Transfer Learning and Optimal Transport - Advanced Machine Learning

Beltrán Castro Gómez
beltran.castro.gomez@gmail.com

Abstract

This document aims to present the results obtained after the implementation and testing of the Subspace alignment and the Entropic regularized Optimal Transport methods on the Office/Caltech dataset.

1. Office/Caltech dataset

The Officee+Caltech10 dataset is the dataset used in this project (Gong et al., 2012). It is an object recognition dataset and a standard benchmark for domain adaptation tasks. It is composed of images belonging to 10 different classes and it includes 4 different visual domains:

- **C**: Caltech domain with 1123 samples.
- **A**: Amazon domain with 958 samples.
- **W**: Webcam domain with 295 samples.
- **D**: DSLR domain with 157 samples.

The domains differ between each other in several aspects, such as scene, lighting, object positioning, view angle, resolution or background.
These make up to 12 adaptation pairs but, for simplicity, in this project we will only use the W and D domains, having the Webcam domain as the source domain and D as the target one.
In addition, this datasets presents more than one feature representations. These are the SURF representation, with 800 features, the GoogleNet1024 representation, with 1024 features and, finally, the CaffeNet4096 representation, with 4096 features.

2. Subspace alignment

Subspace alignment is an algorithm for Domain Adaptation that aligns both the source and target data subspaces, generating a new feature space that is domain independent (Fernando et al., 2014).

For projecting the labeled source data S and unlabeled target data T, we need to calculate the alignment matrix M, which is obtained by multiplying their d principal components. The steps for performing subspace alignment are the following (Fernando et al., 2014):

1. Standardize S and T.
2. Calculate the d principal components of S (X_s) and T (X_t).
3. Calculate the alignment matrix as $M = X_s^T X_t$ and also $X_a = X_s M$.
4. Obtain the projected data $S_a = S X_a^T$ and $T_a = T X_t^T$.

2.1. Results with the SURF features

With the goal of perceiving the improvement achieved thanks to subspace alignment, I initially trained a 1-NN classifier on the source data and tested it on the target data to see how well this model would perform on the raw data (before any projection) when changing domains, which is something that could actually happen in real life applications. The accuracy outputted by this classifier was of 0.5605.

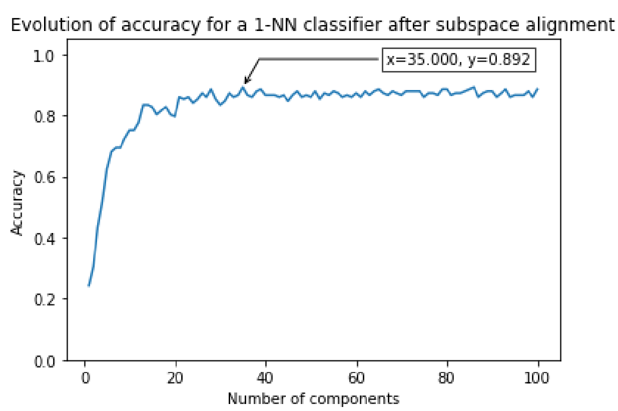

Figure 1. Evolution of test accuracy for a 1-NN classifier after subspace alignment for different d values and the SURF features.

However, after subspace alignment and tuning of the d parameter, the classifier improved up to a 0.8917 of test accuracy for $d = 35$, as it can be seen in Figure 1.

2.2. Results with the GoogleNet1024 features

For the GoogleNet1024 case and, contrarily to the SURF one, the results obtained by the 1-NN classifier on the raw data were already very good, probably due to the expresiveness of the feature, achieving a test accuracy of 0.9809.
After performing subspace alignment and tuning the d hyperparameter, the classifier improved up to a perfect classification and 1.0 test accuracy, for $d = 17$ and for a lot of other d values, as it can be observed in Figure 2.

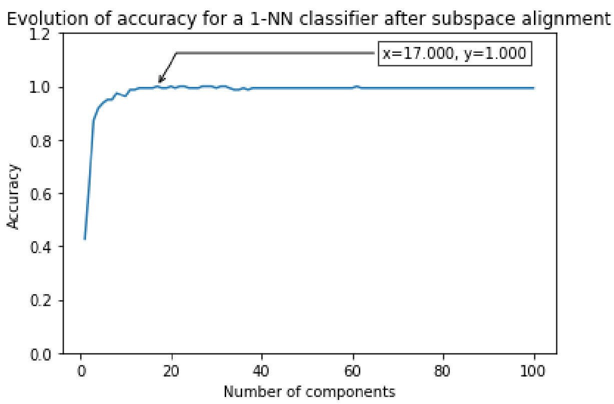

Figure 2. Evolution of test accuracy for a 1-NN classifier after subspace alignment for different d values and the GoogleNet1024 features.

