Supplementary material for
PR Product: A Substitute for Inner Product in Neural Networks

1. Training Curves on CIFAR10
   Figures 1-4 show the training curves of some classification models on CIFAR10 used in the paper, from which we can see that the models of PR Product version get consistent lower error rates than the models of standard inner product version.

2. The Minimum of $|\sin \theta|$ 
   We plot the minimum of $|\sin \theta|$ in some layers of our captioning model, as shown in Figures 5-13. From these plots, we can observe that the minimum of $|\sin \theta|$ in PR Product version is larger than the one in P Product version for most of the layers, which means the weight vector and data vector in PR Product are more orthogonal. We argue this is the reason for PR Product to take effect.
Figure 7. The minimum of $|\sin \theta|$ of the $W_s \Pi_t$ to hidden transfer part in the Attention LSTM.

Figure 8. The minimum of $|\sin \theta|$ of the $v_k$ to hidden transfer part in the Attention LSTM.

Figure 9. The minimum of $|\sin \theta|$ of the hidden to hidden transfer part in the Attention LSTM.

Figure 10. The minimum of $|\sin \theta|$ of the $\hat{v}_t$ to hidden transfer part in the Language LSTM.

Figure 11. The minimum of $|\sin \theta|$ of the $h_1^t$ to hidden transfer part in the Language LSTM.

Figure 12. The minimum of $|\sin \theta|$ of the hidden to hidden transfer part in the Language LSTM.

Figure 13. The minimum of $|\sin \theta|$ of the output layer (softmax layer in the Decoder of our captioning model.)

3. Examples of Image Captioning

To intuitively illustrate the advantage of the PR Product, we show some examples of image captioning in Figure 14. The images are sampled from Karpathy’s test split of MS COCO dataset. All the three models (P version, R version, and PR version) are trained with cross-entropy loss and then fine-tuned for CIDEr optimization. The results show that PR product makes contribution to the descriptiveness of the sentences and prove that the PR Product is effective.
GT: two teddy bears lie propped up against a wall
P: a teddy bear with a ball in front of it
R: a teddy bear holding a group of balloons
PR: two teddy bears sitting next to each other

GT: a large passenger jet flying through a cloudy sky
P: a large airplane flying in the sky
R: a large airplane flying in a cloudy sky
PR: a large airplane flying in a cloudy sky

GT: a group of people in a pool with floating plates of food
P: a group of people sitting at a table
R: a group of people sitting around a table
PR: a group of people in a pool

GT: a group of people gathered at the bottom of a snow mountain
P: a group of people on skis on a ski lift
R: a group of people riding skis on a ski lift
PR: a group of people on skis on a snow covered mountain

GT: a group of people sitting in a shelf
P: a bunch of cats sitting in a shelf
R: a group of cats sitting on top of a shelf
PR: a group of shoes sitting on top of a shelf

GT: a bowl of broccoli sits beside a lemon wedge
P: a white plate of broccoli and orange slices on a table
R: a white plate of broccoli and a table
PR: a white plate of broccoli and a lemon wedge

GT: a bowl of broccoli sits beside a lemon wedge
P: a white plate of broccoli and orange slices on a table
R: a white plate of broccoli and a table
PR: a white plate of broccoli and a lemon wedge

GT: a no parking sign is attached to a traffic cone on a sidewalk
P: a sign on the side of a street
R: a street sign on the side of a road
PR: a no parking sign on the side of a street

GT: a woman standing outside taking a picture with her cellphone
P: a woman holding a cell phone in her hand
R: a woman holding a cell phone in her hand
PR: a woman taking a picture with her cell phone

Figure 14. Examples of captions on MS COCO dataset. GT: human ground truth. P: sentence generated by the P Product version model. R: sentence generated by the R Product version model. PR: sentence generated by the PR Product version model. Obviously, the PR Product performs better than the P Product and the R Product.