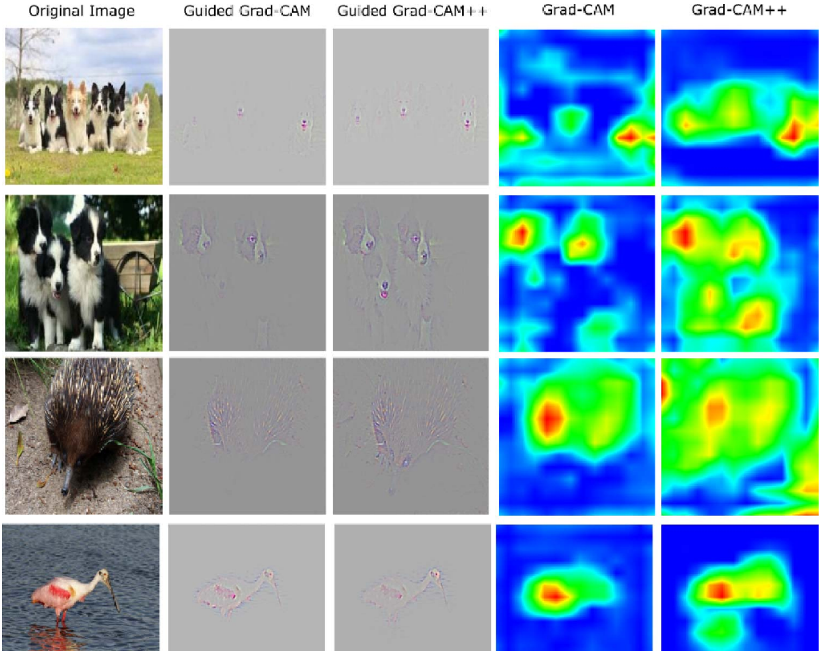
Formal Concept Analysis, Lattice Polinomials and Approximation

Levan Tsinadze

Tbilisi State University

2022

Why Formal Concept Analysis



Aggregation of Common Features

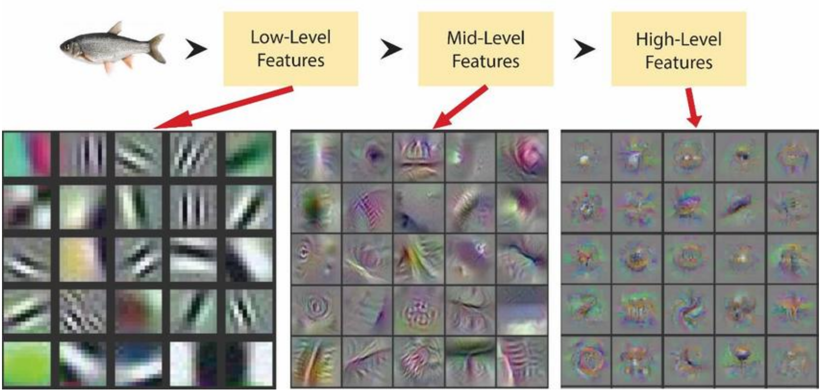


Figure: Hierarchical Aggregation of Features

Concept Partial Order

As it is defined in [1]:

- Formal Context is a triple (G, M, I) where G and M are sets and $I \subset G \times M$ is a relation betweeen them which is callsed incidence relation, G is called objects and M attributes
- For subset of objects $A \subset G$ define $A' = \{m \in M | (g, m) \in I \text{ for all } g \in A\}$
- For subset of attributes $B \subset M$ define $B' = \{g \in G | (g, m) \in I \text{ for all } m \in B\}$
- A concept of the context (G, M, I) is apair (A, B) where $A \subset G$, $B \subset M$ with properties $A' = B$ and $B' = A$
- For two concepts (A_1, B_1) and (A_2, B_2) define order $(A_1, B_1) \leq (A_2, B_2) \iff A_1 \subset A_2 \iff B_1 \subset B_2$

Example of Concepts

	leathery peel	sweet	sour	juicy	contains seeds	made from milk	sandwich filling
banana	×	×			×		
goat cheese						×	×
ice cream		×				×	
kiwifruit		×		×	×		
lemon	×		×	×	×		
orange	×	×		×	×		
strawberry		×		×	×		
yoghurt						×	

Figure: Incidence Table of Fruits.

