Article

Learning Convolutional Neural Layer in High Dimension Space

Jing-Wein Wang [1], $Dinh - ThuanDang^{2}$

[1] Institute of Photonic Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan

[2] Department of Electronic Engineering, National Kaohsiung University of Science and Technology, Kaohsiung 80778, Taiwan

* Correspondence: jwwang@nkust.edu.tw; Tel.: +886-930943143

Version January 3, 2020 submitted to Journal Not Specified

Abstract: we evaluate the replace ReLU non-linear activation layer with the combination between convolution layer and polynomial kernel. If the evaluation results are acceptable, we can remove ReLU layer in neural network model design. This will also exceed the limit deeper the better in neural network architecture. We use LeNet5 baseline CNN model and the NEU surface defect database for defect classification task.

Keywords: convolution neural network; ReLU, Polynomial kernel, classification, NEU surface defect database

---

1. Introduction

https://arxiv.org/pdf/1807.02582.pdf

Role of non-linear activation functions is mean a lot for optimizing in neural network. They allow the model to create complex mapping between the network's inputs and outputs, which are essential for learning and modeling complex data, such as images, video, audio, and data sets which are non-linear or have high dimensionality. Some of non-linear activation functions are used to be applied such as Sigmoid, Tanh, ReLU, Threshold. They are located right behind convolution layer and linear layer.

2. Related Work

3. Representer theorem

3.1. Inner product

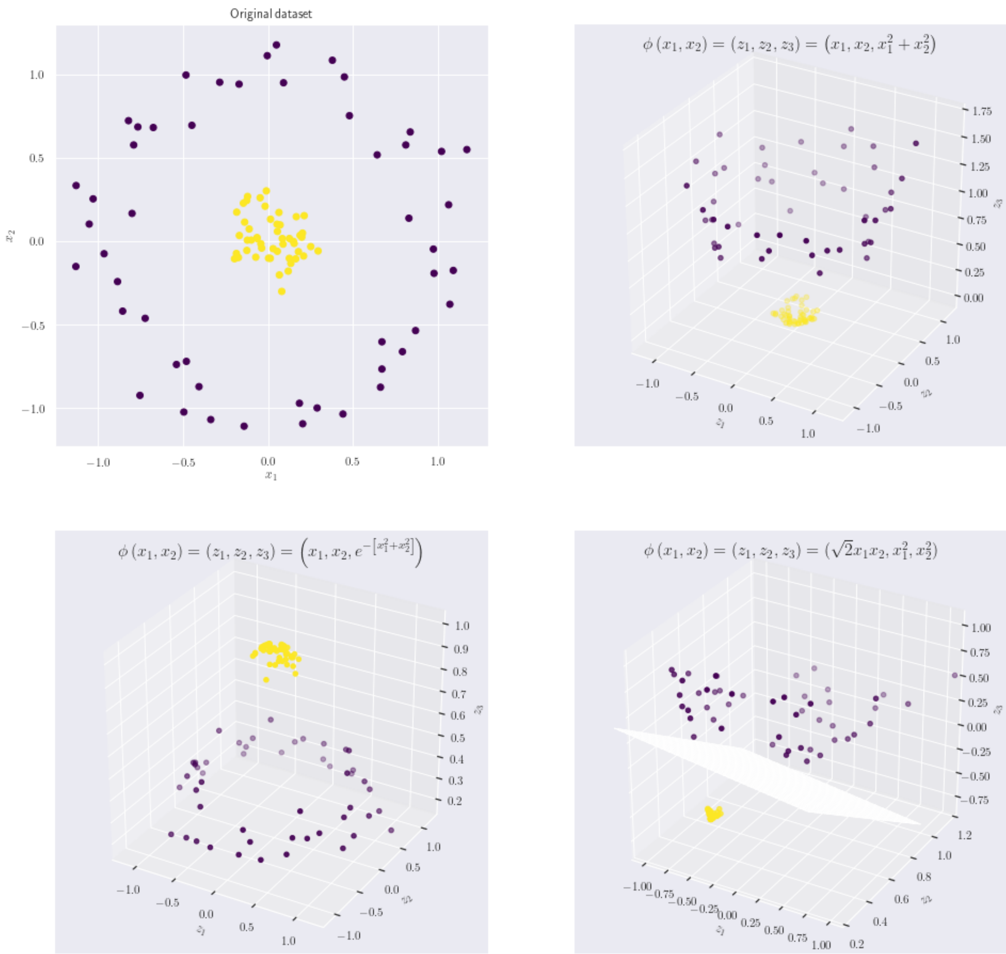
Definition 1: Let $\mathcal{H}$ be a vector space over $\mathbb{R}$. A function $\langle \cdot, \cdot \rangle_{\mathcal{H}} : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is an inner product on $\mathcal{H}$ if all of linear, symmetric, and positive definite properties are satisfied.

