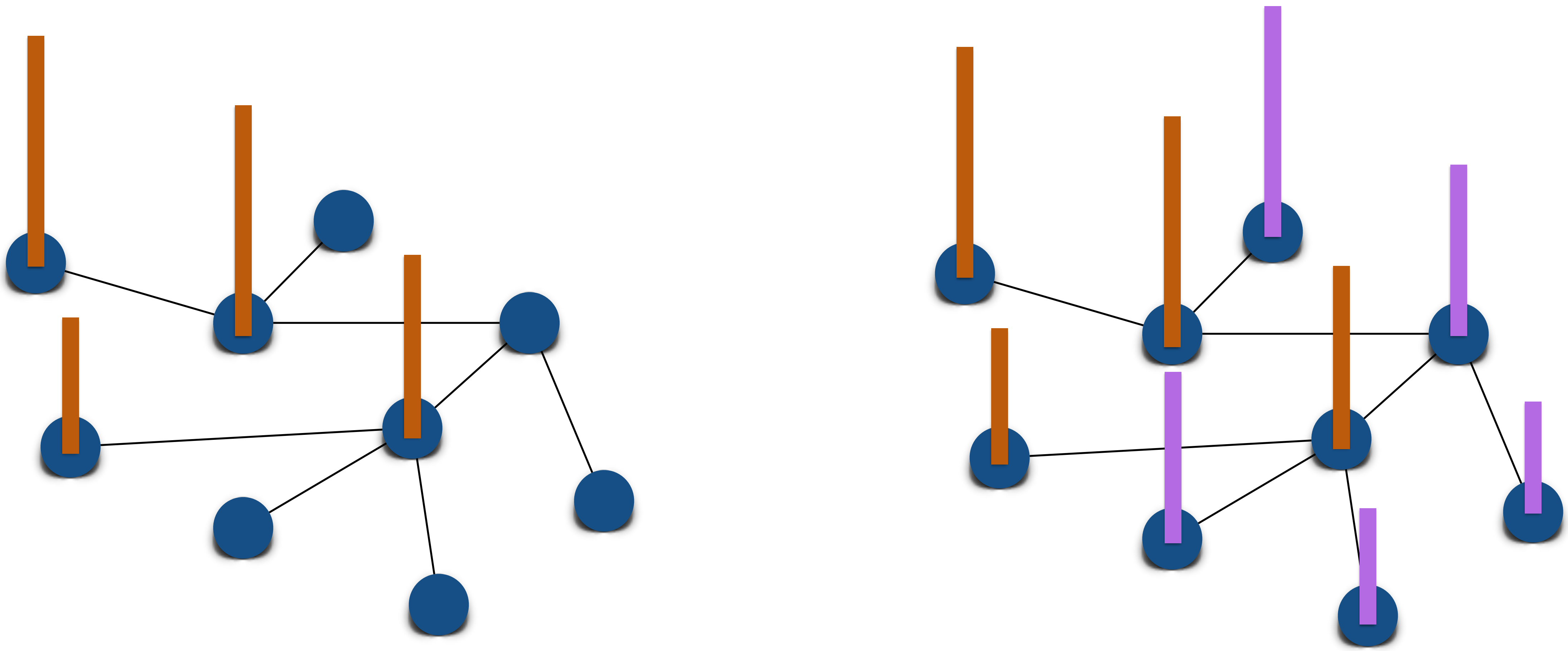
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Transductive learning & RKHS on graphs

Rodrigo Pena
LTS2 Internal Seminar
May 5, 2017

Transductive learning



$f[j], j \in S$ $\longrightarrow$ $f[j], j \in V$

Transductive learning

$$\arg\min_{f \in \mathbb{R}^N} \sum_{j \in S} C(y_j, f[j]) + \mu f^T L f$$

"Manifold assumption"

$$\arg\min_{f \in \mathcal{H}_\phi} \sum_{j \in S} C(y_j, f[j]) + \mu R(f)$$

Function space,
empirical distribution

RKHS

Reproducing

$$f[i] = \langle f, T_i \varphi \rangle_{\mathcal{H}_\varphi}$$

$$f = \sum_{j \in V} \alpha_j T_j \varphi$$

$$\langle f, g \rangle_{\mathcal{H}_\varphi} = \sum_{l \in V} \frac{1}{\varphi(\lambda_l)} \hat{f}(l)^* \hat{g}(l)$$

