Universal Invariant and Equivariant Graph Neural Networks

Matthew Regehr

University of Waterloo, Waterloo, ON, Canada

matt.regehr@uwaterloo.ca

This presentation is a review of Keriven & Peyré (2019).

February 26, 2024

Overview

1. Introduction
2. Background
3. Results
4. Future Work
5. References

Contents

1. Introduction
2. Background
3. Results
4. Future Work
5. References

The Problem and its Importance

- Equivalence and invariance constraints induce *parameter sharing* constraints in neural networks.

The Problem and its Importance

- Equivariance and invariance constraints induce *parameter sharing* constraints in neural networks.
- This leads to more efficient learning.

The Problem and its Importance

- Equivariance and invariance constraints induce *parameter sharing* constraints in neural networks.
- This leads to more efficient learning.
- Since permutation of nodes doesn't affect graph structure, graph neural networks should be invariant or equivariant.

The Problem and its Importance

- Equivariance and invariance constraints induce *parameter sharing* constraints in neural networks.
- This leads to more efficient learning.
- Since permutation of nodes doesn't affect graph structure, graph neural networks should be invariant or equivariant.
- Classical universal approximation theorem (Cybenko 1989, Hornik et al. 1989, Pinkus 1999) states that a single layer MLP can approximate any continuous function on a compact set.

The Problem and its Importance

- Equivariance and invariance constraints induce *parameter sharing* constraints in neural networks.
- This leads to more efficient learning.
- Since permutation of nodes doesn't affect graph structure, graph neural networks should be invariant or equivariant.
- Classical universal approximation theorem (Cybenko 1989, Hornik et al. 1989, Pinkus 1999) states that a single layer MLP can approximate any continuous function on a compact set.
- Universality is important because it ensures that a model does not have "gaps" in its learning capabilities.

The Problem and its Importance

- Equivariance and invariance constraints induce *parameter sharing* constraints in neural networks.
- This leads to more efficient learning.
- Since permutation of nodes doesn't affect graph structure, graph neural networks should be invariant or equivariant.
- Classical universal approximation theorem (Cybenko 1989, Hornik et al. 1989, Pinkus 1999) states that a single layer MLP can approximate any continuous function on a compact set.
- Universality is important because it ensures that a model does not have "gaps" in its learning capabilities.
- Are GNNs still universal under equivariance or invariance?

Prior Work

- Maron et al. (2019) gives a positive result for the invariant case.

Prior Work

- Maron et al. (2019) gives a positive result for the invariant case.
- Theoretical attention in general has focused on the invariant setting (Yarotsky 2022) and special cases such as sets (Zaheer et al. 2017).

Prior Work

- Maron et al. (2019) gives a positive result for the invariant case.
- Theoretical attention in general has focused on the invariant setting (Yarotsky 2022) and special cases such as sets (Zaheer et al. 2017).
- Work on equivariance has mainly been limited to translation as it relates to convolution (Cohen & Welling 2016).

Prior Work

- Maron et al. (2019) gives a positive result for the invariant case.
- Theoretical attention in general has focused on the invariant setting (Yarotsky 2022) and special cases such as sets (Zaheer et al. 2017).
- Work on equivariance has mainly been limited to translation as it relates to convolution (Cohen & Welling 2016).
- Many applications demand equivariance, e.g., community detection (Chen et al. 2017), recommender systems (Ying et al. 2018).

Contents

1. Introduction
2. Background
3. Results
4. Future Work
5. References

Universality

Definition

Given an activation function $\rho : \mathbb{R} \to \mathbb{R}$, a one-layer MLP is a neural network of the form

$$f : \mathbb{R}^n \to \mathbb{R}, x \mapsto \sum_s c_s \rho(\langle a_s, x \rangle + b_s).$$

The set of such MLPs is denoted $N_{\mathrm{MLP}}(\rho)$.

Universality

Definition

Given an activation function $\rho : \mathbb{R} \to \mathbb{R}$, a one-layer MLP is a neural network of the form

$$f : \mathbb{R}^n \to \mathbb{R}, x \mapsto \sum_s c_s \rho(\langle a_s, x\rangle + b_s).$$

The set of such MLPs is denoted $N_{\mathrm{MLP}}(\rho)$.

Theorem (The Universal Approximation Theorem (Pinkus 1999))

For any $\rho \in C[a, b]$ that is not a polynomial, $N_{MLP}(\rho)$ is a dense subspace of $C[a, b]$.