Hesse Diagram of Concepts

	s	t	u	v	w
A		×	×		
B	×			×	
C		×		×	×
D		×	×	×	×

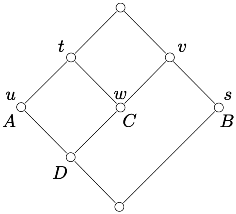


Figure: Cross-table and Hesse Diagram

Context with order might be visualized an Hesse diagram

Concept Lattice

Denote set of all concepts on context (G, M, I) as $\mathcal{B}(G, MI)$
Then it is easy to proofe that $\mathcal{B}(G, MI)$ is complete lattice with meet and join operations:

$$\bigvee_{j \in J} (A_j, B_j) = ((\bigcup_{j \in J} A_j)', \bigcap_{j \in J} B_j)$$

and

$$\bigwedge_{j \in J} (A_j, B_j) = (\bigcap_{j \in J} A_j, (\bigcup_{j \in J} B_j)')$$

With this we can refine context or reduce it, or find appropriate attributes for class of objects for clusterization, classification, knowledge discovery, etc.

Formal Concept Analysis and Decision Trees

If we consider objects as in our dataset and attributes as features, we can define analysis of machine learning model and find similarities with formal concept analysis an machine learning architectures

- In [2] defined several methods for classification using Formal Concept Analysis
- In [3], [4], [5] Formal Concept Analysis is applied for decision tree classification and found different concept classification rules also conversion of lattice to the decision trees

Piecewise Apprimixation Functions

In [13] neural networks are defined as continous piecewise linear functions and representation of each finity pieces continous piecewise linear functions with neural netwoeks established. Which gives isomorphic functor $F : \mathbb{NN} \to \mathbb{CPWL}$ among categories of neural networks and continous piecewise linear functions.
For piecewise approximation defined on X as $\mathcal{F} : X \times P \to Y$ where for each $p \in P$ projection is basis:

$$\mathcal{F}|_{X \times \{p\}} = \varphi : X \to Y$$

we can define $\mathcal{B} \subset 2^X$ as pieces or batches for which approximation happens on one particular basis function.
Define $\mathcal{A} = (\mathcal{A}, \leq)$ partially ordered set and $\alpha : \mathcal{F} \times 2^X \to \mathcal{A}$ which maps each pair $(\varphi, B) = \varphi(B)$ to the single approximation $A \in \mathcal{A}$

- For each $B \in \mathcal{B}$ approximation happens on one particular basis function φ
- $\cup \mathcal{B} = X$

For instance linear regresion or neural network as piecewise universal approximator for l_2 or l_1 with $\mathcal{A} = \mathbb{R}_+$

