Graphs in Machine Learning

Michal Valko

INRIA Lille - Nord Europe, France

Partially based on material by: Branislav Kveton, Partha Niyogi

Last Lecture

- Inductive and transductive semi-supervised learning
- Manifold regularization
- Theory of Laplacian-based manifold methods
- Transductive learning stability based bounds
- SSL Learnability
- Online Semi-Supervised Learning
- Online incremental k-centers

This Lecture

- Analysis of online SSL
- When does graph-based SSL provably help?
- Scaling harmonic functions to millions of samples
- Online decision-making on graphs
- Graph Bandits

Previous Lab Session

- 24. 2. 2015 by Daniele.Calandriello@inria.fr
- Content
  - Semi-supervised learning
  - Graph quantization
  - Online face recognizer
  - Install OpenCV (if you still have problems contact Daniele)
- Short written report (graded, each lab around 5% of grade)
- Daniele: Use CHFS instead of SHFS (material updated)
- Questions to Daniele.Calandriello@inria.fr
- *Deadline:* 10. 3. 2015
- http://researchers.lille.inria.fr/~calandri/ta/graphs/td2_handout.pdf

Next Lab Session

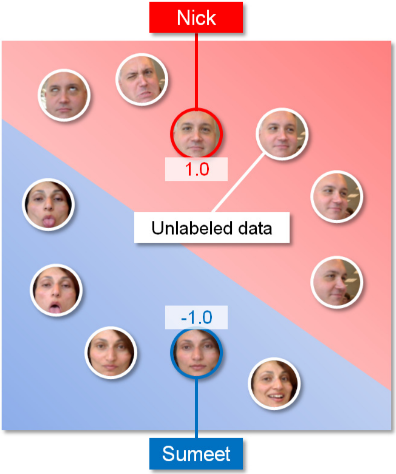
- 10. 3. 2015 by Daniele.Calandriello@inria.fr
- Content
  - GraphLab
  - Large-Scale Graph Learning
- AR1: Get the GraphLab license
- AR2: Install VirtualBox
- AR3: Download virtual machine (online very soon)
- Short written report (graded, each lab around 5% of grade)
- Questions to Daniele.Calandriello@inria.fr
- *Deadline:* 24. 3. 2015
- http://researchers.lille.inria.fr/~calandri/ta/graphs/td3_handout.pdf

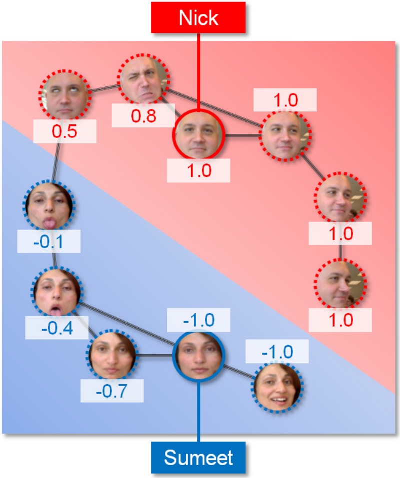
Final Class projects

- preferred option: you come up with the topic
- details and list of suggested topics on the class website
- theory/implementation/review or a combination
- one or two people per project (exceptionally three)
- grade: report + short presentation of the team
- Deadlines
  - taking projects 17. 3. 2015 (recommended 10. 3. 2015)
  - 11. 4. 2015 final report
  - 13. 4. 2015 afternoon (tentative), presentation in class

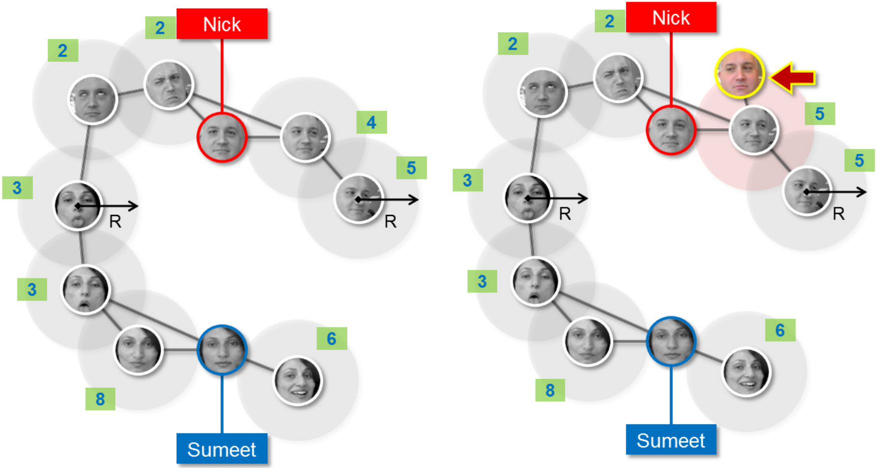
http://researchers.lille.inria.fr/~valko/hp/serve?from=slides&what=mvaprojects