Equivariance/Invariance

Definition

Given a hypergraph $G \in \mathbb{R}^{n^d}$ represented as a tensor and a permutation $\pi \in S_n$, denote by

$$(\pi \cdot G)_{i_1,\ldots,i_d} := G_{\pi^{-1}(i_1),\ldots,\pi^{-1}(i_d)}$$

the group action of π on G.

Equivariance/Invariance

Definition

Given a hypergraph $G \in \mathbb{R}^{n^d}$ represented as a tensor and a permutation $\pi \in S_n$, denote by

$$(\pi \cdot G)_{i_1,\dots,i_d} := G_{\pi^{-1}(i_1),\dots,\pi^{-1}(i_d)}$$

the group action of π on G.

Definition

A map $f : \mathbb{R}^{n^k} \to \mathbb{R}$ is called invariant if $f(\pi \cdot G) = f(G)$ for all π and G.

Equivalence/Invariance

Definition

Given a hypergraph $G \in \mathbb{R}^{n^d}$ represented as a tensor and a permutation $\pi \in S_n$, denote by

$$(\pi \cdot G)_{i_1,\ldots,i_d} := G_{\pi^{-1}(i_1),\ldots,\pi^{-1}(i_d)}$$

the group action of π on G.

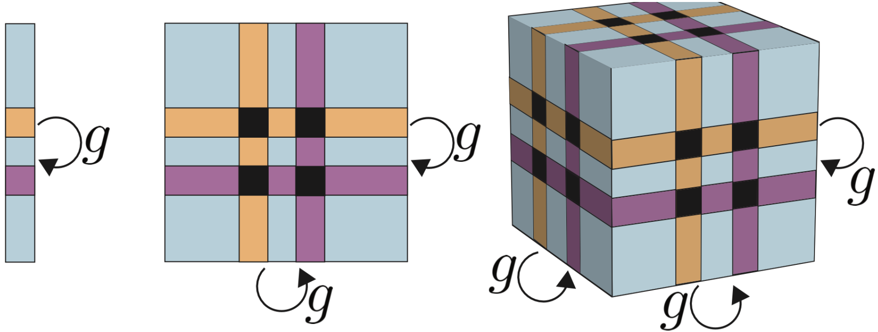
Definition

A map $f : \mathbb{R}^{n^k} \to \mathbb{R}$ is called invariant if $f(\pi \cdot G) = f(G)$ for all π and G.

Definition

A map $f : \mathbb{R}^{n^k} \to \mathbb{R}^{n^\ell}$ is called equivariant if $f(\pi \cdot G) = \pi \cdot f(G)$ for all π and G.

Equivariant and Invariant GNNs



(Maron et al. 2019)

Equivariant and Invariant GNNs

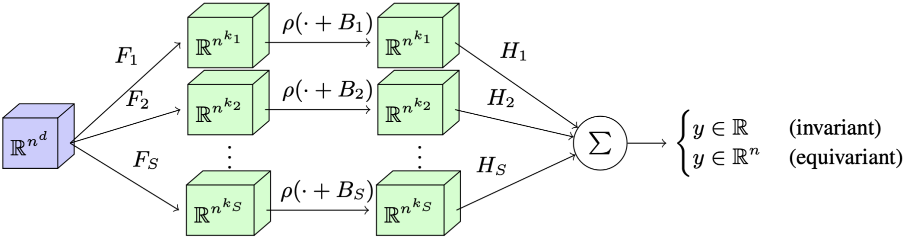
Definition

A one-layer invariant (equivariant) GNN with activation function $\rho : \mathbb{R} \to \mathbb{R}$ is

$$f : \mathbb{R}^{n^d} \to \mathbb{R}(\mathbb{R}^n), G \mapsto \sum_s H_s(\rho(F_s G + B_s)) + b$$

where each $F_s : \mathbb{R}^{n^d} \to \mathbb{R}^{n^{k_s}}$ is a linear equivariance, each $H_s : \mathbb{R}^{n^{k_s}} \to \mathbb{R}(\mathbb{R}^n)$ is an linear invariance (equivariance), and each bias term satisfies the equivariance relation $B_s = \pi \cdot B_s$. The set of these invariant networks is denoted $N_{\mathrm{inv}}(\rho)$ and the equivariant networks are denoted $N_{\mathrm{eq}}(\rho)$.

Equivariant and Invariant GNNs



(Keriven & Peyré 2019)

Contents

1 Introduction

2 Background

3 Results

4 Future Work

5 References

Universality of Invariant GNNs

First contribution is a new real-analytic proof of Theorem 1 from Maron et al. (2019).

