Unsupervised Learning & Transfer Learning

http://nlp.cs.berkeley.edu/comics.shtml

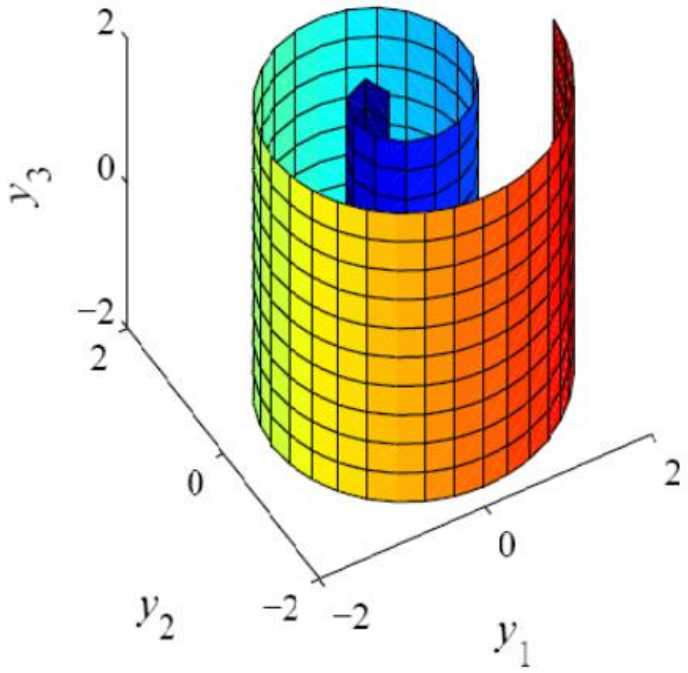
Slides from Simon Carbonelle, Luke Metz

SML310: Research Projects in Data Science, Fall 2019

Michael Guerzhoy

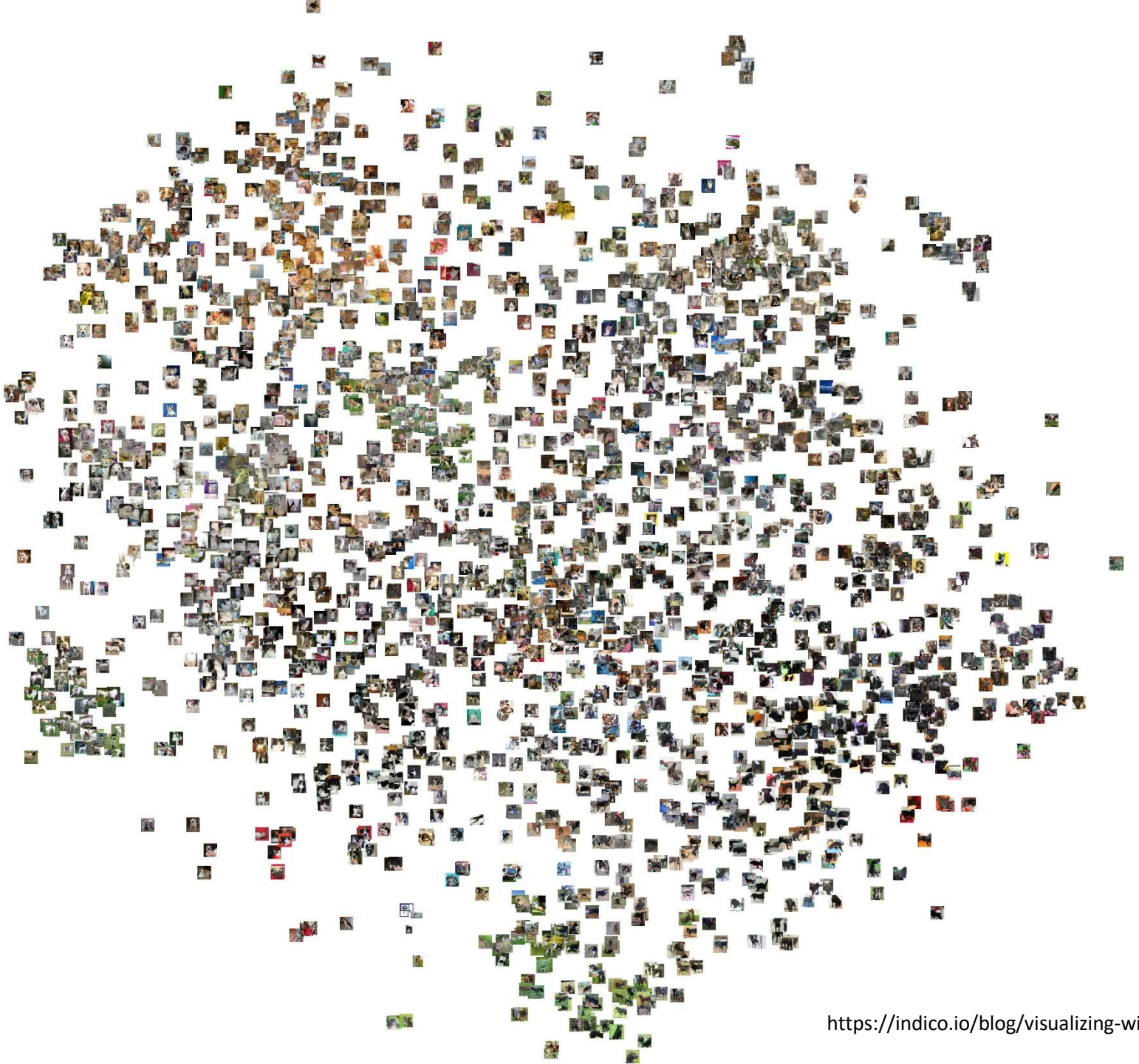
Dimensionality Reduction

- Data in n-dimensional space often lie along manifolds
  - E.g., a “warped” 2-dimensional plane
- Can use this to visualize high-dimensional data
  - Easy to plot stuff in 2D
  - Just un-warp the 2D plane and display it on the screen!