Where we left off: Online SSL with Graphs

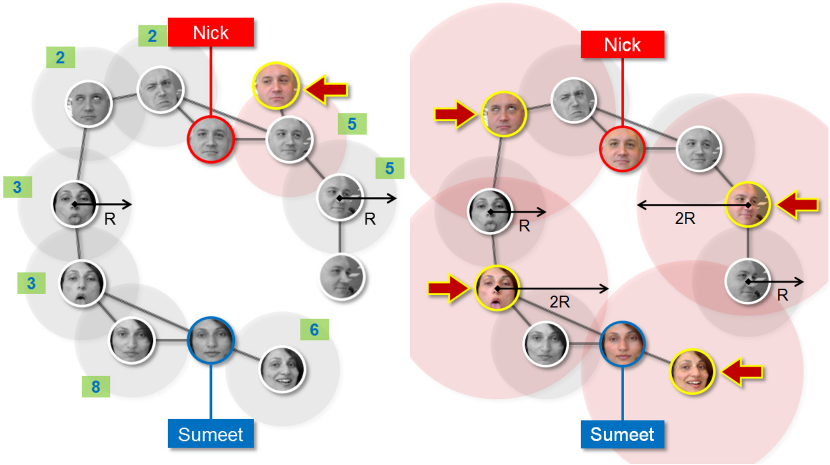




Where we left off: Online SSL with Graphs



Where we left off: Online SSL with Graphs



Where we left off: Online SSL with Graphs

Video examples

http://www.bkveton.com/videos/Coffee.mp4

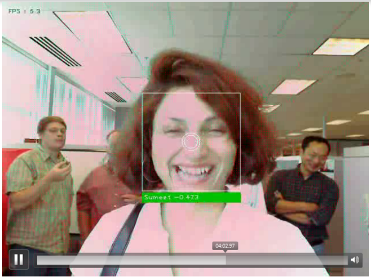
http://www.bkveton.com/videos/Ad.mp4

http://researchers.lille.inria.fr/~valko/hp/serve.php?what=publications/kveton2009nipsdemo.adaptation.mov

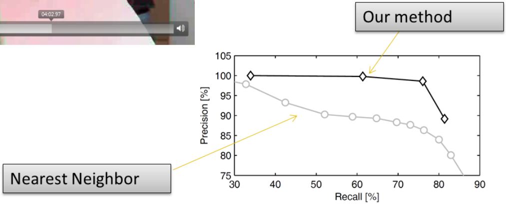
http://researchers.lille.inria.fr/~valko/hp/serve.php?what=publications/kveton2009nipsdemo.officespace.mov

http://bcove.me/a2derjeh

SSL with Graphs: Some experimental results

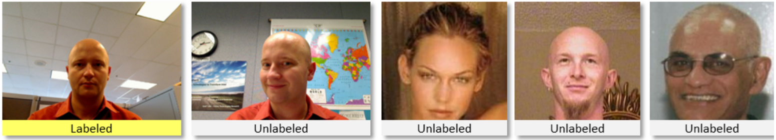


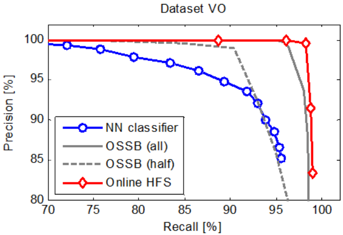
- 8 people classification
- Making funny faces
- 4 faces/person are labeled