Spetwise Approximation

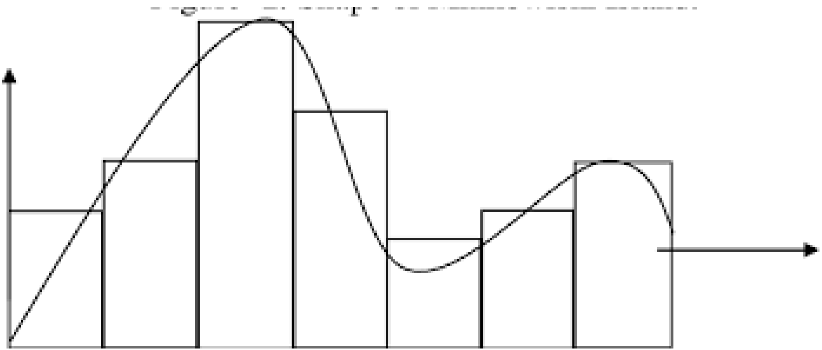
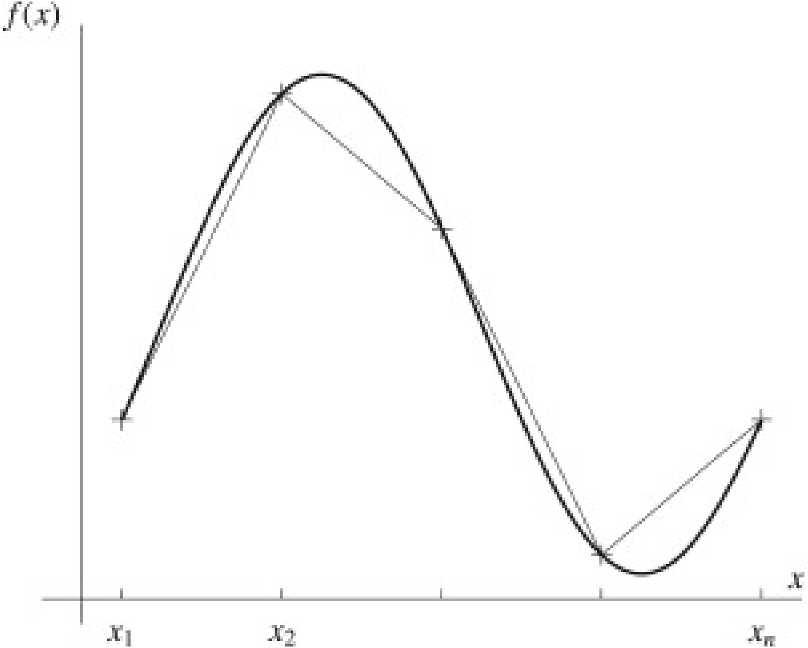


Figure: Step by step approximation (MLP).

Priecwise Linear Approximation



Sigmoid Approximation

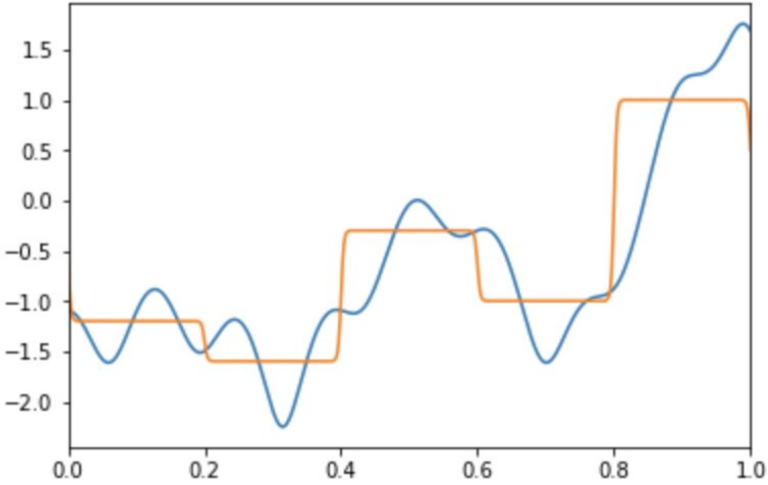


Figure: Piecewise continous nonlinear approximation (Sigmoid).

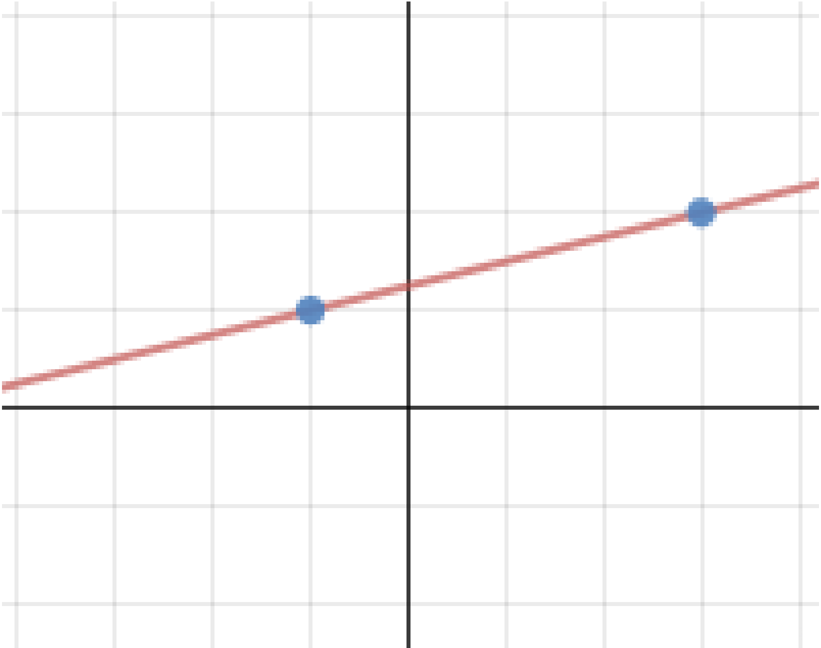
Piecewise linear / continous piecewise linear function as lattice polinomials