Universality of Invariant GNNs

First contribution is a new real-analytic proof of Theorem 1 from Maron et al. (2019).

Theorem

Let $\mathcal{G}_{N,R} := \{G \in \mathbb{R}^{n^d} : n \leq N, ||G|| \leq R\}$ and let $C_{inv}(\mathcal{G}_{N,R})$ denote the set of continuous and invariant functions $f : \mathcal{G}_{N,R} \to \mathbb{R}$. Then, for any locally Lipschitz non-polynomial ρ, $N_{inv}(\rho)$ is dense in $C_{inv}(\mathcal{G}_{N,R})$.

Stone–Weierstraß

The new proof works by reduction to the the following generalized Stone–Weierstraß theorem.

Stone–Weierstraß

The new proof works by reduction to the the following generalized Stone–Weierstraß theorem.

Theorem (Stone–Weierstraß)

Let $\mathcal{X}$ be a compact metric space and let $\mathcal{F}$ be a subalgebra of $C(\mathcal{X})$ that contains at least one non-zero constant function. If $\mathcal{F}$ separates points, i.e. $\forall x \neq y$ there is $f \in \mathcal{F}$ s.t. $f(x) \neq f(y)$, then $\mathcal{F}$ is dense in $C(\mathcal{X})$.

Stone–Weierstraß

The new proof works by reduction to the the following generalized Stone–Weierstraß theorem.

Theorem (Stone–Weierstraß)

Let $\mathcal{X}$ be a compact metric space and let $\mathcal{F}$ be a subalgebra of $C(\mathcal{X})$ that contains at least one non-zero constant function. If $\mathcal{F}$ separates points, i.e. $\forall x \neq y$ there is $f \in \mathcal{F}$ s.t. $f(x) \neq f(y)$, then $\mathcal{F}$ is dense in $C(\mathcal{X})$.

Note: A subalgebra needs to be closed under pointwise multiplication, addition, as well as scalar multiplication.

How to Incorporate Invariance?

- Applying Stone–Weierstraß directly has several problems.

How to Incorporate Invariance?

- Applying Stone–Weierstraß directly has several problems.
- First, $C_{\mathrm{inv}}(\mathcal{G}_{N,R}) \neq C(\mathcal{G}_{N,R})$.

How to Incorporate Invariance?

- Applying Stone–Weierstraß directly has several problems.
- First, $C_{\mathrm{inv}}(\mathcal{G}_{N,R}) \neq C(\mathcal{G}_{N,R})$.
- The solution is to replace $\mathcal{G}_{N,R}$ with a quotient space $\hat{\mathcal{G}}_{N,R}$ under permutation equivalence.

How to Incorporate Invariance?

- Applying Stone–Weierstraß directly has several problems.
- First, $C_{\mathrm{inv}}(\mathcal{G}_{N,R}) \neq C(\mathcal{G}_{N,R})$.
- The solution is to replace $\mathcal{G}_{N,R}$ with a quotient space $\hat{\mathcal{G}}_{N,R}$ under permutation equivalence.
- Next, equip the new space with an appropriate metric d_{edit} based on edit distance.

How to Incorporate Invariance?

- Applying Stone–Weierstraß directly has several problems.
- First, $C_{\mathrm{inv}}(\mathcal{G}_{N,R}) \neq C(\mathcal{G}_{N,R})$.
- The solution is to replace $\mathcal{G}_{N,R}$ with a quotient space $\hat{\mathcal{G}}_{N,R}$ under permutation equivalence.
- Next, equip the new space with an appropriate metric d_{edit} based on edit distance.
- Can show $d_{\mathrm{edit}}([G_1],[G_2]) \leq ||G_1 - G_2||_1$, meaning the embedding $G \mapsto [G]$ is continuous and thus no loss of generality working in the new space.

Passing to a Subalgebra

- The next problem is that $N_{\mathrm{inv}}(\rho)$ is not a subalgebra because it is not closed under pointwise multiplication.

Passing to a Subalgebra

- The next problem is that $N_{\mathrm{inv}}(\rho)$ is not a subalgebra because it is not closed under pointwise multiplication.
- The solution is to replace with $N_{\mathrm{inv}}(\rho)$ with its closure under pointwise multiplication $N_{\mathrm{inv}}^{\otimes}(\rho)$.