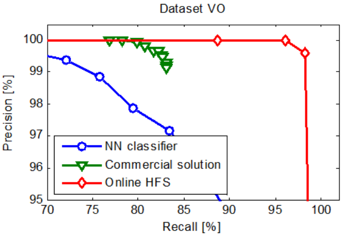


SSL with Graphs: Some experimental results

- One person moves among various indoor locations
- 4 labeled examples of a person in the cubicle







Online HFS outperforms OSSB (even when the weak learners are chosen using future data)

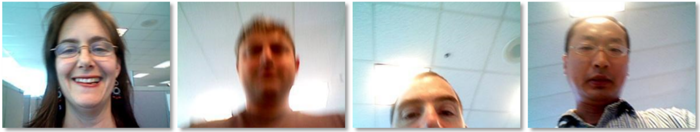
Online HFS yields better results than a commercial solution at 20% of the computational cost

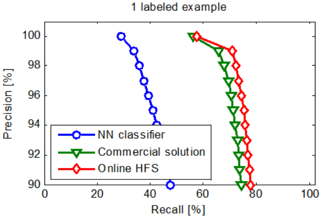
SSL with Graphs: Some experimental results

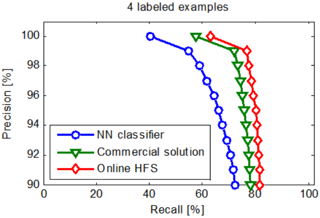
- **Logging in** with faces instead of password
- Able to **learn** and improve

SSL with Graphs: Some experimental results

- 16 people log twice into a tablet PC at 10 locations







Online HFS yields better results than a commercial solution at 20% of the computational cost

Online SSL with Graphs: Analysis

What can we guarantee?

Three sources of error

- generalization error — if all data: $(\ell_t^\star - y_t)^2$
- online error — data only incrementally: $(\ell_t^{\mathrm{o}}[t] - \ell_t^\star)^2$
- quantization error — memory limitation: $(\ell_t^{\mathrm{q}}[t] - \ell_t^{\mathrm{o}}[t])^2$

All together:

$$\frac{1}{n}\sum_{t=1}^{n}(\ell_t^{\mathrm{q}}[t]-y_t)^2 \le \frac{9}{2n}\sum_{t=1}^{n}(\ell_t^\star-y_t)^2 + \frac{9}{2n}\sum_{t=1}^{n}(\ell_t^{\mathrm{o}}[t]-\ell_t^\star)^2 + \frac{9}{2n}\sum_{t=1}^{n}(\ell_t^{\mathrm{q}}[t]-\ell_t^{\mathrm{o}}[t])^2$$

Since for any a, b, c, $d \in [-1, 1]$:

$$(a-b)^2 \le \frac{9}{2}\left[(a-c)^2 + (c-d)^2 + (d-b)^2\right]$$

Online SSL with Graphs: Analysis

Bounding transduction error $(\ell_t^\star - y_t)^2$

If all labeled examples l are i.i.d., $c_l = 1$ and $c_l \gg c_u$, then

$$R(\ell^\star) \;\leq\; \widehat{R}(\ell^\star) + \underbrace{\beta + \sqrt{\frac{2\ln(2/\delta)}{n_l}}(n_l\beta + 4)}_{\text{transductive error } \Delta_T(\beta, n_l, \delta)}$$

$$\beta \;\leq\; 2\left[\frac{\sqrt{2}}{\gamma_g + 1} + \sqrt{2n_l}\frac{1-\sqrt{c_u}}{\sqrt{c_u}}\frac{\lambda_M(\mathbf{L}) + \gamma_g}{\gamma_g^2 + 1}\right]$$

holds with the probability of $1-\delta$, where

$$R(\ell^\star) = \frac{1}{n}\sum_t (\ell_t^\star - y_t)^2 \quad \text{and} \quad \widehat{R}(\ell^\star) = \frac{1}{n_l}\sum_{t\in l}(\ell_t^\star - y_t)^2$$

How should we set γ_g?

Want $\Delta_T(\beta, n_l, \delta) = o(1) \to \beta = o\left(n_l^{-1/2}\right)$

$\to \gamma_g = \Omega\left(n_l^{1+\alpha}\right)$ for any $\alpha > 0$.