Consider pairs (φ, B) on batches and rder:

- $(\varphi, B) \leq (\varphi, D)$ if $\alpha(\varphi(B)) \leq \alpha(\varphi(D)) \iff B \subset D$
- $\varphi(x) \wedge \psi(y) = \mu(\{x, y\})$ where φ, ψ and μ are basis of $\mathcal{F}$

Of course $\wedge$ operation is defined with respect to $\mathcal{A}$ and it is more categorical limit which is unique up to isomorphism

Data Fitting on Two Items



Data Fitting on Two Items

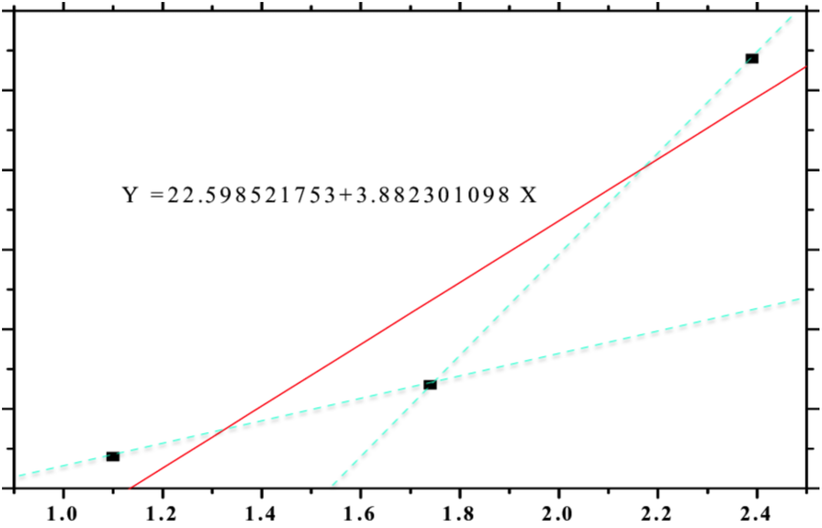


Figure: Approximation on three points.

The Best Approximation Batches

Consider pairs (φ, B) on batches and order:

- $\bigwedge_{\varphi \in \mathcal{F}} (\varphi(B))$
- $\bigvee_{B \in \mathcal{B}} (\varphi, B)$

This gives the best approximation batches $\mathcal{B}$ for $\mathcal{F}$ function and $\mathcal{A}$ approximation.

Lattice Polinomial and Formal Concept

So we can consider neural network, linear regression and multi-linear regressions as:

1. $K(B) = \bigwedge_{\varphi \in \mathcal{F}} \varphi(B)$. fitting on batch
2. $K(\mathcal{A}, \mathcal{F}) = \bigvee_{B \in \mathcal{B}} K(B)$ fitting on functions where $\varphi(x) \vee \psi(x) = \max\left(\alpha(\varphi(x)), \alpha(\psi(x))\right)$ as in [6], [7] and [8] but towards approximation.

