Transfer Learning of Statistical Models on Riemannian Manifolds

Tien-Tam Tran, Ines Adouani and Chafik Samir

UCA CNRS LIMOS (UMR 6158)

Introduction

What?

- **Transfer Learning:** A model learned from one set can be a prior for the second.
- **Statistics on Manifolds:** The data on manifold is mapped into tangent space by the Log map, from this we consider the covariance, PCA, Linear Regression.

Why?

- Is model learned in one region are utilized in another?

How?

- Let M be manifold and $T_p(M)$ and $T_q(M)$ be tangent spaces.
- Suppose we have a data in $T_p(M)$, **Parallel Transport data** to $T_q(M)$ then learn and use this model in $T_q(M)$. This scales poorly.
- Different approach is **Model Transport**: the model is learned in $T_p(M)$ then transport the model to $T_q(M)$. Computational complexity is free from the size of data set.

Geometry on the sphere

- A geodesic on the sphere is a great circle, parallel transport given by rotation.

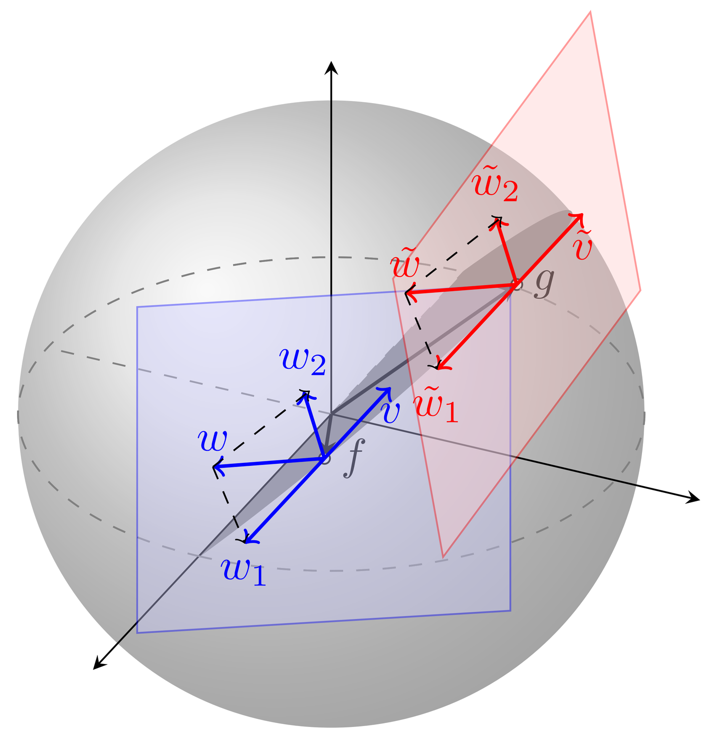


Figure 1: Parallel transport of w at T_f is factored in to two perpendicular components w_1 and w_2.

Discussion and conclusion

- We have studied the geometry of finite probability measures and its Transfer learning.
- The tangent space $T\mathcal{P}_+(I)$ is trivial.
- Geodesics and Parallel transport are solved explicitly.
- In general, Parallel transport has no closed form, we have to approximate it.
- In the Box-plots, we see that the transfered models are comparable to the usual one.

References

[1] O. Freifeld, S. Hauberg, and M. J. Black. Model transport: Towards scalable transfer learning on manifolds. *CVPR*, pages 1378–1385, 2014.

[2] N. Ay, J. Jost, H. Le, and L. Schwachhofer. *Information geometry*. Springer, 2017.

[3] M. Itoh and H. Satoh. Geometry of fisher information metric and the barycenter map. *entropy*, pages 1814–1849, 2015.

The geometry of finite probability measures

Let $I = \{1, \dots, n, n+1\}$ be a finite set.

