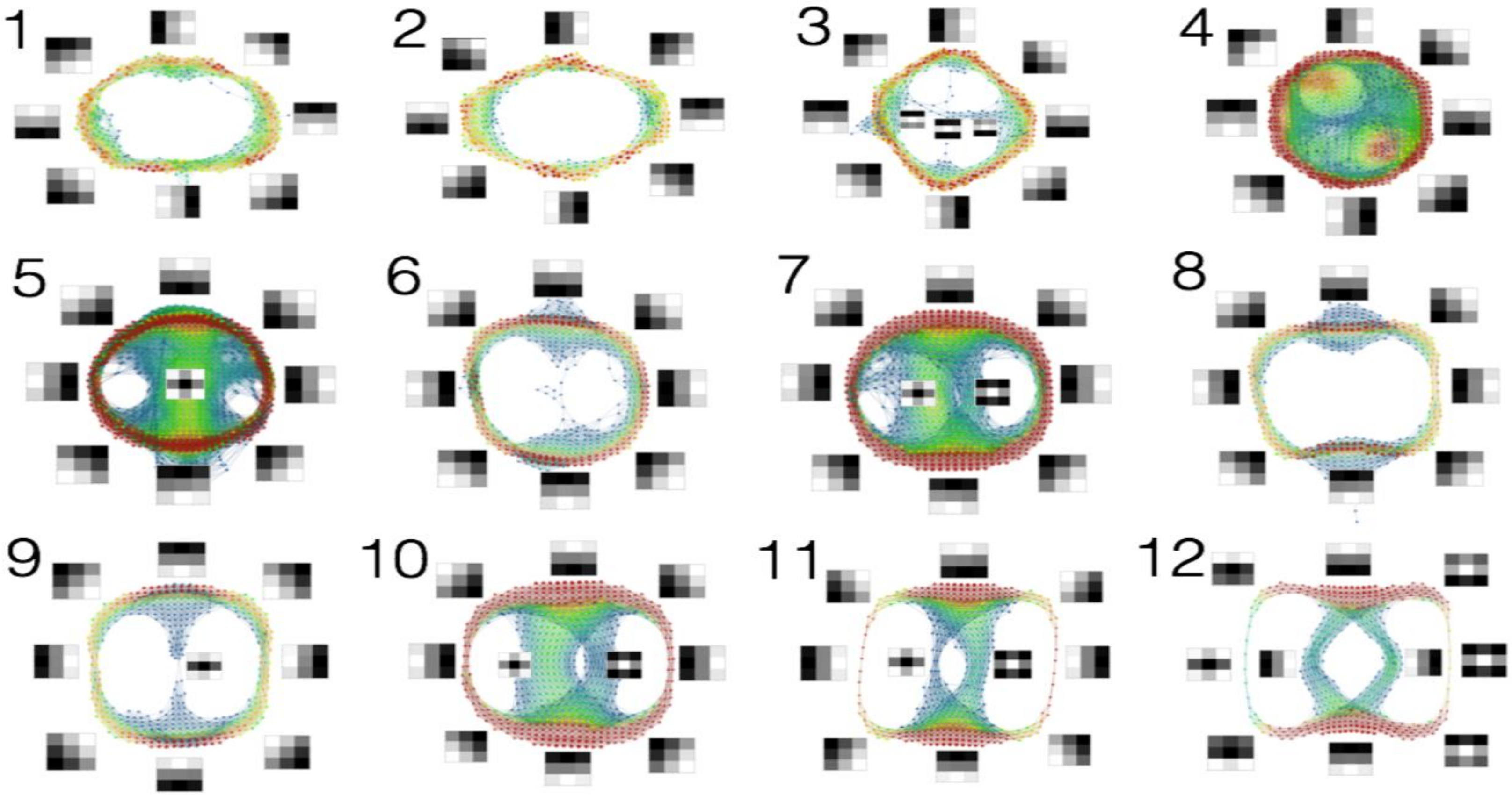
1st layer, looser density thresholding, more localized density estimator, color retained

Topological Analysis of Weight Spaces (Cifar10)



2^{nd} layer, fine density thresholding, color retained

Topological Analysis of Weight Spaces (VGG16)



Mapper findings from each of 13 layers, same density thresholding, relatively local estimator

Hard Code Primary Circle and Klein Bottle

- Allows both speed up and generalization
- Speed up of 2X for MNIST and 3.5X with SVHN
- Generalization from MNIST to SVHN doubles accuracy of standard method (10% to 22%)

TDA and Deep Architectures

Convolutional Situation

- Convolutional nets use the grid structure on the set of pixels in a critical way
- Pixels are the features for image data sets
- We use this a priori geometry and use it in designing the neural net
- Important in restricting the class of formulas that are considered
- Restrict to formulas that are “local” in the geometry of feature space
- Thought experiment: Time series – learn derivatives
- Images - learn directional derivatives
- What if there is no a priori geometry?