Hierarchical Lattice Polinomials

If we define refinement of $\mathcal{B}^l$ by the B^{l+1}:

1. For each $B^l \in \mathcal{B}^l$ there exists $B^{l+1} \in \mathcal{B}^{l+1}$ such that $B^{l+1} \subset B^l$
2. For each $B^l \in \mathcal{B}^l$ we have $\bigcup_{b^{l+1} \subset B^l} B^{l+1} = B^l$

We can distinguish layer of the lattice polinomial:

$$K^{l+1}(\mathcal{A}) = \bigvee_{B^{l+1} \in \mathcal{B}^{l+1}} K^l(B^{l+1})$$

Representation Learning

For later layers we can investigate lattice polinomial of this layer:

$$K^{l+1}(\mathcal{A})$$

If we fix the layer l we can consider approximation / representation power of the appropriated polinomial

Further Work 1

Investigate lattice polinomial and learning objective along with natural topology / metric / (quasi) uniform structures on X and Y spaces

- Investigate model capacity by amount of batches and memorization capacity [12]
- Investigation if kernel machines can be represented as a lattice polinomials
- Compare lattice polinomials of different models and analysis of similarities [9], [10]
- Investigate overparametrization [11]
- Self-supervised representation learning and transfer learning as lattice polinomials [27], [28], [29], [30], [31], [32], [33], [35], [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46].
- Investigate (uniformly) continous property for each lattice polinomial
- Compact and locally compact structures

Further Work 2

- Completness and completion of X and Y spaces
- Tropical geometry of machine learning and deep learning [16], [17]
- Analysis on generalization capaciry and overfitting

Whole area for research.

Thank You

Questions

Thank You Angain

- B. Ganter, R. Wille, Formal Concept Analysis: Mathematical Foundations, Springer-Verlag Berlin/Heidelberg/New York, 1999.
- N. Meddouri, M. Maddouri, Classification Methods based on Formal Concept Analysis, Conference: Concept Lattices and their ApplicationsAt: Olomouc, Czech Republic
- R. Belohlavek, B. De Baets, J. Outrata, V. Vychodil, Inducing decision trees via concept lattices, International Journal of General Systems, volume 38, 2009 - Issue 4: Concept-lattice applications
- E. Dudyrev, S. O. Kuznetsov, Decision Concept Lattice vs. Decision Trees and Random Forests, arXiv:2106.00387
- László Kovács, Generating decision tree from lattice for classification, Proceedings of the 7th International Conference on Applied Informatics Eger, Hungary, January 28-31, 2007. Vol. 2. pp. 377-384.

- J.M. Tarela and M.V. Martinez, Region configurations for realizability of lattice piecewise-linear models, Mathematical and Computer Modelling, 30(1999), 17-27
- J.M. Tarela, J.M. Pérez, V. Aleixandre, Minimization of lattice polynomials on piecewise linear functions (Part I) Annales de l'Association Internationale pour le Calcul Analogique (2) (1975), pp. 79-85
- J.M. Tarela, J.M. Pérez, V. Aleixandre, Minimization of lattice polynomials on piecewise linear functions (Part II) Annales de l'Association Internationale pour le Calcul Analogique (2) (1975), pp. 121-127
- P. Domingos, Every Model Learned by Gradient Descent Is Approximately a Kernel Machine, arXiv:2012.00152
- A. Jacot, F. Gabriel, C. Hongler, Neural Tangent Kernel: Convergence and Generalization in Neural Networks, arXiv:1806.07572

- S. S. Du, X. Zhai, B. Poczos, A. Singh, Gradient Descent Provably Optimizes Over-parameterized Neural Networks, arXiv:1810.02054
- C. Zhang, S. Bengio, M. Hardt, B. Recht, O. Vinyals, Understanding deep learning requires rethinking generalization, arXiv:1611.03530
- J. He, L. Li, J. Xu, C. Zheng, ReLU Deep Neural Networks and Linear Finite Elements, arXiv:1807.03973v2
- K. Hornik, M. Stinchcombe and H. White, Multilayer feedforward networks are universal approximators, Neural networks, 2(1989), 359-366.
- G. Cybenko, Approximation by superpositions of a sigmoidal function, Mathematics of control, signals and systems, 2(1989), 303-314.
- P. Maragos; V. Charisopoulos; E. Theodosis, Tropical Geometry and Machine Learning, Proceedings of the IEEE, Volume: 109 Issue: 5