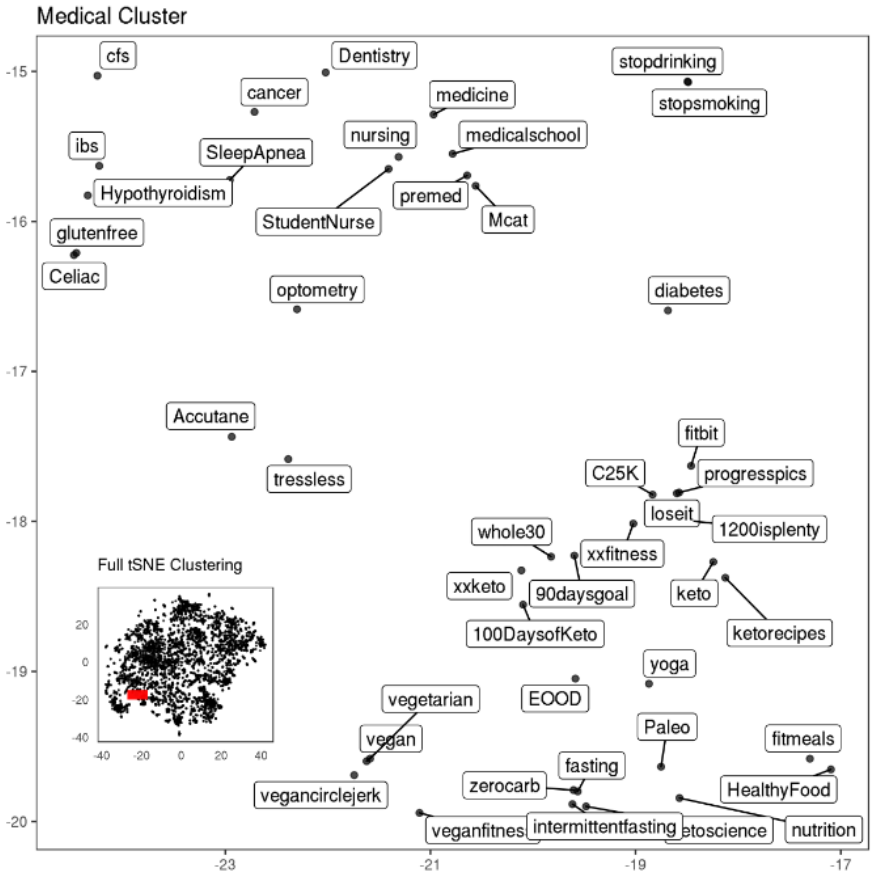
t-SNE

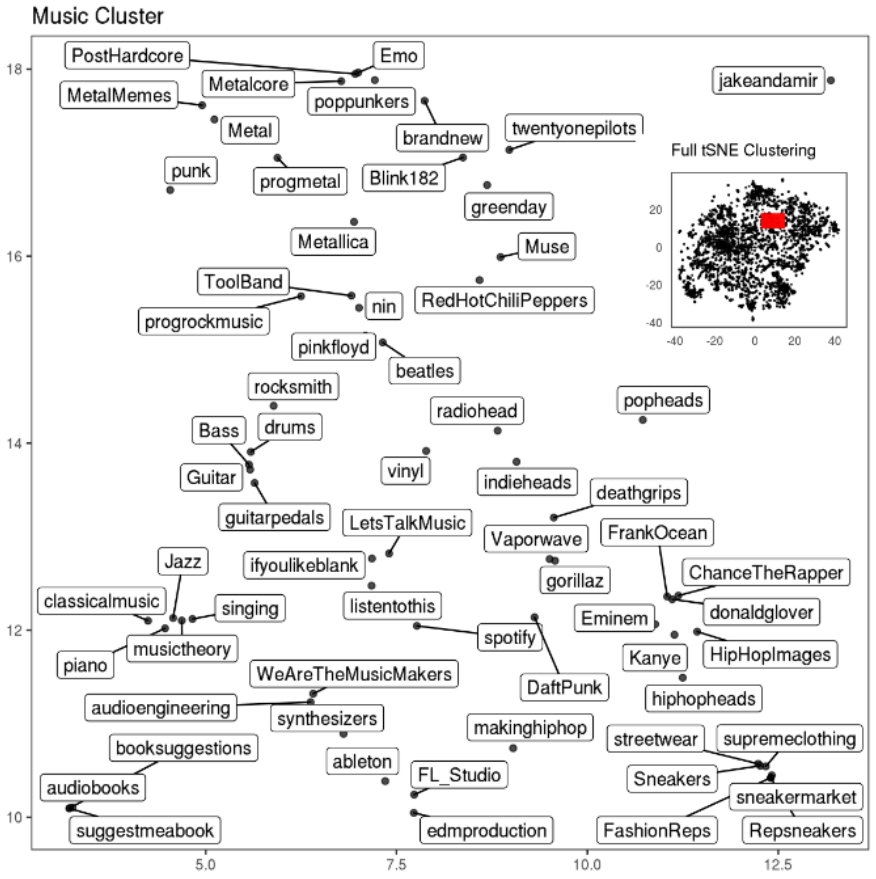
- Original dataset: $x^{(i)}$, with n-dimensional $x^{(i)}$
- Transformed dataset: $z^{(i)}$, with 2-dimensional $z^{(i)}$
- Objective function: if $|x^{(i)} - x^{(j)}|$ is small, $\left|z^{(i)} - z^{(j)}\right|$ should be (usually) small as well



https://indico.io/blog/visualizing-with-t-sne/

Internet communities





T. Martin, community2vec: Vector representations of online communities encode semantic relationships, Proceedings of the Second Workshop on Natural Language Processing and Computational Social Science