Discovered Geometry

- We have discovered feature sets that carry geometry
- Primary circle , Klein bottle, etc.
- Can one use geometry to design architectures?
- Need a formalism

Feature Space Modeling

- Given a data matrix, one can also consider the transpose matrix
- The rows of the transpose are the features of the data set
- When there are many features, very useful to create Mapper models
- Compresses and recognizes correlations among features
- Each row of original matrix gives a function on feature set, and on nodes of the topological model

Microarray Analysis of Breast Cancer

Cohort B

Cohort C

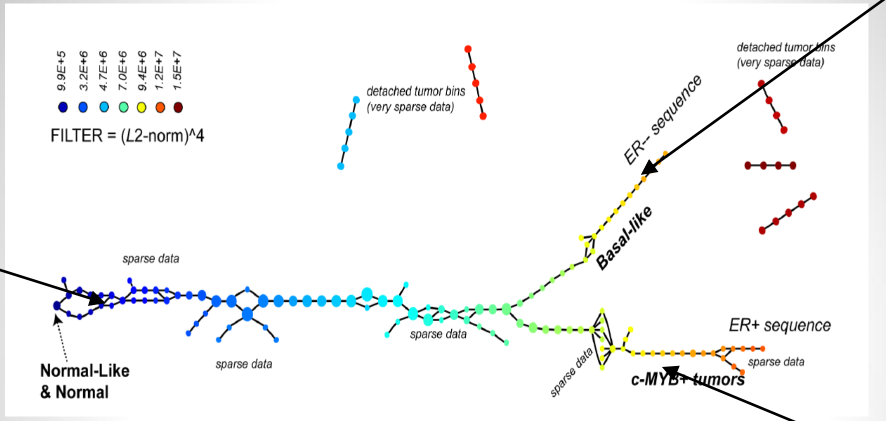
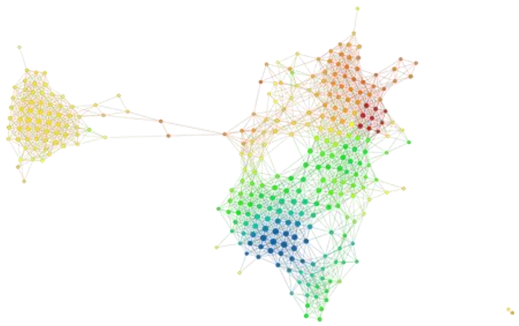


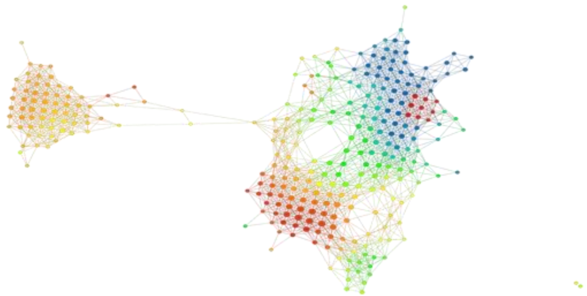
Diagram of gene expression profiles for breast cancer
M. Nicolau, A. Levine, and G. Carlsson, PNAS 2011

