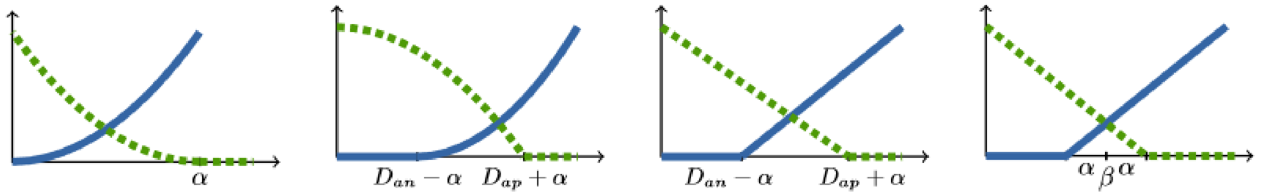
URL: https://arxiv.org/abs/1706.07567

Main idea

个人感觉这是对度量学习中的sampling机制和triplet loss理解很透彻的一篇论文，也被用在私有数据库的ReID Codebase上了。论文主要在强调两个问题：（1）对deep embedding learning来说，sampling方法很重要，负样本距离太小则负样本的梯度方向容易受噪声影响；（2）contrasive loss的margin限制太强，batch hard的point to point的triplet loss存在明显的弊端：难的负样本挖掘不足导致负样本梯度接近0，网络很容易收敛到一个点，从而网络崩溃。因此论文提出了一种基于距离加权的采样方法和一种新的Margin based loss。

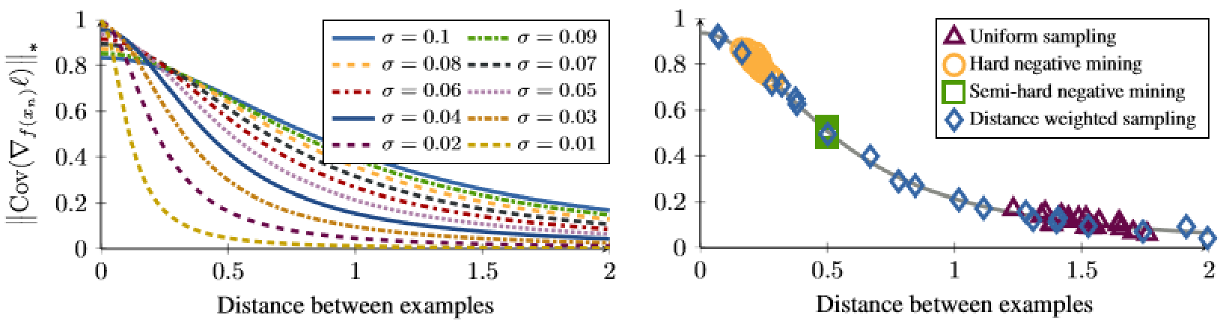


(a) Contrastive loss [11] (b) Triplet loss ℓ_2^2 [25] (c) Triplet loss ℓ_2 (d) Margin based loss

Figure 4: Loss vs. pairwise distance. The solid blue lines show the loss function for positive pairs, the dotted green for negative pairs. Our loss finds an optimal boundary β between positive and negative pairs, and α ensures that they are separated by a large margin.

Experiment Detail

- 难负样本如果太近，会给梯度引入很多噪声，而基于距离加权的采样方法可以由更大的距离采样区间。



(a) Variance of gradient at different noise levels. (b) Sample distribution for different strategies.

Figure 3: (a) shows the nuclear norm of a noisy gradient estimate for various levels of noise. High variance means the gradient is close to random, while low variance implies a deterministic gradient estimate. Lower is better. Note that higher noise levels have a lower variance at distance 0. This is due to the spherical projection imposed by the normalization. (b) shows the empirical distribution of samples drawn for different strategies. Distance weighted sampling selects a wide range of samples, while all other approaches are biased towards certain distances.

- 负样本的采样方法为：around a small region. To avoid noisy samples, we clip the weighted sampling. Formally, given an anchor example a, distance weighted sampling samples negative pair $(a, n^\star)$ with

$$\Pr\left(n^\star = n|a\right) \propto \min\left(\lambda, q^{-1}\left(D_{an}\right)\right).$$

- 结合 contrastive loss 的相对高效与 triplet loss的margin限制更soft的优势，提出改进的margin based loss。 $\ell^{\text{margin}}(i,j) := \left(\alpha + y_{ij}(D_{ij}-\beta)\right)_+$. beta是决定正负样本对的分界线，alpha是决定分割间距的margin。且beta是在线自适应学习的。

Thoughts

这篇文章对sampling和triplet loss的思考很深入，我目前还没有完全领悟，还需要反复斟酌。且论文行文思路清晰，实验论证严谨，值得学习。