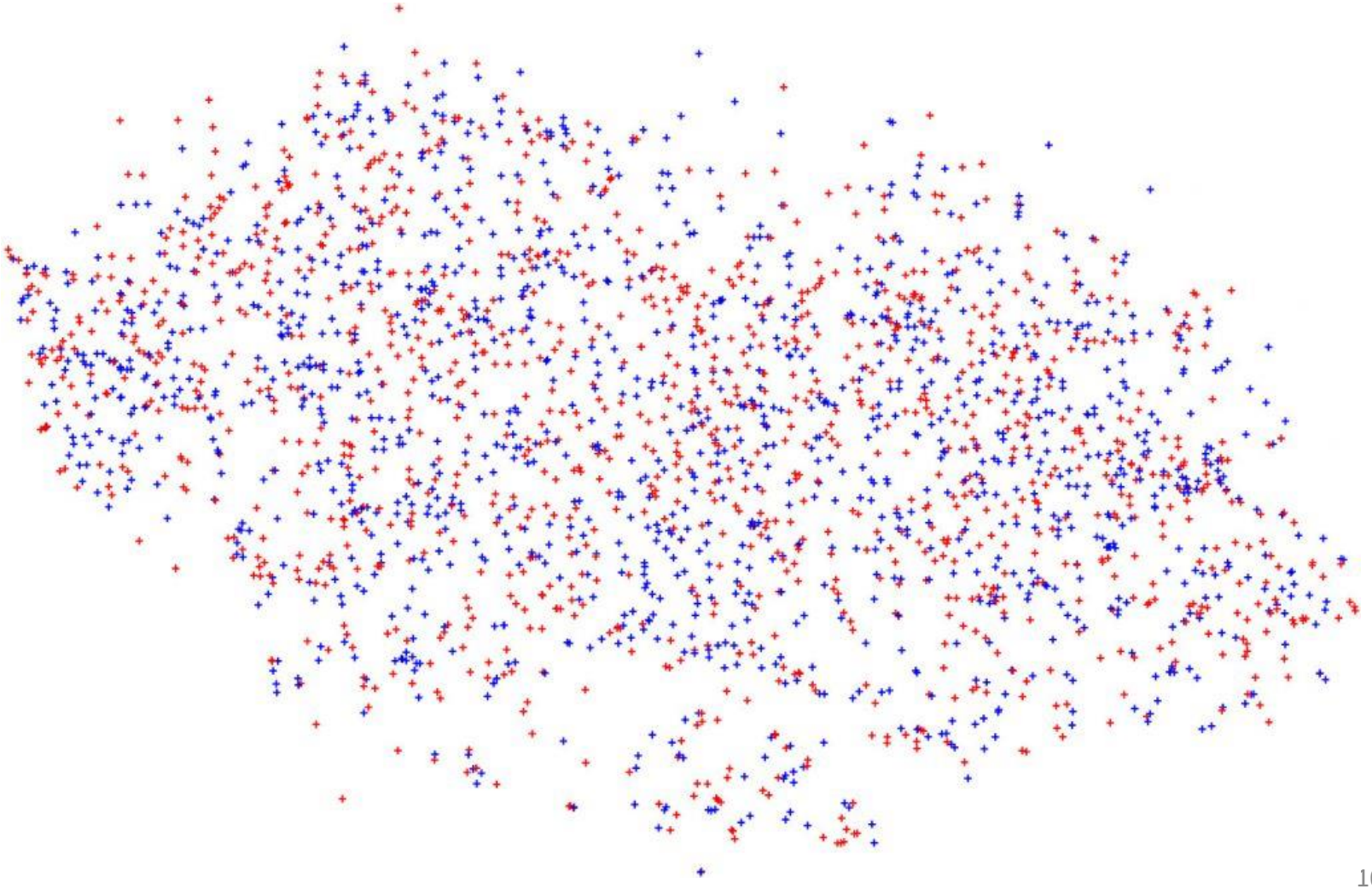
Applications of t-SNE

- Images
- Text
- Genetic data
  - SNP

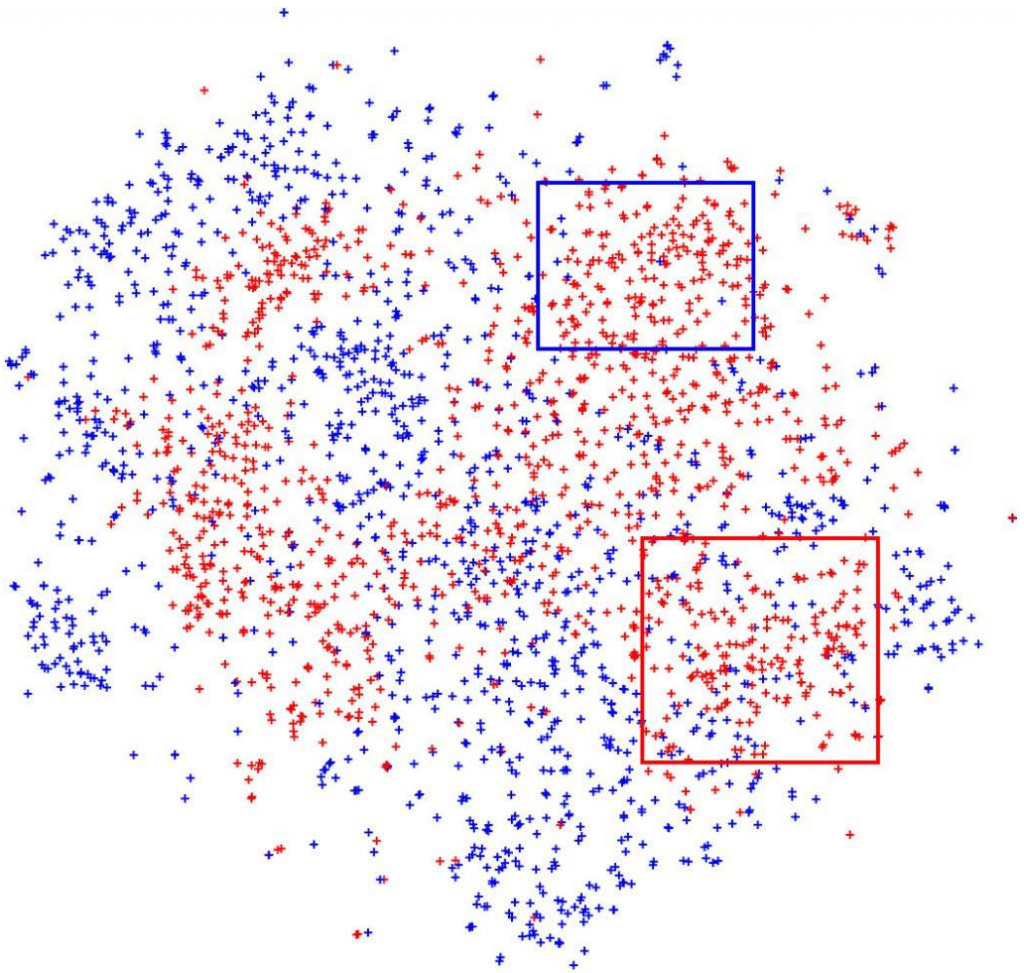
Key Assumption Behind t-SNE

- If $|x^{(i)} - x^{(j)}|$ is small, then the i-th and j-th datapoints are probably similar