3. Entropic Regularized Optimal Transport

The classical optimal transport problem contemplates a scenario where we have to transport earth from one place to another one in an optimal way (Monge, 1781).
The Kantorovich Problem (Kantorovitch, 1958) introduces a relaxation to the classical OT problem by allowing to split the mass of probability. The Kantorovich formulation aims to calculate the γ probabilistic coupling described in the Equation 1 below:

$$\gamma^* = \min_{\gamma} \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ij} d_{ij} = \min_{\gamma} \langle \gamma, D \rangle_F$$
$$s.t. \gamma \in U_{ab} = \{\gamma \in \mathbb{R}_+^{nxm}; \gamma 1_m = a; \gamma^T 1_n = b\} \quad (1)$$

However, this formulation still has some limitations, since γ^* is unstable and not always unique. To cope with that, we can introduce Entropic Regularization (Cuturi, 2013), that will impose uniqueness due to the strong convexity of the entropy:

$$W_\gamma(\mu, \nu) = \min_{\gamma \in U(a,b)} \langle \gamma, D \rangle_F - \lambda E(\gamma) \quad (2)$$

The steps we will take to solve the entropic regularized OT problem between our S and T domains presented before are the following:

1. Standardize S and T.
2. Define two uniform vectors a of size n_s and b of size n_t.
3. Calculate the cost matrix as M with the distances between S and T instances.
4. Learn the coupling matrix γ by using the Sinkhorn-Knopp algorithm (Cuturi, 2013).
5. Transform the points from source to target by $S_a = \gamma T$.

3.1. Results with the SURF features

After tuning the regularization parameter with $\lambda \in [\,0.0001, 0.01]$, the best test accuracy achieved by the 1-NN classifier was of 0.9745 for $\lambda = 0.006$, as it can be seen in Figure 3.

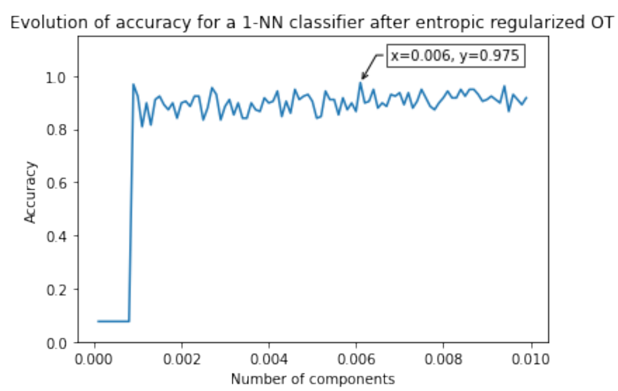

Figure 3. Evolution of test accuracy for a 1-NN classifier after entropic regularized OT for different λ values and the SURF features.

This clearly improved the best result of subspace alignment from the previous section of 0.8917 of test acuraccy. It also took more time to compute for the same amount of tuning values.

3.2. Results with the GoogleNet1024 features

When using the GoogleNet1024 features, the results obtained by the 1-NN classifier after tuning the regularization parameter (see Figure 4) are surprisingly a bit worse than in the previous case for the subspace alignment. The best score achieved was of 0.9554 for $\lambda = 0.0075$. It was of 1.0 for subspace alignment and it was even better with the SURF features (0.9745).

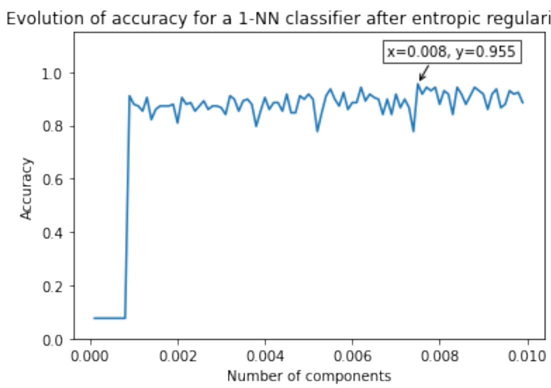

Figure 4. Evolution of test accuracy for a 1-NN classifier after entropic regularized OT for different d values and the GoogleNet1024 features.

References

Cuturi, M. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.

Fernando, B., Habrard, A., Sebban, M., and Tuytelaars, T. Subspace alignment for domain adaptation, 2014. URL https://arxiv.org/abs/1409.5241.

Gong, B., Shi, Y., Sha, F., and Grauman, K. Geodesic flow kernel for unsupervised domain adaptation. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2066–2073, 2012. doi: 10.1109/CVPR.2012.6247911.

Kantorovitch, L. On the translocation of masses. *Management science*, 5(1):1–4, 1958.

Monge, G. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pp. 666–704, 1781.