ICML2016
International Conference on Machine Learning
June 19 – June 24, 2016, New York, United States

Reviews For Paper

Paper ID	1368
Title	Dimensionality Reduction Using Generalized Artificial Neural Networks

Masked Reviewer ID: Assigned_Reviewer_1
Review:

Question	
Summary of the paper (Summarize the main claims/contributions of the paper.)	This paper discusses neural networks with uncountably many inputs. This is motivated as a natural generalization of neural networks with countably many inputs and of a previous paper by Le Roux and Bengio discussing networks with uncountably many hidden units.
Clarity (Assess the clarity of the presentation and reproducibility of the results.)	Below Average
Clarity - Justification	Technical details and formal statements are not formulated in sufficient clarity. The conditions on the integrability of parameters for instance. See the specific comments below.
Significance (Does the paper contribute a major breakthrough or an incremental advance?)	Below Average
Significance - Justification	The topic is interesting. However, in my opinion this paper is not ready for show time. Large parts of the discussion are focused on restating bounded linear operator results form functional analysis.
	* Line 075: functional space -> function space * Line 132: max\{(x,y)... this needs an ordering of pairs. * Line 206: discretized neural networks. Maybe the notation could be more intuitive. The set of variables is discrete, but the variable values are not discrete. * References: Roux -> Le Roux * Line 216: this statement appears incomplete. Also, the notation appears scrambled. Functionals map functions to numbers, and can be expressed in terms of measures. Note also that Cybenko's proof requires discriminatory activation functions. * Line 219: Actually the result does: feedforward networks with one layer of hidden units. * Line 225-233: Why do Lp spaces not provide a natural generalization of finite dimensional vector spaces? * Line 236: Theorem 1 -> Theorem 3

Detailed comments. (Explain the basis for your ratings while providing constructive feedback.)	* In equation 2 it seems that some conditions on integrability should hold. That is, restrictions on the possible types of weight functions. * Theorem 3 seems to follow directly from the fact that $\sum x_i w_i$ can be written as $\int \xi(i) w(i)$. * Corollary 4 appears to be making an unintended uninformative statement. In the current formulation the statement is that for some input the output of the network can approximate some continuous function uniformly. Is ξ some or any function / from what set? Is f some or any continuous function? * Theorem 6 appears to be a direct consequence of the Riesz representation theorem and the Weierstrass polynomial approximation theorem. * Line 411: The statement of Theorem 6 refers to bounded linear operators. Compare this with the statement: A neural network can approximate linear functions $[0,1]^n \to \mathbb{R}$. ... * Line 511: the non-linearity seems to be missing. * Line 700: in a traditional neural network the weights are not discrete but continuous. * Line 704: the ``functional neural networks'' approximate bounded linear operators by construction (and Riesz theorem).
Overall Rating	Weak reject
Reviewer confidence	Reviewer is an expert

Masked Reviewer ID: Assigned_Reviewer_4

Review:

Question	
Summary of the paper (Summarize the main claims/contributions of the paper.)	The paper extends ANNs to infinite dimensional input and output spaces, leading to a NN class called functional neural networks (FNN). It is shown that ANNs are a subset of FNNs and that FNNs are universal approcimators of bounded linear operators. Furtheremore, an even broader set of NNs, generealized NNs, and a sub class, continuous classifier networks, are defined. The letter need less parametets than their ANN analogs, which are shown to exist. Finally, pseudocode for an backprop algorirhm for the continuous classifier is given.
Clarity (Assess the clarity of the presentation and reproducibility of the results.)	Above Average
Clarity - Justification	The paper is quite dense but still good to follow. The derivation of the backprop algorirhm is only in the suplement and thus this part is difficult to understand without it.
Significance (Does the paper contribute a major breakthrough or an incremental advance?)	Above Average

Significance - Justification	The provided generealizarions and theorems are an interesting contribution to the theoretical understanding of NNs and may serve as an interesting link to kernel methods. It would have been nice to see an practical application of the backprop algorirhm.
Detailed comments. (Explain the basis for your ratings while providing constructive feedback.)	I found the title of the paper confusing because "dimensionality reduction" usually refers to the reduction of the dimension of input data and not of the parameters needed by the model. What do the + signs in the end or the equations (5) to (9) indicate? Or are this just typos? In corolarry 4 you could write "set of wight functions w^{(0)}.... w^{(L-1)}" instead of "set of wights W" to distinguish from the wights matrices W in ANNs. In Theorem 6 it should be "operation of layer l+1 on layer l" instead of "operation on layer l on layer l-1". In the notation of the operation g(....) it would be nice to use w(i, j) to make the dependence of g on w direcly visible (same in eq. 14). What is k in eq. 18? Shoudnt this be G[K_t] (s) ? You could swap the order of definition 7 und 8, because 7 needs the definition of T to be well defined. Strictly seen, Definition 8 only defines T_l for l=L-1 and l=0. What about the other l? Minor comments: - regarding eq (6): Using sigma for the sigmoid rather than for the output would be more standard, and the wights w_ij are not defined. - the x's in line 215 and 216 should be fet printed. - line 218: this is result->this result - a subscript m is missing for R in line 239 - komma in beginning of line 246 - line 344: the let - > let - line 445: two dots - line 500: identically equal - line 536: shoul x_n be x_t? - line 681: analohgues
Overall Rating	Weak accept
Reviewer confidence	Reviewer is knowledgeable

Masked Reviewer ID: Assigned_Reviewer_5

Review:

Question	
Summary of the paper (Summarize the main claims/contributions	This paper proposes a generalized neural network formulation, which extends neural networks to the case of an infinite number of hidden nodes. Theoretically, the authors show the universal approximation capability for this type of networks, under mild conditions. Practically, a dimensional reduction method is proposed to make the generalized network tractable

of the paper.)	for continuous data.
Clarity (Assess the clarity of the presentation and reproducibility of the results.)	Above Average
Clarity - Justification	This is a theory paper, and all the theorems are presented clearly, and well explained.
Significance (Does the paper contribute a major breakthrough or an incremental advance?)	Excellent (substantial, novel contribution)
Significance - Justification	Overall, I think this paper gives an important generalization of neural networks, and provides answers to several fundamental questions for this model, i.e., the universal approximation capability and practical implementation.
Detailed comments. (Explain the basis for your ratings while providing constructive feedback.)	1. The motivation of this paper is not clearly explained, and few related works are reviewed. Therefore, it is difficult to evaluate whether this is a timely contribution, and whether it will be potentially interesting to neural network researchers. 2. Since a tractable algorithm is proposed in this paper, it could be interesting to see some empirical studies, which and provide some insights about this algorithm.
Overall Rating	Strong accept
Reviewer confidence	Reviewer's evaluation is an educated guess