- **Measure space:** $\mathcal{P}_+(I) = \left\{\mu = \sum_{i\in I} \mu_i \delta^i \mid \mu_i > 0, \quad \forall i \in I, \text{ and } \sum_{i\in I}\mu_i = 1\right\}$.
- **Tangent space:** $T_\mu\mathcal{P}_+(I) = \{\mu\} \times \mathcal{S}_0(I)$, where $\mathcal{S}_0(I) = \left\{\sum_{i\in I}\mu_i\delta^i \mid \sum_{i\in I}\mu_i = 0\right\}$.
- **Fisher-Rao metric:** $\mathfrak{g}_\mu(X,Y) = \sum_{i\in I}\frac{X_iY_i}{\mu_i}, \quad \forall X = \sum_{i\in I} X_i\delta^i, Y = \sum_{i\in I} Y_i\delta^i \in T_\mu\mathcal{P}_+(I)$.
- **Levi-Civita connection ∇:** Let X, Y be constant vector fields, $\frac{dX}{d\mu}$ is Radon-Nykodym derivative. Then

$$\nabla_X Y|_\mu = -\frac{1}{2}\left(\frac{dX}{d\mu}\frac{dY}{d\mu} - \mathfrak{g}_\mu(X,Y)\right)\mu, \tag{1}$$

- **Geodesics:** Given $\mu \in \mathcal{P}_+$ and a unit tangent vector $X \in T_\mu\mathcal{P}_+$. Then there exist a unique geodesic $\alpha(t) = \sum_{i\in I}\alpha_i(t)\delta^i$ in $\mathcal{P}_+(I)$, starting from a point μ with direction X, where

$$\alpha_i(t) = \left(\cos\frac{t}{2} + \frac{X_i}{\mu_i}\sin\frac{t}{2}\right)^2 \mu_i, \quad \text{for } i \in I. \tag{2}$$

- **Distance:** For every $\mu, \nu \in \mathcal{P}_+(I)$ we have $d(\mu,\nu) = 2\arccos\left(\sum_{i\in I}\sqrt{\mu_i\nu_i}\right)$.
- **Exponential map:** $\exp_\mu(X) = \sum_{i\in I}\left(\cos\frac{\|X\|_\mu}{2} + \frac{X_i}{\mu_i\|X\|_\mu}\sin\frac{\|X\|_\mu}{2}\right)^2\mu_i\delta^i$.
- **Log map:** $\log_\mu(\nu) = \frac{\ell}{\sin\frac{\ell}{2}}\sum_{i\in I}\left(\sqrt{\frac{\nu_i}{\mu_i}} - \sum_{j\in I}\sqrt{\nu_j\mu_j}\right)\mu_i\delta^i$, where $\ell = d(\mu,\nu)$.
- **Parallel transport:** Let $\mu, \nu \in \mathcal{P}_+(I)$ and $\alpha : [0,\ell] \to \mathcal{P}_+(I)$ be a geodesic curve such that $\alpha(0) = \mu$ and $\alpha(\ell) = \nu$. The parallel transport, $\Gamma_{\mu\to\nu} : T_\mu\mathcal{P}_+(I) \to T_\nu\mathcal{P}_+(I)$, given by

$$\Gamma_{\mu\to\nu}(X) = \sum_{i\in I}\sqrt{\nu_i}\left(-C\sqrt{\mu_i}\left(2\sin\frac{\ell}{2} - 2\frac{\tau_i}{\mu_i}\cos\frac{\ell}{2}\right) + \frac{X_i}{\sqrt{\mu_i}} - 2C\frac{\tau_i}{\sqrt{\mu_i}}\right)\delta^i, \tag{3}$$

where τ is the unit tangent vector $\tau = \log_\mu(\nu)/\ell$ and $C = \frac{1}{2}\mathfrak{g}_\mu(X,\tau)$.

- **Isometry:** By the map $\Phi(\mu) = 2\sum_{i\in I}\sqrt{\mu_i}e_i$, $\mathcal{P}_+(I)$ is isometric to the sphere

$$\mathbb{S}_{(0,2),+}(I) = \left\{f \in \mathbb{R}^{n+1} \mid f^i > 0, \quad \forall i \in I, \text{ and } \sum_{i\in I}(f^i)^2 = 4\right\}. \tag{4}$$

Pull-back from the sphere, we find again the geodesics and parallel transport.