1. Linear: $\langle \alpha_1 f_1 + \alpha_2 f_2, g \rangle_{\mathcal{H}} = \alpha_1 \langle f_1, g \rangle_{\mathcal{H}} + \alpha_2 \langle f_2, g \rangle_{\mathcal{H}}$
2. Symmetry: $\langle f, g \rangle_{\mathcal{H}} = \langle g, f \rangle_{\mathcal{H}}$
3. Positive definiteness: $\langle f, f \rangle_{\mathcal{H}} \geq 0$ and $\langle f, f \rangle_{\mathcal{H}} = 0$ if and only if $f = 0$

In this paper the vector space is assumed to be real.
We can define a norm using the inner product as $\|f\|_{\mathcal{H}} := \sqrt{\langle f, f \rangle_{\mathcal{H}}}$. A Hilbert space is a space on which an inner product is defined, along with an Cauchy sequence limits condition.

3.2. Feature mapping

A feature map is a function which maps a input feature. In neural network: it means map input features to hidden units to form new features to feed to the next layer. In Kernel machine: feature mapping means a mapping of features from input space to a reproducing kernel hilbert space, where usually it is very high dimension, or even infinite dimension.



3.3. Kernel

Definition 2: Let $\mathcal{X}$ be a non-empty set. A function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called a kernel if there exists an $\mathbb{R}$-Hilbert space and a feature map $\phi : \mathcal{X} \to \mathcal{H}$ such that $\forall x, x' \in \mathcal{X}$,

$$k\left(x, x'\right) := \left\langle \phi(x), \phi\left(x'\right) \right\rangle_{\mathcal{H}}.$$

Let $\mathcal{X} = \mathcal{H} = \mathbb{R}^2$ and for $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ let $\Phi(x) = \begin{pmatrix} x_1^2 \\ x_2^2 \end{pmatrix}$ Then the kernel is:

$$K\left(x, x'\right) = \Phi(x)^{\top}\Phi\left(x'\right) = \left(x_1\right)^2\left(x_1'\right)^2 + \left(x_2\right)^2\left(x_2'\right)^2$$

3.4. Polynomial kernels

Let $x, x' \in \mathbb{R}^d$ for $d \geq 1$, and let $m \geq 1$ be an integer and $c \geq 0$ be a positive real. Then polynomial kernel is below: $k\left(x, x'\right) := \left(\left\langle x, x'\right\rangle + c\right)^m$

3.5. Kernel tricks

The problem is that the features may live in very high dimensional space, possibly infinite, which makes the computation of the inner product $\left\langle \Phi\left(x_i\right), \Phi\left(x_j\right)\right\rangle$ very difficult. This is where we introduce the notion of a Kernel which will greatly help us perform these computations.

when you can replace $\left\langle x_i, x_j \right\rangle$ with $K\left(x_i, x_j\right)$ and get a nonlinear version of the algorithm. This is equivalent to replacing x with $\Phi(x)$ and replacing $\left\langle x_i, x_j \right\rangle$ with $\left\langle \Phi\left(x_i\right), \Phi\left(x_j\right)\right\rangle$. However, $K\left(x_i, x_j\right) = \left\langle \Phi\left(x_i\right), \Phi\left(x_j\right)\right\rangle$ and $K\left(x_i, x_j\right)$ is much easier to compute.

For $x = \left(x_1, x_2\right)^\top \in \mathbb{R}^2$, let $\Phi(x) = \left(1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2\right) \in \mathbb{R}^5$, we calculate two operator below:

Operator 1: inner product

$$\begin{aligned}\left\langle \Phi\left(\mathbf{x}_i\right), \Phi\left(\mathbf{x}_j\right)\right\rangle &= \left\langle \left\{1, \sqrt{2}x_{i1}, \sqrt{2}x_{i2}, x_{i1}^2, x_{i2}^2, \sqrt{2}x_{i1}x_{i2}\right\}, \left\{1, \sqrt{2}x_{j1}, \sqrt{2}x_{j2}, x_{j1}^2, x_{j2}^2, \sqrt{2}x_{i1}x_{i2}\}\right\rangle \\ &= 1 + 2x_{i1}x_{j1} + 2x_{i2}x_{j2} + x_{i1}^2x_{j1}^2 + x_{i2}^2x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2}\end{aligned}$$

