Neural Network Classifier as Mutual Information Evaluator

Zhenyue Qin[1], Dongwoo Kim[1,2], Tom Gedeon[1] Australian National University[1], POSTECH[2]

Paper Code

Motivation

- The combination of softmax with cross-entropy is a standard choice to train neural network classifiers.
- It measures the cross-entropy between label and network output.
- We aim to know what relation the network aims to model between input x and label y.
- Also, arguably, it seem to be artificial to view the output of softmax as probabilities.
- We aim to rethink the meaning of softmax and cross-entropy.

Connecting Softmax with Cross-Entropy to Mutual Information

- Definition of mutual information:

$$I(X,Y) = \int_{x\in X}\left[\sum_{y\in Y} P(x,y)\log\frac{P(x,y)}{P(x)P(y)}\right]dx$$
$$= E_{(X,Y)}[\log\frac{P(y|x)}{P(y)}]$$
$$\geq E_{(X,Y)}\left[\log\frac{Q(y|x)}{P(y)}\right]$$

- If we make $Q(x,y) = \frac{P(x)P(y)}{E_{y'}[e^{f_\phi(x,y')}]}e^{f_\phi(x,y)}$, we have:

$$I(X,Y) \geq E_{(X,Y)}[\log\frac{e^{f_\phi(x,y)}}{E_{y'}e^{f_\phi(x,y')}}]$$

- For the denominator, if the dataset is balanced:

$$E_{y'}\left[e^{f_\phi(x,y')}\right] = \frac{1}{M}\sum_{y'=1}^{M} e^{f_\phi(x,y')}$$

- Then, the RHS of (1) becomes softmax + M.
- Thus, given a balanced dataset, minimizing softmax with cross-entropy is to maximize the mutual information between input and label.
- If the dataset is imbalanced, we propose PC-softmax:

$$\frac{e^{f_\phi(x,y)}}{\sum_{y'=1}^{M} P(y')e^{f_\phi(x,y')}}$$

Experimental Performance: Estimating Mutual Information

Dim	Accuracy		Mutual Information			
	Softmax	PC-Softmax	GT	MINE	Softmax	PC-Softmax
1	79	79	1.02	0.99	1.11	0.96
2	87	88	1.23	1.17	1.31	1.20
5	93	95	1.44	1.27	1.41	1.31
10	95	96	1.48	1.22	1.36	1.34

Mutual information estimation results with softmax-based classification neural networks.

Experimental Performance: Classification

Dataset	MNIST		CUB-200-2011	
	Balanced	Imbalanced	Balanced	Imbalanced
Softmax	97.95	95.05	89.21	84.63
PC-Softmax	97.91	96.30	89.16	87.69

Average-per-class classification accuracy of using softmax and PC-softmax.