Passing to a Subalgebra

- The next problem is that $N_{\text{inv}}(\rho)$ is not a subalgebra because it is not closed under pointwise multiplication.
- The solution is to replace with $N_{\text{inv}}(\rho)$ with its closure under pointwise multiplication $N^{\otimes}_{\text{inv}}(\rho)$.
- Once we can show the other conditions for the sigmoid function σ, we can apply the same trick as in Hornik et al. (1989) and argue that

Passing to a Subalgebra

- The next problem is that $N_{\mathrm{inv}}(\rho)$ is not a subalgebra because it is not closed under pointwise multiplication.
- The solution is to replace with $N_{\mathrm{inv}}(\rho)$ with its closure under pointwise multiplication $N_{\mathrm{inv}}^{\otimes}(\rho)$.
- Once we can show the other conditions for the sigmoid function σ, we can apply the same trick as in Hornik et al. (1989) and argue that
  1. $N_{\mathrm{inv}}(\rho)$ is dense in $N_{\mathrm{inv}}(\cos)$ (by the classical universal approximation theorem);

Passing to a Subalgebra

- The next problem is that $N_{\mathrm{inv}}(\rho)$ is not a subalgebra because it is not closed under pointwise multiplication.
- The solution is to replace with $N_{\mathrm{inv}}(\rho)$ with its closure under pointwise multiplication $N_{\mathrm{inv}}^{\otimes}(\rho)$.
- Once we can show the other conditions for the sigmoid function σ, we can apply the same trick as in Hornik et al. (1989) and argue that
  1. $N_{\mathrm{inv}}(\rho)$ is dense in $N_{\mathrm{inv}}(\cos)$ (by the classical universal approximation theorem);
  2. $N_{\mathrm{inv}}(\cos) = N_{\mathrm{inv}}^{\otimes}(\cos)$ (because a product of cos is also a sum of cos); and, finally,

Passing to a Subalgebra

- The next problem is that $N_{\mathrm{inv}}(\rho)$ is not a subalgebra because it is not closed under pointwise multiplication.
- The solution is to replace with $N_{\mathrm{inv}}(\rho)$ with its closure under pointwise multiplication $N_{\mathrm{inv}}^{\otimes}(\rho)$.
- Once we can show the other conditions for the sigmoid function σ, we can apply the same trick as in Hornik et al. (1989) and argue that
  1. $N_{\mathrm{inv}}(\rho)$ is dense in $N_{\mathrm{inv}}(\cos)$ (by the classical universal approximation theorem);
  2. $N_{\mathrm{inv}}(\cos) = N_{\mathrm{inv}}^{\otimes}(\cos)$ (because a product of cos is also a sum of cos); and, finally,
  3. $N_{\mathrm{inv}}^{\otimes}(\cos)$ is dense in $N_{\mathrm{inv}}^{\otimes}(\sigma)$ where (by Fourier decomposition of σ).

Passing to a Subalgebra

- The next problem is that $N_{\mathrm{inv}}(\rho)$ is not a subalgebra because it is not closed under pointwise multiplication.
- The solution is to replace with $N_{\mathrm{inv}}(\rho)$ with its closure under pointwise multiplication $N_{\mathrm{inv}}^{\otimes}(\rho)$.
- Once we can show the other conditions for the sigmoid function σ, we can apply the same trick as in Hornik et al. (1989) and argue that
  1. $N_{\mathrm{inv}}(\rho)$ is dense in $N_{\mathrm{inv}}(\cos)$ (by the classical universal approximation theorem);
  2. $N_{\mathrm{inv}}(\cos) = N_{\mathrm{inv}}^{\otimes}(\cos)$ (because a product of cos is also a sum of cos); and, finally,
  3. $N_{\mathrm{inv}}^{\otimes}(\cos)$ is dense in $N_{\mathrm{inv}}^{\otimes}(\sigma)$ where (by Fourier decomposition of σ).
- That is, if Stone–Weierstraß applies to $N_{\mathrm{inv}}^{\otimes}(\sigma)$, then the theorem is proved for $N_{\mathrm{inv}}(\rho)$.

Separation of Points

- The only remaining challenge is to argue that $N_{\mathrm{inv}}^{\otimes}(\sigma)$ separates points in $\hat{\mathcal{G}}$.