Cohort A

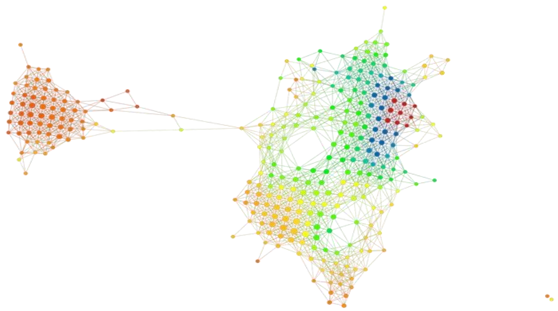
Explaining the Different Cohorts



Cohort A



Cohort B



Cohort C

UCSD Microbiome

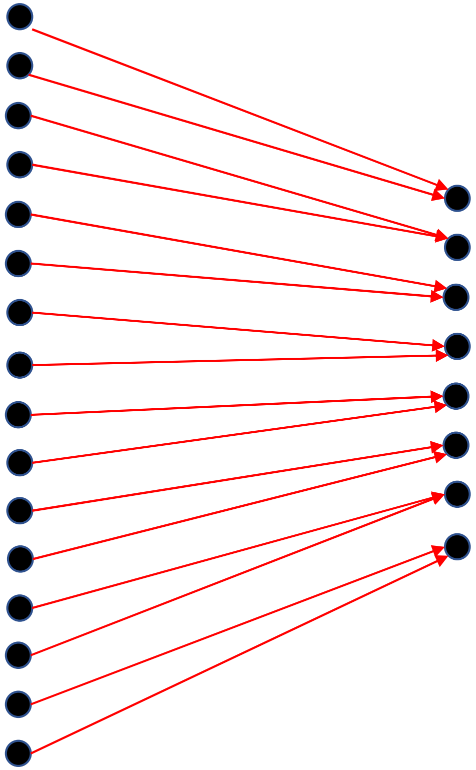
Generalized Convolutional Nets

- Feed-forward structure: $\{X_0,\ldots,X_n,C_{1,}\ldots, C_n\}$
- X_i's are sets, $C_i:X_{i-1} \longrightarrow X_i$ is a correspondence
- This data can be used to create a network
- Nodes are elements in all X_i's, each X_i is a layer
- Correspondence determines connections from $X_{i-1 \text{ to}}\, X_i$
- $v \longrightarrow w$ if and only if (v,w) is in C_i

Building Correspondences

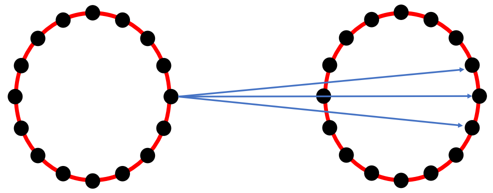
- Complete correspondences build “fully connected” networks. Can be used in conjunction with other correspondences to make useful things.
- Functional correspondences – graphs of functions.
- Products, compositions of correspondences
- Pooling correspondences

Pooling Correspondence



Metric and Graph Correspondences

- For any metric space X and threshold *r*, have a metric space correspondence from X to X, with (v,w) a member if d(x,y)$\leq r$
- Convolutional nets are an example
- For any graph, have graph correspondence from vertex set V to itself, with (v,w) in the correspondence if and only it (v,w) is an edge
- E.G. discretized circles and Klein Bottles



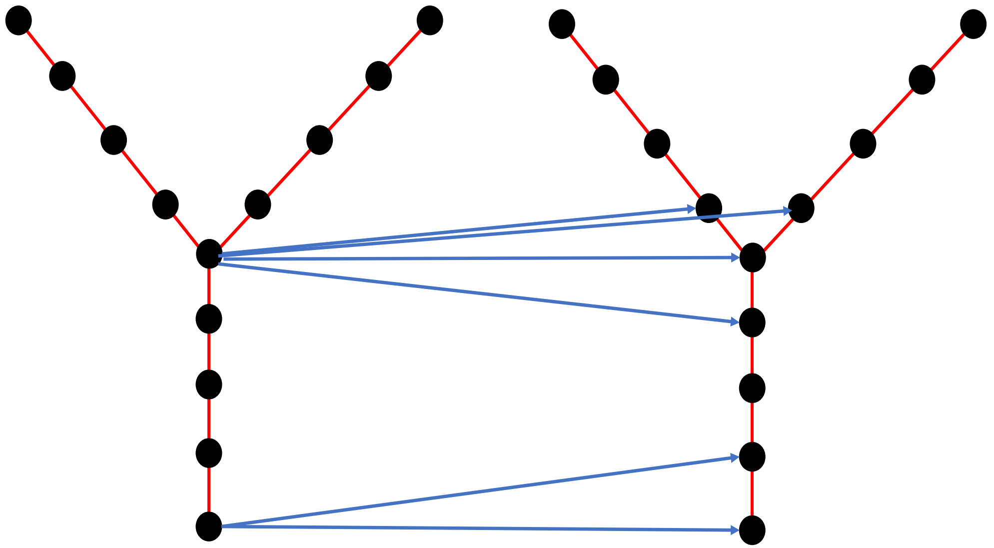
The Mapper Correspondences

- For a data set X, may have two Mapper models Γ_0 and Γ_1
- Can construct a natural correspondence $C(\Gamma_0, \Gamma_1)$ from V_0 to V_1 .
- Vertices correspond to subsets of X
- $v_0 \in \Gamma_0, v_1 \in \Gamma_1$ lie in the correspondence if the corresponding collections have a point in common.

