Error Correction Output Codes for Robust Neural Networks against Weight-errors: A Neural Tangent Kernel Point of View

Anlan Yu*, Shusen Jing*, Ning Lyu, Wujie Wen, Zhiyuan Yan

*Equal contribution

LEHIGH UNIVERSITY

What's ECOC?

One-hot labeling

Label 0	1	0	0	0	0	0	0	0
Label 1	0	1	0	0	0	0	0	0
Label 2	0	0	1	0	0	0	0	0
…	…							
Label 6	0	0	0	0	0	0	1	0
Label 7	0	0	0	0	0	0	0	1

Minimum distance = 2

1. Output i: predict the probability of **the given input belongs to class i**
2. Num output equal to num class
3. Code matrix as identity matrix

ECOC labeling

Label 0	1	1	1	1	1	1	1	1
Label 1	1	0	1	0	1	0	1	0
Label 2	1	1	0	0	1	1	0	0
…	…							
Label 6	1	1	0	0	0	0	1	1
Label 7	1	0	0	1	0	1	1	0

Minimum distance = 4

1. Output i: predict the probability of **the i-th output being 1**
2. Num output **self-defined**
3. Code matrix **self-defined**

ECOC: an ensemble of binary classifiers

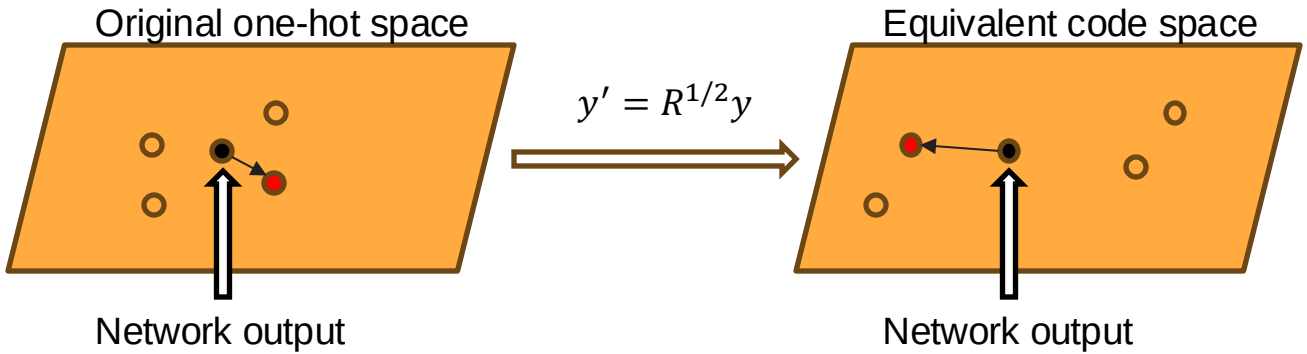
Fundamental problems of ECOC

Lack theoretical foundations

- 🧐 Why does ECOC work?
- 🤔 How good can ECOC be?
- 🤯 How to make ECOC work better?

Efficacy of ECOC in Absence of Weight-errors

- Theoretically prove that ECOCs affect the performance of DNNs through the correlation matrix R of the code matrix



Efficacy of ECOC on the Robustness of DNNs

Conclusion:

The ratio of Hamming distance to code length is crucial for the robustness of DNNs with ECOCs.

ECOC Construction: Directly Optimize

- **Weight-error free:** correlation matrix R of code matrix
- **Robustness:** Hamming distance of codes

Objective function:

$$\min_{Z\in\{-1,1\}^{n_L\times C}} \underbrace{-\sum_{i\neq j}\|Z[i]-Z[j]\|^2}_{\text{pair-wise codeword distance}} + \lambda \underbrace{\left(\sum_{i\neq j}(Z[i]^T Z[j])^2 - \beta\sum_i \|Z[i]\|^2\right)}_{\text{correlation}}.$$

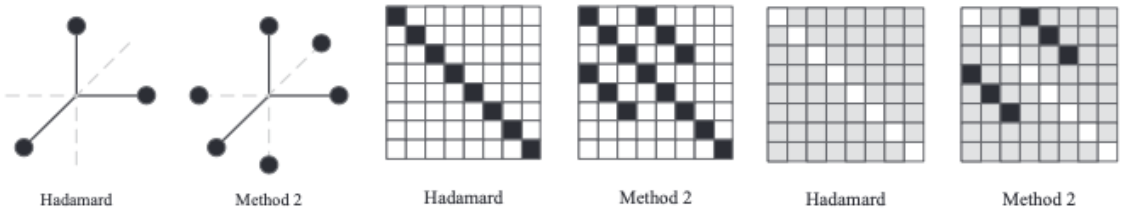
Z is the ECOC codebook

ECOC Construction: Picking from Hadamard

Properties of Hadamard code:

1. Power of 2 code length
2. Codewords orthogonal
3. Hamming distance half of code length

$$H(8) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}$$



(a) Geometric Structure (b) Correlation Matrix (c) Distance Matrix

Thank you

LEHIGH UNIVERSITY