t-SNE on raw images of cats and dogs



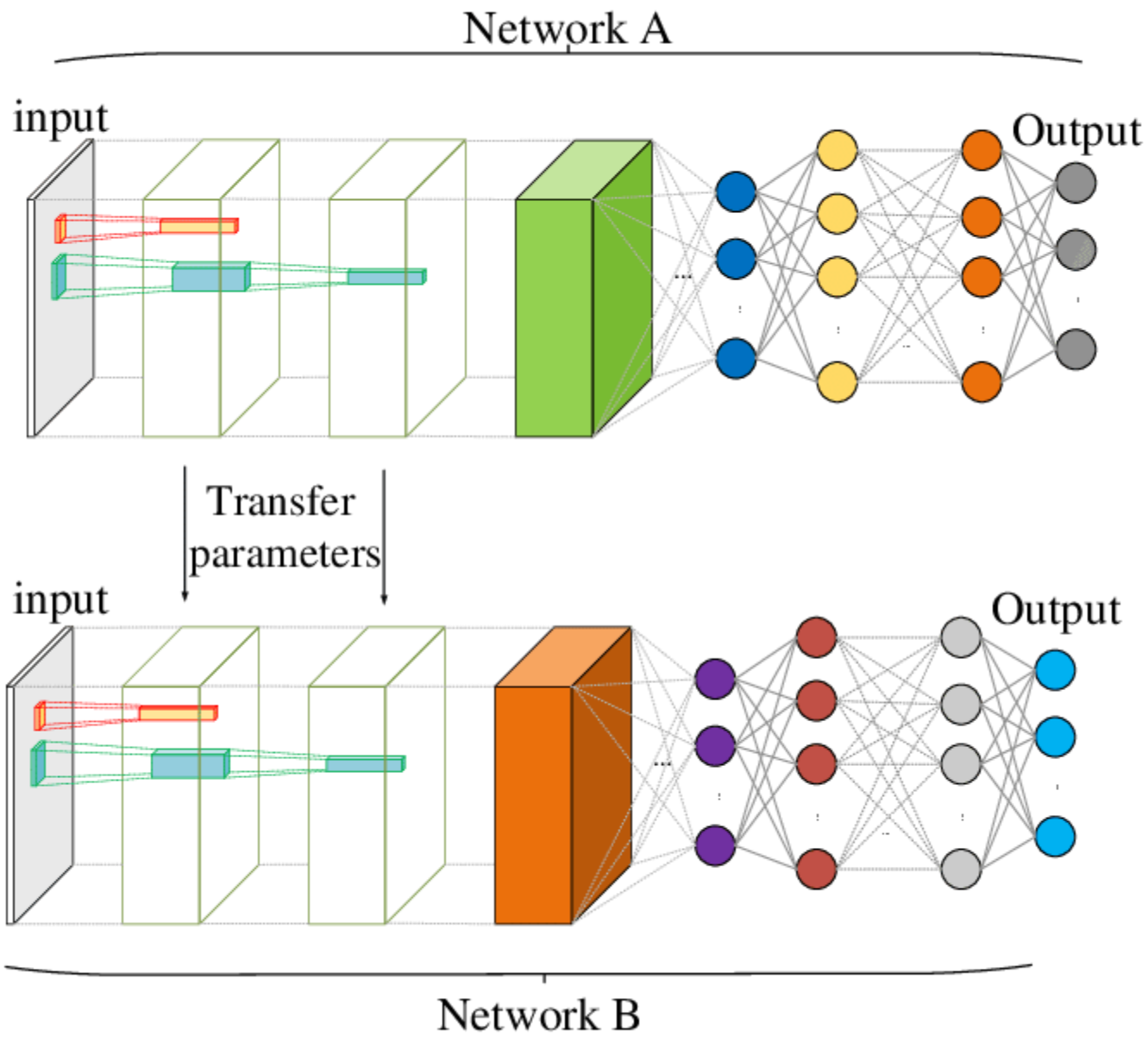
t-SNE on the activations of deep layer of a neural network, for cats and dogs



What we saw before!

Transfer Learning

- Idea: train a large network on a large amount of labelled data
  - Learn to compute useful network activations deep in the network
- Take a small dataset, and use the same network to obtain activations/embeddings
- Train just the top layer(s)
  - Or train the entire network, but starting with the ready-made network


Network A
input
Output
Transfer
parameters
input
Output
Network B

Why are activations more useful than raw pixels?

Why are activations more useful than raw pixels?

- An activation in a layer high-up in a neural networks indicates something like “there is a dog’s ear in the top-right corner of the image”
- A raw pixel value in the image indicates “the pixels at (150, 200) is gray”
- Easier to compare images if you have high-level semantic information
  - True for words as well – that’s why word embeddings are useful

Recap

- For multi-dimensional data, the key is often to obtain a useful representation/embedding
  - Once the embedding is obtained, it can be visualized and/or used for classification
- Unsupervised learning can be used to obtain representations
- Data can be
  - *Clustered:* each data point is assigned to a cluster of points similar to it
  - *Embedded:* each data point is mapped to a point in a new space, and similar points (semantically) are mapped to nearby points in the space into which we are embedding the data