Deep Neural Network Topology

Jee Dong Jun

10.11.2020

Blackbox

- Change input and observe output

Figure 1: From wikipedia

Neural Network

- Neural Network is called blackbox for several reasons.
- We have no idea how neural network is doing prediction.
- We cannot interpret weight of neural network.
- Neural network is not identifiable.

Open Blackbox

- Instead of observing how a single image change through layers, observe whole data set.

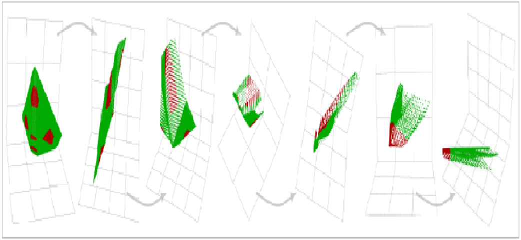


Figure 2: From Topology of Deep Neural Networks

Track manifold M

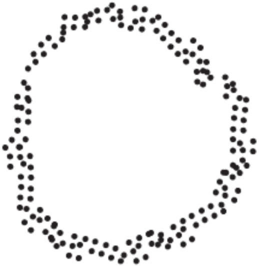
- Consider supervised binary classification problem with $M = M_a \cup M_b$
- Suppose the neural network is fully trained with very low generalization error.
- Observe how topology of M changes within output of each layer.
- Think of kernel method.

Two Problems

- How to quantify topology of manifold?
- Use Betti numbers that capture the shape of a space.
- For $M \in \mathbb{R}^d$, Betti numbers $\beta_k(M)$ represents the number of k-dimensional holes in M.
- For d-dimensional space $k = 0, 1, ..., d - 1$ and β_0 represents the number of connected components.
- This quantity is topologically invariant and intuitively, higher the Betti numbers, the more complex its topology.

Two Problems

- We do not have whole manifold.
- Instead only have finite number of points sampled from manifold with some noises.
- We need to estimate $\beta(M)$ from a point cloud data set.
- Betti numbers of M are estimated by constructing a simplicial complex from a point cloud data set.



Methodology

- Generate Data sets from topologies known in advance.
- Supervised train neural network.
- Track the topology of the respective point cloud data set.
- Change neural network structure and repeat.

Results of Experiment

- All Betti numbers decrease across the layers.
- Using ReLU layers, result in most rapid decreases in all Betti numbers compared to tanh layes.
- As depth of neural network decreases, most of change in topology occurs at final layers.
- Overall, fully trained neural network makes prediction by reducing complex manifold of dataset to simple structure.

Three Questions Asked?

- Why non smooth activation works better than smooth activation?
- Homeomorphic map preserves topology so tanh cannot work well compared to ReLU.
- Universal Approximation Theorem but why deep network is trained well?
- For complex dataset, more layers are required to make topological changes.

How to characterize shape

Figure 4: From Topological Pattern Recognition for Point Cloud Data

- Characterize spaces by pattern occurrence.
- Intuitively, there is one loop.
- But there are infinitely many loops.
- Consider equivalence class.

Homotopy

- Intuitively, loop can be defined by f :$S^1 \to X$ continuous.
- However, counting occurrence (f) is not possible.
- We use homotopy as equivalence relation.
- Two continuous map f,g :$X \to Y$ are homotopic if there exists continuous H : $X \times [0, 1] \to Y$ such that $H(x, 0) = f(x)$ and $H(x, 1) = g(x)$ for all x.
- Based maps are maps with additional constraint such that $f(x_0) = y_0$.
- $\pi_n(Y, y_0)$ is a group of equivalence class of f from $S^n \to Y$

Homology

- Homotopy groups are easy to define but difficult to calculate
- On the other hand, homology is easy to calculate but difficult to define.
- Homology is also topologically invariant.
- It is initially defined for simplicial complexes.

Simplicial Complexes

- k-simplex σ is the convex hull of k+1 affinely independent points $v_0, v_1, ..., v_k \in \mathbb{R}^d$
- k-simplex is represented by its vertices and denoted by $\sigma = [v_0, ..., v_k]$
- Faces of a k-simplex σ are simplices of dimensions 0 to k-1 formed by subset of vertex of σ.

Geometrical Simplicial Complex K

- m-dimensional simplicial complex K is a finite collection of simplicies with dimension at most m.
- Any intersection between two simplicies in K is a face of both of them.
- Include all faces of all its simplicies.
- Geometrical realization of abstract simplicial complex.

Abstract Simplicial Complex

- For every geometrical simplicial complex K, there is abstract simplicial complex behind it.
- List of simplicies $K = \sigma_1, \sigma_2 ..., \sigma_n$ with the property that if $\tau \subset \sigma \in K$ then $\tau \in K$
- Note given abstract simplicial complex, it might not be geometrical simplicial complex.
- We may need to embed vertex set to higher dimension to ensure vertex are affinely independent.
- Called geometrical realization $|K|$

Example of 3d simplicial complex

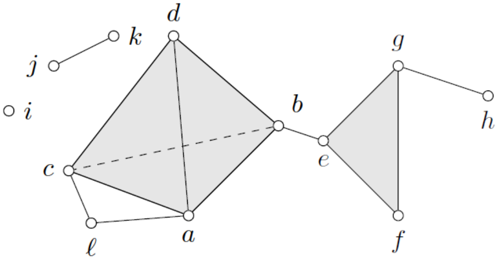


Figure 5: From Topology of Deep Neural Network

Geometrical Representation |K|

- Any simplicial complex is homemorphic to geometric realization of abstract simplicial complex.
- A map of abstract simplicial complexes f from X to Y is map such that $f : V(X) \to V(Y)$ such that if $\sigma \in \Sigma(X)$ then $f_v(\sigma) \in \Sigma(Y)$
- Geometric realization is functorial in a sense that if f,g are abstract complexes map, it induces map $|f| : |X| \to |Y|$ such that $|f \circ g| = |f| \circ |g|$