Separation of Points

- The only remaining challenge is to argue that $N_{\text{inv}}^{\otimes}(\sigma)$ separates points in $\hat{\mathcal{G}}$.
- The proof is by contrapositive: Let $G \in \mathbb{R}^{n^d}$ and $G' \in \mathbb{R}^{(n')^d}$ be s.t. $f(G) = f(G')$ for every $f \in N_{\text{inv}}^{\otimes}(\sigma)$. Then $[G] = [G']$.

Separation of Points

- The only remaining challenge is to argue that $N_{\mathrm{inv}}^{\otimes}(\sigma)$ separates points in $\hat{\mathcal{G}}$.
- The proof is by contrapositive: Let $G \in \mathbb{R}^{n^d}$ and $G' \in \mathbb{R}^{(n')^d}$ be s.t. $f(G) = f(G')$ for every $f \in N_{\mathrm{inv}}^{\otimes}(\sigma)$. Then $[G] = [G']$.
- Start by showing that $n = n'$.

Separation of Points

- The only remaining challenge is to argue that $N_{\mathrm{inv}}^{\otimes}(\sigma)$ separates points in $\hat{\mathcal{G}}$.
- The proof is by contrapositive: Let $G \in \mathbb{R}^{n^d}$ and $G' \in \mathbb{R}^{(n')^d}$ be s.t. $f(G) = f(G')$ for every $f \in N_{\mathrm{inv}}^{\otimes}(\sigma)$. Then $[G] = [G']$.
- Start by showing that $n = n'$.
- Let $\tau \in \mathbb{R}$ be strictly smaller than all elements in G and G' and consider

$$f_\lambda(G) = \sum_{i_1,\ldots,i_d} \sigma(\lambda(G - \tau))_{i_1,\ldots,i_d} \in N_{\mathrm{inv}}^{\otimes}(\sigma).$$

Separation of Points

- The only remaining challenge is to argue that $N_{\text{inv}}^{\otimes}(\sigma)$ separates points in $\hat{\mathcal{G}}$.
- The proof is by contrapositive: Let $G \in \mathbb{R}^{n^d}$ and $G' \in \mathbb{R}^{(n')^d}$ be s.t. $f(G) = f(G')$ for every $f \in N_{\text{inv}}^{\otimes}(\sigma)$. Then $[G] = [G']$.
- Start by showing that $n = n'$.
- Let $\tau \in \mathbb{R}$ be strictly smaller than all elements in G and G' and consider

$$f_\lambda(G) = \sum_{i_1,\dots,i_d} \sigma(\lambda(G - \tau))_{i_1,\dots,i_d} \in N_{\text{inv}}^{\otimes}(\sigma).$$

- As $\lambda \to \infty$, each $\sigma(\dots)_{i_1,\dots,i_d} \to 1$, so $f_\lambda(G) \to n$ and $f_\lambda(G') \to n'$.

Separation of Points

- The only remaining challenge is to argue that $N_{\mathrm{inv}}^{\otimes}(\sigma)$ separates points in $\hat{\mathcal{G}}$.
- The proof is by contrapositive: Let $G \in \mathbb{R}^{n^d}$ and $G' \in \mathbb{R}^{(n')^d}$ be s.t. $f(G) = f(G')$ for every $f \in N_{\mathrm{inv}}^{\otimes}(\sigma)$. Then $[G] = [G']$.
- Start by showing that $n = n'$.
- Let $\tau \in \mathbb{R}$ be strictly smaller than all elements in G and G' and consider

$$f_\lambda(G) = \sum_{i_1,\dots,i_d} \sigma(\lambda(G - \tau))_{i_1,\dots,i_d} \in N_{\mathrm{inv}}^{\otimes}(\sigma).$$

- As $\lambda \to \infty$, each $\sigma(\dots)_{i_1,\dots,i_d} \to 1$, so $f_\lambda(G) \to n$ and $f_\lambda(G') \to n'$.
- Remainder of argument is more technically involved but proceeds similarly.

Generalization to Equivariant GNNs

Main contribution is a universality result for equivariant GNNs.

Generalization to Equivariant GNNs

Main contribution is a universality result for equivariant GNNs.

Theorem

Let $C_{eq}(\mathcal{G}_{N,R})$ denote the set of continuous and equivariant functions $f : \mathcal{G}_{N,R} \to \mathbb{R}^n$. Then, for any locally Lipschitz non-polynomial ρ, $N_{eq}(\rho)$ is dense in $C_{eq}(\mathcal{G}_{N,R})$.

Generalization to Equivariant GNNs

Main contribution is a universality result for equivariant GNNs.