Operator 2: polynomial kernel $K(\mathbf{x}_i, \mathbf{x}_j) = \left(1 + \left\langle \mathbf{x}_i, \mathbf{x}_j \right\rangle\right)^2$

$$\begin{aligned}\left(1 + \left\langle \mathbf{x}_i, \mathbf{x}_j \right\rangle\right)^2 &= \left(1 + x_{i1}x_{j1} + x_{i2}x_{j2}\right)^2 \\ &= 1 + 2x_{i1}x_{j1} + 2x_{i2}x_{j2} + x_{i1}^2x_{j1}^2 + x_{i2}^2x_{j2}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2}\end{aligned}$$

The result of inner product $\left\langle \Phi\left(\mathbf{x}_i\right), \Phi\left(\mathbf{x}_j\right)\right\rangle$ and polynomial kernel $K(\mathbf{x}_i, \mathbf{x}_j)$ is equal, but $K(\mathbf{x}_i, \mathbf{x}_j)$ is much easier to compute. It is only relate to calculate the low dimension space $\left\langle \mathbf{x}_i, \mathbf{x}_j \right\rangle$

4. Method

4.1. Polynomial kernel in convolution layer

Let $\mathbf{x} \in \mathbb{R}^n$ is a vectorized input, $\mathbf{w} \in \mathbb{R}^n$ is the filter, and $\mathbf{f}(\mathbf{x})$ is output. The convolutional operator $\oplus$ is definited:

$$\mathbf{f}(\mathbf{x}) = \mathbf{x} \oplus \mathbf{w} \tag{1}$$

Specifically, the i_{th} element of the convolution output $\mathbf{f}(\mathbf{x})$ is calculated as:

$$\mathbf{f}_i(\mathbf{x}) = \left\langle \mathbf{x}_{(i)}, \mathbf{w} \right\rangle \tag{2}$$

$\mathbf{X}_{(i)}$ the circular shift of $\mathbf{x}$ by $\dot{i}$ elements.

We will custom in convolution operator (1) by replace (2) by (3). that mean we use the feature mapping Φ to transform i_{th} element and filter $\mathbf{w}$ before applying inner product (2).

$$\mathbf{g}_i(\mathbf{x}) = \left\langle \Phi\left(\mathbf{x}_{(i)}\right), \Phi(\mathbf{w}) \right\rangle \tag{3}$$

The feature mapping Φ is calculated by using polynomial kernel easily.

4.2. convergence

4.3. Learnable Kernel

5. Result

Choose a corresponding Kernel. working in the high-dimension while doing computation in the original low dimensional space

6. Conclusions

7. Patents

This section is not mandatory, but may be added if there are patents resulting from the work reported in this manuscript.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used "conceptualization, X.X. and Y.Y.; methodology, X.X.; software, X.X.; validation, X.X., Y.Y. and Z.Z.; formal analysis, X.X.; investigation, X.X.; resources, X.X.; data curation, X.X.; writing–original draft preparation, X.X.; writing–review and editing, X.X.; visualization, X.X.; supervision, X.X.; project administration, X.X.; funding acquisition, Y.Y.", please turn to the CRediT taxonomy for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

Funding: Please add: "This research received no external funding" or "This research was funded by NAME OF FUNDER grant number XXX." and and "The APC was funded by XXX". Check carefully that the details given are accurate and use the standard spelling of funding agency names at https://search.crossref.org/funding, any errors may affect your future funding.

Acknowledgments: In this section you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

Conflicts of Interest: Declare conflicts of interest or state "The authors declare no conflict of interest." Authors must identify and declare any personal circumstances or interest that may be perceived as inappropriately influencing the representation or interpretation of reported research results. Any role of the funders in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript, or in the decision to publish the results must be declared in this section. If there is no role, please state "The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results".

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	linear dichroism

Appendix A

Appendix A.1

The appendix is an optional section that can contain details and data supplemental to the main text. For example, explanations of experimental details that would disrupt the flow of the main text, but nonetheless remain crucial to understanding and reproducing the research shown; figures of replicates for experiments of which representative data is shown in the main text can be added here if brief, or as Supplementary data. Mathematical proofs of results not central to the paper can be added as an appendix.

Appendix B

All appendix sections must be cited in the main text. In the appendixes, Figures, Tables, etc. should be labeled starting with 'A', e.g., Figure A1, Figure A2, etc.

References

1. Author1, T. The title of the cited article. *Journal Abbreviation* **2008**, *10*, 142–149.
2. Author2, L. The title of the cited contribution. In *The Book Title*; Editor1, F., Editor2, A., Eds.; Publishing House: City, Country, 2007; pp. 32–58.

Sample Availability: Samples of the compounds ...... are available from the authors.