Definition of the Betti number

- For simplicity restrict field to the field of two elements.
- Let $C_0, C_1, ..., C_d$ be vector spaces over finite field. Also $C_{d+1}, C_{-1} = 0$
- Let $\partial_k : C_k \to C_{k-1}$ be linear maps called boundary such that $\partial_k \circ \partial_{k+1} = 0$.

$$0 \xrightarrow{\partial_{d+1}} C_d \xrightarrow{\partial_d} C_{d-1} \xrightarrow{\partial_{d-1}} \cdots \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} C_{k-1} \xrightarrow{\partial_{k-1}} \cdots \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0,$$

- Such sequence is called chain complex.
- Image of ∂_k is called boundaries and kernel of ∂_{k-1} is called cycles.

Definition of the Betti number

- kernel and image of linear map are subspace.

$$B_k := im(\partial_{k+1}) \subset ker(\partial_k) =: Z_k$$

- We can form the quotient vector space $H_k = Z_k / B_k$ where we treat B_k as zero vector.
- Equivalence Class is defined as $[z] = z + B_k$
- Note vector space is also abelian group with addition. H_k as group is called kth homology group.
- As vector space consider $\dim(H_k)$ called β_k
- If a chain complex considered is associated with a simplical complex K , β_k is called the kth Betti number of K.

Connected Components of simplicial complex K

- Consider vertex set of K $K^{(0)}$
- Define equivalence relation of vertex by $v \sim v'$ if $(v, v') \in K$
- Then, number of connected components equal number of equivalence class.

Free Vector Space V(S)

- When S is finite set, we want to form vector space with S as basis.
- We consider a formal linear combination (it is not actual linear combination)
- If $S = \{v_0, v_1, ..., v_n\}$, then $V(S) = \Sigma_{j=0}^{n} n_j v_j$ where $n_j = 0$ or 1
- Can think of it as one hot encoding φ_{v_k}.
- Dimension of V(S) is number of elements in S

Free vector space of relation

- Given binary relationship, R on X, we can define $V(R) \subset V(X)$
- V(R) is defined as linear span of $\{\varphi_{v_j} - \varphi_{v_i} | (v_i, v_j) \in R\}$
- Note as field considered as the field of two elements, $\varphi_{v_j} - \varphi_{v_i} = \varphi_{v_j} + \varphi_{v_i}$
- There is an isomorphism between $V(X)/V(R) \cong V(X/R)$

An Example

$$\partial_1 = \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{cccc} AB & AC & BC & DE \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

- Note each column describes V(R) where $R = \{(v_i, v_j)|(v_i, v_j) \in K\}$
- $V(K)/V(R) = V(K)/col(\partial_1) \cong V(K/R)$
- Note ∂_1 maps span of 1-simplicies in K to 0-simplicies in K.
- By above explanation, β_0 = number of connected component = number of equivalence class = dimension of K/R = dimension of $V(K)/col(\partial_1)$ = dimension of $ker(\partial_0)/col(\partial_1)$

How to form a chain complex for a simplicial complex

- Given an abstract simplicial complex K, define vector space $C_k(K)$ as a formal linear combination of $K^{(k)}$ be the set of all k-dimensional simplicies in K.
- Note we can define $\partial_k : C_k(K) \to C_{k-1}(K)$ on a k-simplex σ by

  $$\partial_k(\sigma) := \sum_{j=0}^{k} \sigma^{(j)}$$

  where $\sigma^(j)$ is k-1 simplex obtained by removing one vertex from σ
- Note $\delta_k \circ \delta_{k+1} = 0$ as each vertex appears twice so sum becomes 0. (we are still in finite field)
- Now the kth Betti number can be found by $nullitiy(\partial_k) - rank(\partial_{k+1})$
- As each ∂_k can be expressed as matrices, we can calculate them.

How to form a simplicial complex from a point cloud data set

- A point cloud data set is a finite set of n points X.
- After some noise reduction steps, we can form the simplicial complex whose vertex set is X
- One commonly used method is Vietoris-Rips Complex.
- Suppose a measure δ is given and scale ϵ

Vietoris-Rips Complex

- Vietoris-Rips complex at scale ϵ is defined as

$$VR_{\epsilon}(X) := \{[x_0, ..., x_k]\delta(x_i, x_j) \leq 2\epsilon\}$$

- Note this complex depends on ϵ and δ
- There is a theorem that for a dense enough sample, and at sufficiently small scale, the topology recovers the true topology.

Persistent Barcode

- At scale of 0, $VR_0(X)$ is a collection of 0-dimension simplicies with each points in X being a 0-dimension simplicies.
- On the other hand, if ϵ goes to infinity, we get only one big simplex.
- The true answer lies somewhere in between.
- Persistence barcodes provide a summary of the evolution of topology across all scales.
- As ϵ increases, $VR(X, \epsilon_j)$ forms the chain of a nested simplicial complexes.
- Generally speaing, a persistence barcode is tracking of when new homology group appears and dies.