Transfer Learning

- We have the data $\{x_i\}_{i=1}^N \subset T_\mu\mathcal{P}_+(I)$ and its labels $\{y_i\}_{i=1}^N$.
- A linear regression model in $T_\mu\mathcal{P}_+(I)$ has the following form

$$X \mapsto Y(X) = X^Ta + a_0 = \left\langle X, G_\mu^{-1}a\right\rangle_\mu + a_0, \tag{5}$$

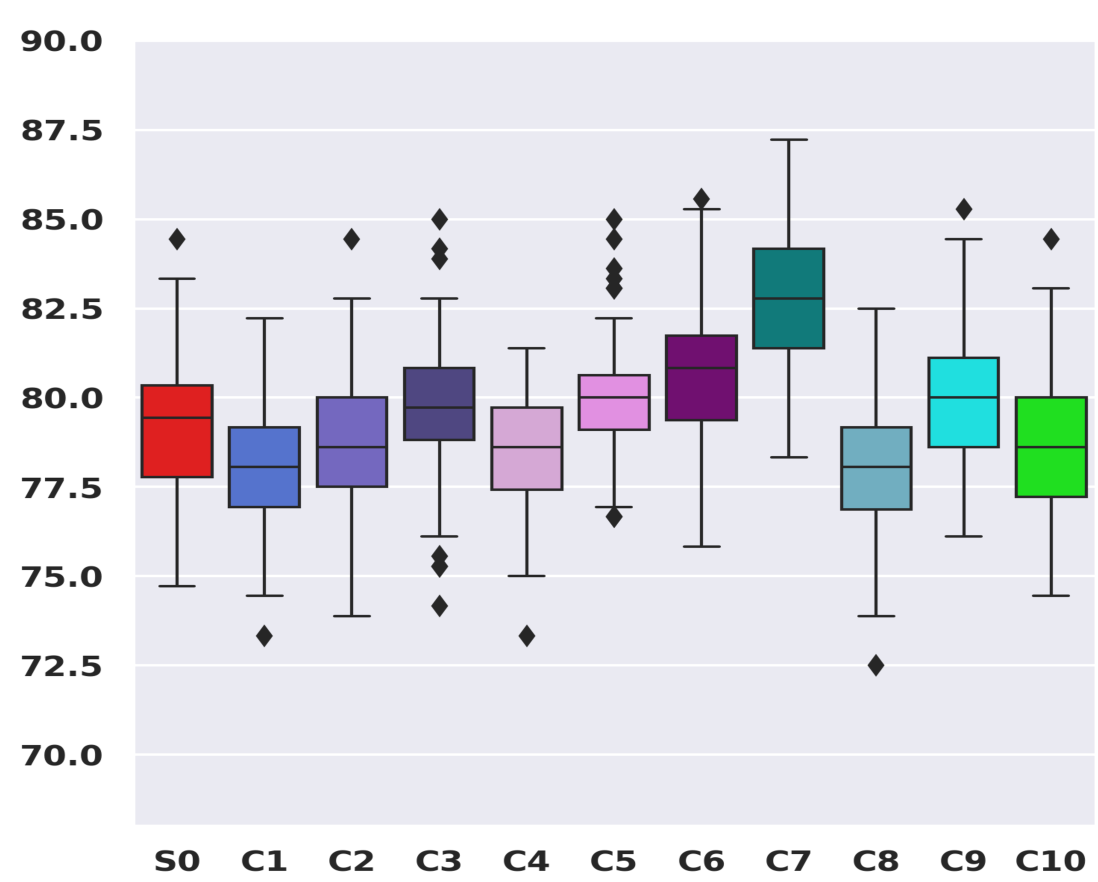
where $a_0 \in \mathbb{R}$, while a and $G_\mu^{-1}a$ are considered as tangents vectors on $T_\mu\mathcal{P}_+(I)$, and G_μ is the matrix induced by Riemanninan metric.