Building Architectures

- When one discovered a model for feature space, such as primary circle or Klein bottle, can use discretized version as finite metric space
- Create metric correspondences, perhaps with products with complete correspondences
- If start with primary circle, and restrict in this fashion, should encourage learning second order information in Klein bottle
- Second order directional derivatives, not angular derivatives

The Mapper Architectures



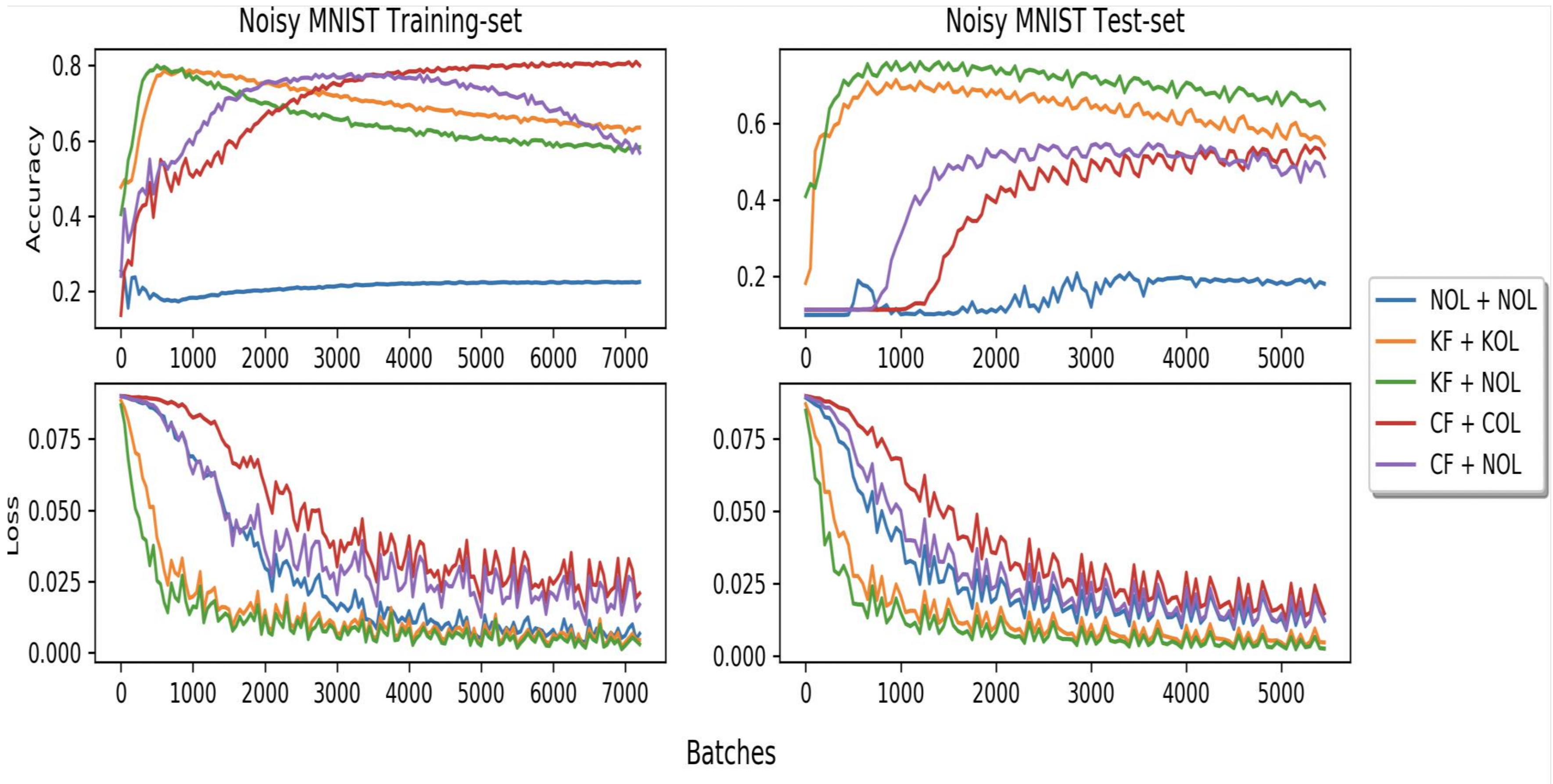
