Uncorrelation and Evenness: A New Diversity-Promoting Regularizer Supplementary Material

Pengtao Xie [1 2] **Aarti Singh** [1] **Eric P. Xing** [2]

1. Additional Explanation of Uncorrelation

If components (denoted by horizontal and vertical axis in Figure 2a) are correlated, then samples (points in Figure 2a) are in a non-spherical shape, then eigenvalues are mutually different. Hence correlation leads to non-uniformity of eigenvalues. Since the eigenvectors are orthogonal by design, it suffices to focus on eigenvalues only. To reduce correlation, we encourage the eigenvalues to be uniform (Figure 2b). Rotation does not affect eigenvalues or uncorrelation. For a component matrix $\mathbf{A}$ and rotation matrix $\mathbf{R}$, $\mathbf{A}^\top\mathbf{A}$ equals to $\mathbf{A}^\top\mathbf{R}^\top\mathbf{R}\mathbf{A}$ and they have the same eigen-decomposition (say $\mathbf{U}\mathbf{E}\mathbf{U}^\top$). Ensuring the eigenvalue matrix $\mathbf{E}$ is close to identity implies the latent components are rotations of the orthonormal (and hence uncorrelated) eigenvectors.

2. Additional Experiments

Other than distance metric learning and LSTM network, we also applied our proposed regularizer to latent Dirichlet allocation (LDA) (Blei et al., 2003) and ensemble of support vector machines (SVM) (Yu et al., 2011). The "diversification" of these two models has been studied in (Zou & Adams, 2012) and (Yu et al., 2011) respectively. We made a comparison with them.

2.1. "Diversifying" Latent Dirichlet Allocation

An LDA consists of m topics, each parameterized by a vector $\boldsymbol{\beta} \in \mathbb{R}^d$. We apply the proposed UER to the m topic vectors $\mathcal{B} = \{\boldsymbol{\beta}_j\}_{j=1}^m$ and compare with the DPP regularized LDA proposed in (Zou & Adams, 2012). The experimental settings follow those in (Zou & Adams, 2012). The dataset is 20-Newsgroups and the number of topics is fixed to 25. The inferred topic proportion vectors are fed into a SVM to perform document classification. Without regularization, the classification accuracy is 23.1%. Under DPP, the accuracy is 23.8%. Under our regularizer, the accuracy is 24.7%.

[1]Machine Learning Department, Carnegie Mellon University [2]Petuum Inc. Correspondence to: Pengtao Xie, Eric P. Xing <pengtaox@cs.cmu.edu, eric.xing@petuum.com>.

Proceedings of the 34^{th} International Conference on Machine Learning, Sydney, Australia, PMLR 70, 2017.

	No Regularization	Yu et al. (2011)	UER
Clean1	50.0	35.8	30.2
Ethn	32.1	14.9	11.4
German	30.2	28.0	27.3
Haberman	36.0	30.3	26.9
Vehicle	25.0	22.6	21.1

Table 1. Classification errors (%)

2.2. "Diversifying" Ensemble of SVMs

In an SVM ensemble, there are m base SVM classifiers, each parameterized by a weight vector $\mathbf{w} \in \mathbb{R}^d$. We apply UER to "diversify" the m vectors and compare with the regularizer proposed by (Yu et al., 2011). The experimental settings follow those in (Yu et al., 2011). The number of base SVMs is set to 21 and the ensemble method is Bagging. Table 1 shows the classification errors on 5 UCI datasets.

References

Blei, David M, Ng, Andrew Y, and Jordan, Michael I. Latent dirichlet allocation. *the Journal of machine Learning research*, 2003.

Yu, Yang, Li, Yu-Feng, and Zhou, Zhi-Hua. Diversity regularized machine. 2011.

Zou, James Y and Adams, Ryan P. Priors for diversity in generative latent variable models. In *Advances in Neural Information Processing Systems*, pp. 2996–3004, 2012.