$$\|f\|_{\mathcal{H}_\varphi} = \alpha^* \Phi \alpha$$

Kernel

Positive-definite

$$\Phi[i,j] = \varphi(L)[i,j] = (T_i \varphi)[j]$$

Hilbert Space

Vectors

+

Completeness

+

Inner product

$$\langle \cdot, \cdot \rangle_{\mathcal{H}_\varphi}$$

A representer theorem

$$\arg\min_{f\in\mathcal{H}_\varphi} \sum_{j\in S}(y_j - f[j])^2 + \mu\|f\|^2_{\mathcal{H}_\varphi}$$

"Ridge regression"

1) Decompose the function space

$$\mathcal{H}_\varphi = \mathcal{H}_S \oplus \mathcal{H}_S^\perp$$

$$\mathcal{H}_S = \left\{f \in \mathcal{H}_\varphi : f = \sum_{k\in S} \alpha_k T_k \varphi\right\}$$

A representer theorem

$$\arg \min_{f_S \in \mathcal{H}_S, f_S^\perp \in \mathcal{H}_S^\perp} \sum_{j\in S} (y_j - (f_S + f_S^\perp)[j])^2 + \mu \| f_S + f_S^\perp \|_{\mathcal{H}_\varphi}^{\mathbf{2}}$$

"Ridge regression"

2) Properties of functions in each space

$$f_S \in \mathcal{H}_S \implies \|f_S\|_{\mathcal{H}} = (\alpha|_S)^T \Phi|_S (\alpha|_S)$$

$$f_S^\perp \in \mathcal{H}_S^\perp \implies \forall k \in S, 0 = \langle f_S^\perp, T_k\varphi \rangle \underset{\text{r.p.}}{=} f_S^\perp[k]$$

A representer theorem

$$\arg\min_{f_S \in \mathcal{H}_S, f_S^{\perp} \in \mathcal{H}_S^{\perp}} \sum_{j \in S} (y_j - f_S[j])^2 + \mu(\|f_S\|^2_{\mathcal{H}_\varphi} + \|f_S^{\perp}\|^2_{\mathcal{H}_\varphi})$$

"Ridge regression"

3) Property of the global minimum

$$\|f_S\|^2_{\mathcal{H}_\varphi} + \|f_S^{\perp}\|^2_{\mathcal{H}_\varphi} \geq \|f_S\|^2_{\mathcal{H}_\varphi}, \forall f_S^{\perp} \in \mathcal{H}_S^{\perp}$$

$$\tilde{f} = \sum_{j \in S} \beta_j T_j \varphi \text{ (!)}$$

A representer theorem

$$\arg\min_{\beta\in\mathbb{R}^{|S|}} \sum_{j\in S}(y_j - (\Phi|_S\beta)[j])^2 + \mu\beta^T\Phi|_S\beta$$

"Ridge regression"

4) Algorithm

1. Compute optimal coefficients: $\tilde{\beta} = (\Phi|_S + \mu I_{|S|})^{-1} y$
2. Compute the regression: $\tilde{f} = \varphi(L)\left\{\sum_{j\in S}\tilde{\beta}_j\delta_j\right\}$

Can we do it for other problems?

$$\arg\min_{f \in \mathcal{H}_\varphi} \sum_{j \in S} C(y_j, f[j]) + \mu \|Rf\|_1$$

"Generalized TV"

Informal preliminary results (based on [Unser et al. 2016]):

$$f[\cdot] = \sum_{k \leq M} a_k \Phi \rho_{(R\Phi)}[\cdot, j_k] + \sum_{n=1}^{N_0} b_n \Phi p_n[\cdot]$$

If R = g(L):

$$f[\cdot] = \sum_{k \leq M} a_k \rho_{g(L)}[\cdot, j_k] + \sum_{n=1}^{N_0} b_n \Phi p_n[\cdot]$$

So what?

- Knowing form of the solution simplifies the search space.
- Attaching a probability distribution to the data:
  - Consistency.
  - Error as function of ratio unlabeled/labeled.
  - Where to sample / how many sample labeled data for a given accuracy level.

Thank you

Q & A