The Mapper Architectures

- $\Gamma_0, \Gamma_1, \dots, \Gamma_n$ a collection of Mapper models with same projections, but different resolution. Decreasing resolution with n.
- Get a feed forward system ($\Gamma_0, \Gamma_1, \dots, \Gamma_n$, $C(\Gamma_0, \Gamma_1)$, …, $C(\Gamma_{n-1}, \Gamma_n)$)
- Uses locality in the feature space
- Incorporates pooling behavior
- Bespoke for individual data sets

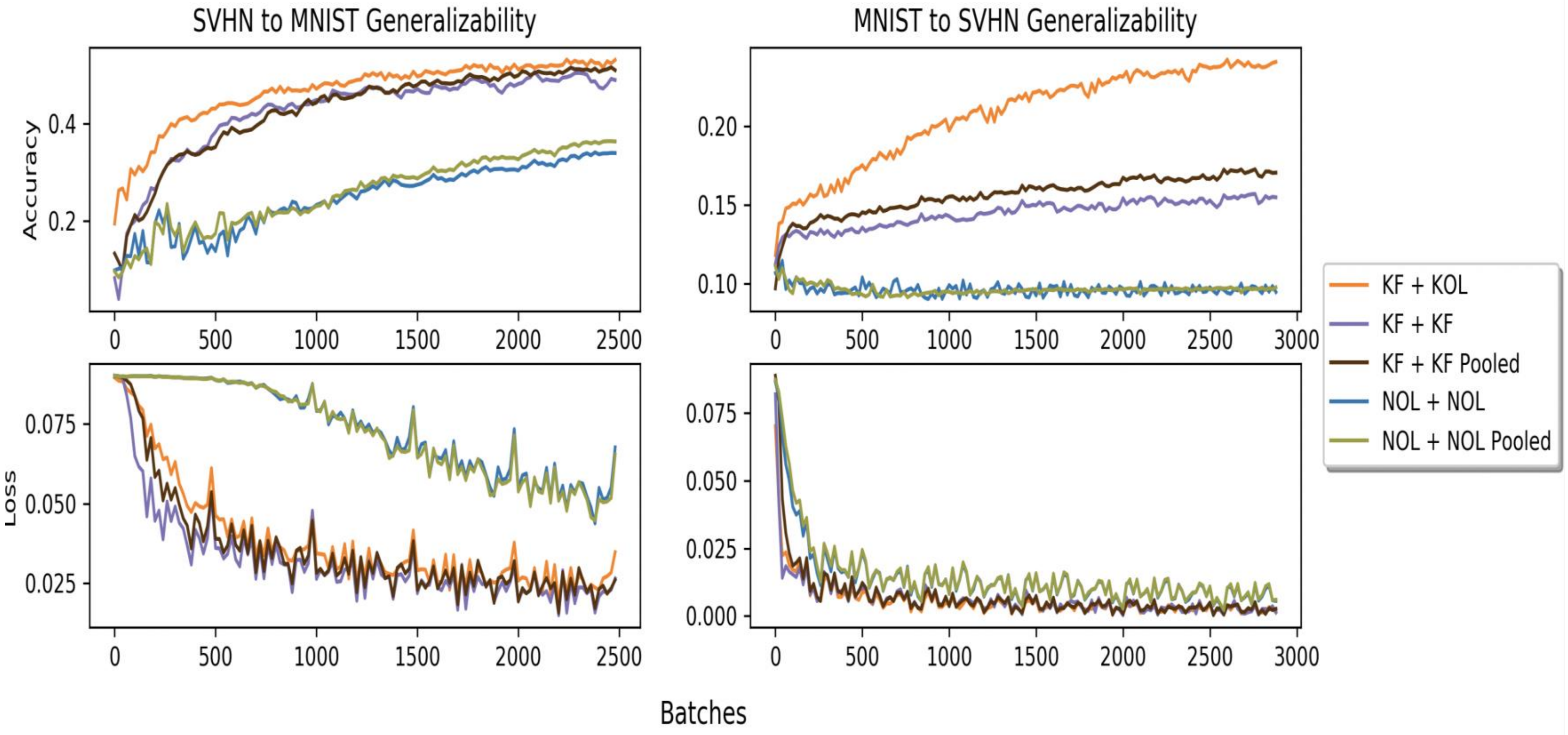
Experiments with Topological Neural Nets