- L. Zhang, G. Naitzat, L. H. Lim, Tropical Geometry of Deep Neural Networks, arXiv:1805.07091
- J. Adámek, H. Herrlich and G.E. Strecker, Abstract and concrete categories, Wiley Interscience, New York, 1990.
- G.C.L. Brúmmer, Categorical aspects of the theory of quasi-uniform spaces, Rend. Ist. Mat. Univ. Trieste 30 Suppl., 1999, 45-74.
- P. Fletcher and W.F. Lindgren, Quasi-uniform spaces, Lecture Notes Pure Appl. Math. 77, Dekker, New York, 1982.
- P. Fletcher and W.F. Lindgren, Locally quasi-uniform spaces with countable bases, Duke Math. J. 41, 1971, 369-372.
- T.L. Hincks and S.M. Huffuman, A note on locally quasi-uniform spaces, Canad. Math. Bull. Vol. 19(4), 1976, 501-504.
- S. Mac Lane, Categories for the working mathematician, Springer-Verlag, New York-Heidelberg-Berlin, 1971.

- L. Tsinadze, Several types of locally quasi-uniform spaces and their categories, functorial local quasi-uniformities, in preparation.
- L. Tsinadze, Epireflection of completion of localy quasi-uniform spaces, in preparation.
- J. Williams, Locally uniform spaces, Transactions of the American Mathematical Society, Volume 168, 1972, 435-469.
- D.N. Perkins and G. Salomon, Transfer of Learning. Oxford, England: Pergamon, 1992.
- S.J. Pan and Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng., vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Senior Member, IEEE, Hui Xiong, Fellow, IEEE, and Qing HeA, Comprehensive Survey on Transfer Learning

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. A simple framework for contrastive learning of visual representations. arXiv preprint arXiv:2002.05709, 2020.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross B. Girshick. Momentum contrast for unsupervised visual representation learning. arXiv preprint arXiv:1911.05722, 2019

Aäron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748, 2018.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, Michal Valko, Bootstrap your own latent: A new approach to self-supervised Learning

Ian Goodfellow, Yoshua Bengio, Aron Courvile, Deep Learning

Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, Veselin Stoyanov, RoBERTa: A Robustly Optimized BERT Pretraining Approach

Petar Veličković, William Fedus, William L. Hamilton, Pietro Liò, Yoshua Bengio, R Devon Hjelm, Deep Graph Infomax

A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in SIGKDD, 2016, pp. 855–864.

T. N. Kipf and M. Welling, "Variational graph auto-encoders," in NeurIPS Workshop, 2016, pp. 1-3.

Y. Zhu, Y. Xu, F. Yu, Q. Liu, S. Wu, and L. Wang, "Deep Graph Contrastive Representation Learning," in ICML Workshop, 2020.

X. Liu, F. Zhang, Z. Hou, L. Mian, Z. Wang, J. Zhang, and J. Tang, "Self-supervised learning: Generative or contrastive," IEEE TKDE, 2021.

A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon, "A survey on contrastive self-supervised learning," Technologies, vol. 9, no. 1, p. 2, 2021.

Y. Xie, Z. Xu, J. Zhang, Z. Wang, and S. Ji, "Self-supervised learning of graph neural networks: A unified review," arXiv:2102.10757, 2021.

L. Wu, H. Lin, Z. Gao, C. Tan, S. Li et al., "Self-supervised on graphs: Contrastive, generative, or predictive," arXiv:2105.07342, 2021.

Y. You, T. Chen, Y. Sui, T. Chen, Z. Wang, and Y. Shen, "Graph contrastive learning with augmentations," in NeurIPS, vol. 33, 2020, pp. 5812–5823.

Shantanu Thakoor, Corentin Tallec, Mohammad Gheshlaghi Azar, Remi Munos, Petar Veličković, Michal Valko, Bootstrapped Representation Learning on Graphs

L. Tsinadze, Universal Representators in Model Spaces, in preparation.