Theorem

Let $C_{eq}(\mathcal{G}_{N,R})$ denote the set of continuous and equivariant functions $f : \mathcal{G}_{N,R} \to \mathbb{R}^n$. Then, for any locally Lipschitz non-polynomial ρ, $N_{eq}(\rho)$ is dense in $C_{eq}(\mathcal{G}_{N,R})$.

Ideally, the proof should be analogous to the invariant case.

The Problem with Stone–Weierstraß

- Equivariant functions act differently on members of the same equivalence class.

The Problem with Stone–Weierstraß

- Equivariant functions act differently on members of the same equivalence class.
- Thus unlike in the invariant case, we cannot replace $\mathcal{G}$ by the quotient space $\hat{\mathcal{G}}$.

The Problem with Stone–Weierstraß

- Equivariant functions act differently on members of the same equivalence class.
- Thus unlike in the invariant case, we cannot replace $\mathcal{G}$ by the quotient space $\hat{\mathcal{G}}$.
- Therefore we cannot reduce to Stone–Weierstraß and need a brand new version.

A New Stone–Weierstraß

- Separability condition remains the same but new "self"-separability condition is introduced. Intuitively, the coordinates of $f(G)$ should separate from one another.

A New Stone–Weierstraß

- Separability condition remains the same but new "self"-separability condition is introduced. Intuitively, the coordinates of $f(G)$ should separate from one another.
- Proving self-separability for $N_{\mathrm{eq}}^{\otimes}(\sigma)$ is similar to separation of points.

A New Stone–Weierstraß

- Separability condition remains the same but new "self"-separability condition is introduced. Intuitively, the coordinates of $f(G)$ should separate from one another.
- Proving self-separability for $N_{\mathrm{eq}}^{\otimes}(\sigma)$ is similar to separation of points.
- Reduction $N_{\mathrm{eq}}(\rho) \mapsto N_{\mathrm{eq}}(\cos) = N_{\mathrm{eq}}^{\otimes}(\cos) \mapsto N_{\mathrm{eq}}^{\otimes}(\sigma)$ same as before.

Contents

1. Introduction
2. Background
3. Results
4. Future Work
5. References

Possible Directions

- Universality of invariant networks is proved constructively in Maron et al. (2019), whereas the Stone–Weierstraß approach is non-constructive. Can we generalize to the equivariant case by an explicit constructive method?

Possible Directions

- Universality of invariant networks is proved constructively in Maron et al. (2019), whereas the Stone–Weierstraß approach is non-constructive. Can we generalize to the equivariant case by an explicit constructive method?
- Can we apply the new Stone–Weierstraß tool to more general settings? In particular, can we show universality for equivariant neural networks operating on dynamic graphs where the structure evolves over time?

Contents

1. Introduction
2. Background
3. Results
4. Future Work
5. References

References I

Chen, Z., Li, X. & Bruna, J. (2017), 'Supervised community detection with line graph neural networks', *arXiv preprint arXiv:1705.08415* .

Cohen, T. & Welling, M. (2016), Group equivariant convolutional networks, *in* 'International conference on machine learning', PMLR, pp. 2990–2999.

Cybenko, G. (1989), 'Approximation by superpositions of a sigmoidal function', *Mathematics of control, signals and systems* **2**(4), 303–314.

Hornik, K., Stinchcombe, M. & White, H. (1989), 'Multilayer feedforward networks are universal approximators', *Neural networks* **2**(5), 359–366.

Keriven, N. & Peyré, G. (2019), 'Universal invariant and equivariant graph neural networks', *Advances in Neural Information Processing Systems* **32**.

Maron, H., Fetaya, E., Segol, N. & Lipman, Y. (2019), On the universality of invariant networks, *in* 'International conference on machine learning', PMLR, pp. 4363–4371.

References II

Pinkus, A. (1999), 'Approximation theory of the mlp model in neural networks', *Acta numerica* **8**, 143–195.

Yarotsky, D. (2022), 'Universal approximations of invariant maps by neural networks', *Constructive Approximation* **55**(1), 407–474.

Ying, R., He, R., Chen, K., Eksombatchai, P., Hamilton, W. L. & Leskovec, J. (2018), Graph convolutional neural networks for web-scale recommender systems, *in* 'Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining', pp. 974–983.

Zaheer, M., Kottur, S., Ravanbakhsh, S., Poczos, B., Salakhutdinov, R. R. & Smola, A. J. (2017), 'Deep sets', *Advances in neural information processing systems* **30**.

Thank You

Thank you :)