- Let $loss_i$ be the loss function associated with y_i, e.g., $loss_i : \bar{y}_i \mapsto (\bar{y}_i - y_i)^2$.
- **Transport the Model:**

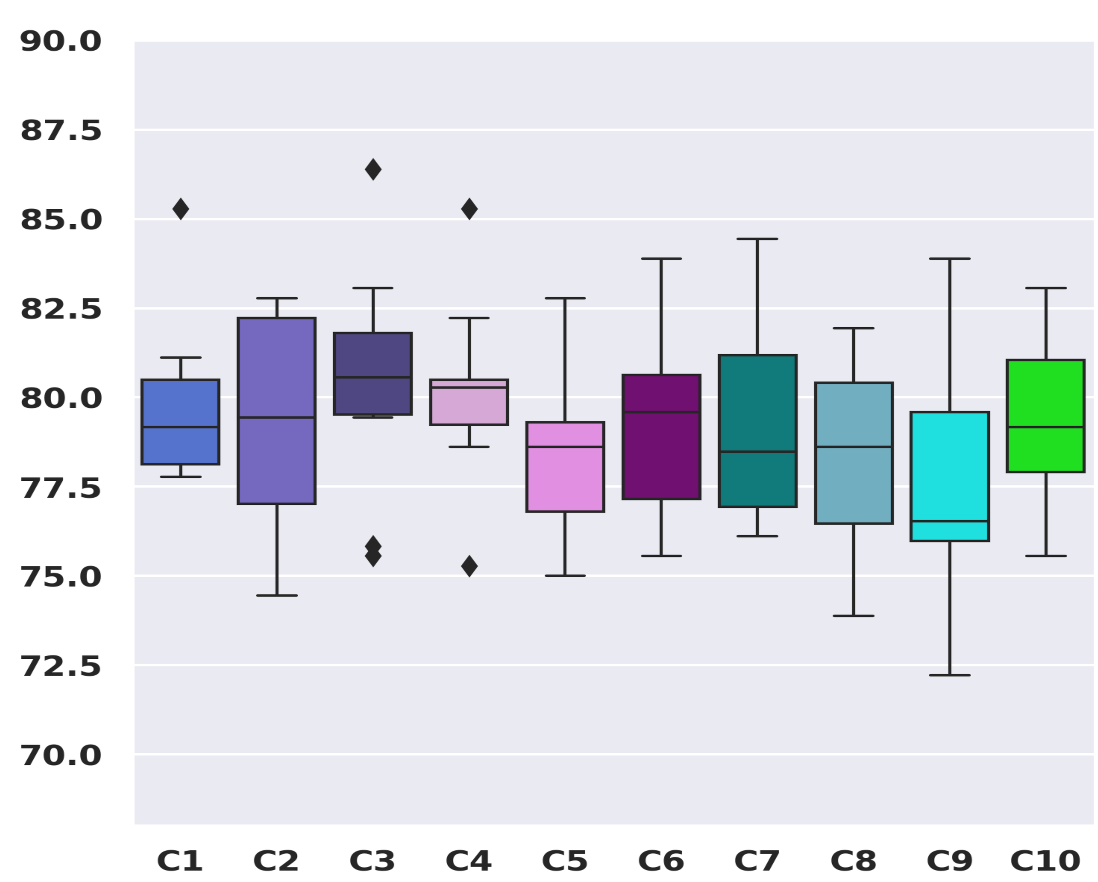
$$(\beta, \beta_0) = \underset{a\in T_\mu\mathcal{P}_+(I), a_0\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^N loss_i(x_i^Ta + a_0) \Rightarrow G_\nu\Gamma G_\mu^{-1}\beta = \underset{b\in T_\nu\mathcal{P}_+(I)}{\operatorname{argmin}} \sum_{i=1}^N loss_i(\Gamma(x_i)^Tb + \beta_0),$$

where Γ is the parallel transport from $T_\mu\mathcal{P}_+(I)$ to $T_\nu\mathcal{P}_+(I)$.

Box-plots of the score for the Logistic Regression Model



The Models learned and test in the same spaces



The transfered models learned from $S0$

$S0$ is the uniform measure, while $C1, \dots, C10$ are chosen randomly.

Acknowledgments

- This work was funded by ANR-IA project.

Tien_Tam.TRAN@doctorant.uca.fr