Joint work with Ephy Love and Vasileos Maroulas

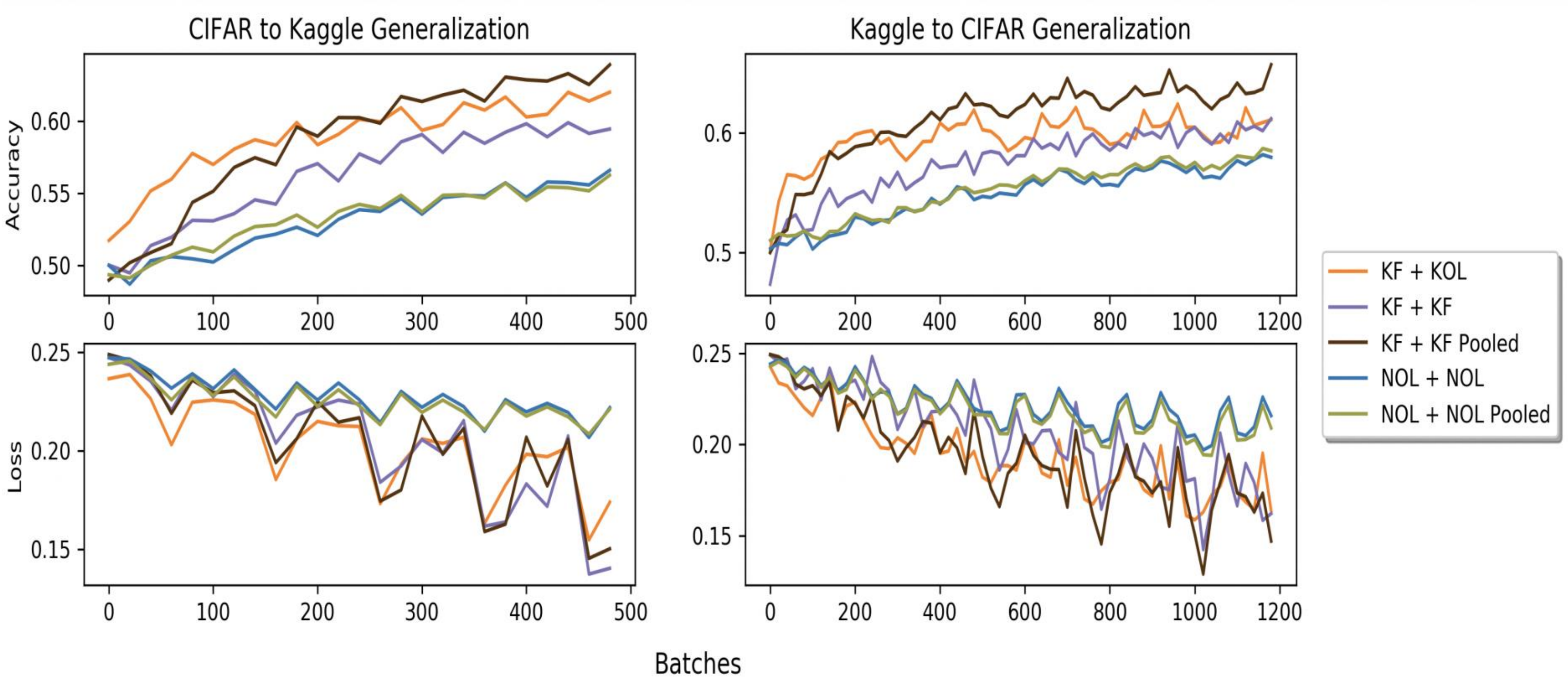
Generalization



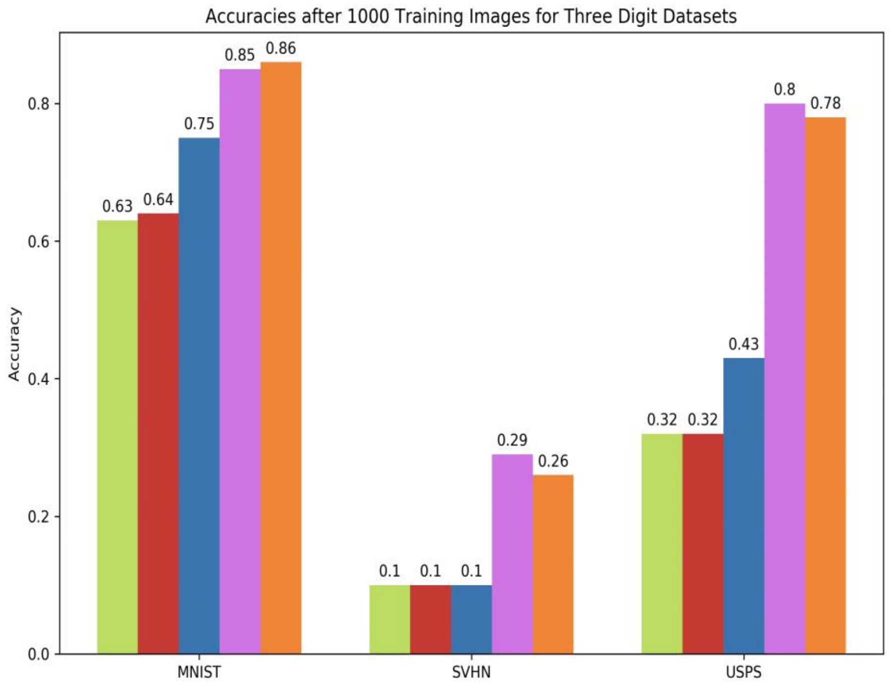
Generalization

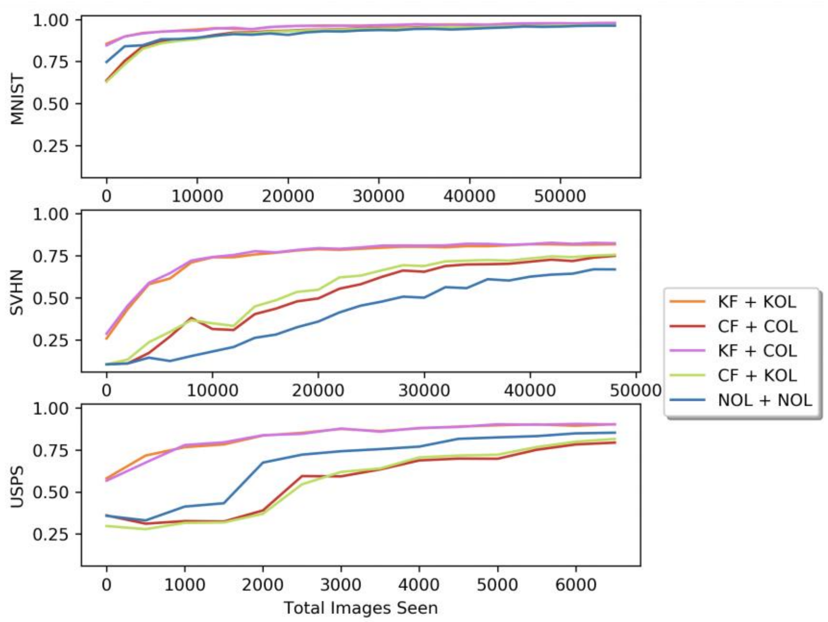


Generalization



Learning


Accuracies after 1000 Training Images for Three Digit Datasets
Accuracy
0.0
0.2
0.4
0.6
0.8
0.63
0.64
0.75
0.85
0.86
0.1
0.1
0.1
0.29
0.26
0.32
0.32
0.43
0.8
0.78
MNIST
SVHN
USPS


MNIST
SVHN
USPS
1.00
0.75
0.50
0.25
0
10000
20000
30000
40000
50000
1000
2000
3000
4000
5000
6000
Total Images Seen
KF + KOL
CF + COL
KF + COL
CF + KOL
NOL + NOL

Thank You