Online SSL with Graphs: Analysis

Bounding online error $(\ell^{\mathrm{o}}_t[t] - \ell^\star_t)^2$

Idea: If $\mathbf{L}$ and $\mathbf{L}^{\mathrm{o}}$ are regularized, then HFSs get closer together.

since they get closer to zero

Recall $\boldsymbol{\ell} = (\mathbf{C}^{-1}\mathbf{Q} + \mathbf{I})^{-1}\mathbf{y}$, where $\mathbf{Q} = \mathbf{L} + \gamma_g \mathbf{I}$

and also $\mathbf{v} \in \mathbb{R}^{n\times 1}$, $\lambda_m(A)\|\mathbf{v}\|_2 \leq \|A\mathbf{v}\|_2 \leq \lambda_M(A)\|\mathbf{v}\|_2$

$$\|\boldsymbol{\ell}\|_2 \leq \frac{\|\mathbf{y}\|_2}{\lambda_m(\mathbf{C}^{-1}\mathbf{Q} + I)} = \frac{\|\mathbf{y}\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C})} + 1} \leq \frac{\sqrt{n_l}}{\gamma_g + 1}$$

Difference between offline and online solutions:

$$(\ell^{\mathrm{o}}_t[t] - \ell^\star_t)^2 \leq \|\boldsymbol{\ell}^{\mathrm{o}}[t] - \boldsymbol{\ell}^\star\|_\infty^2 \leq \|\boldsymbol{\ell}^{\mathrm{o}}[t] - \boldsymbol{\ell}^\star\|_2^2 \leq \left(\frac{2\sqrt{n_l}}{\gamma_g + 1}\right)^2$$

Again, how should we set γ_g? If we want $O\left(n^{-1/2}\right)$? Then $\gamma_g = \Omega\left(n^{1/4}\right)$

Online SSL with Graphs: Analysis

Bounding quantization error $(\ell_t^{\mathrm{q}}[t]-\ell_t^{\mathrm{o}}[t])^2$

How are the quantized and full solution different?

$$\ell^\star = \min_{\boldsymbol{\ell}\in\mathbb{R}^n} (\boldsymbol{\ell}-\mathbf{y})^\mathsf{T}\mathbf{C}(\boldsymbol{\ell}-\mathbf{y}) + \boldsymbol{\ell}^\mathsf{T}\mathbf{Q}\boldsymbol{\ell}$$

In $\mathbf{Q}$! $\mathbf{Q}^{\mathrm{o}}$ (online) vs. $\mathbf{Q}^{\mathrm{q}}$ (quantized)

We have: $\boldsymbol{\ell}^{\mathrm{o}} = (\mathbf{C}^{-1}\mathbf{Q}^{\mathrm{o}}+\mathbf{I})^{-1}\mathbf{y}$ vs. $\boldsymbol{\ell}^{\mathrm{q}} = (\mathbf{C}^{-1}\mathbf{Q}^{\mathrm{q}}+\mathbf{I})^{-1}\mathbf{y}$

Let $\mathbf{Z}^{\mathrm{q}} = \mathbf{C}^{-1}\mathbf{Q}^{\mathrm{q}}+\mathbf{I}$ and $\mathbf{Z}^{\mathrm{o}} = \mathbf{C}^{-1}\mathbf{Q}^{\mathrm{o}}+\mathbf{I}$.

$$\begin{aligned}\boldsymbol{\ell}^{\mathrm{q}}-\boldsymbol{\ell}^{\mathrm{o}} &= (\mathbf{Z}^{\mathrm{q}})^{-1}\mathbf{y}-(\mathbf{Z}^{\mathrm{o}})^{-1}\mathbf{y} = (\mathbf{Z}^{\mathrm{q}}\mathbf{Z}^{\mathrm{o}})^{-1}(\mathbf{Z}^{\mathrm{o}}-\mathbf{Z}^{\mathrm{q}})\mathbf{y}\\ &= (\mathbf{Z}^{\mathrm{q}}\mathbf{Z}^{\mathrm{o}})^{-1}\mathbf{C}^{-1}(\mathbf{Q}^{\mathrm{o}}-\mathbf{Q}^{\mathrm{q}})\mathbf{y}\end{aligned}$$

Online SSL with Graphs: Analysis

Bounding quantization error $(\ell_t^{\mathrm{q}}[t] - \ell_t^{\mathrm{o}}[t])^2$

$$\boldsymbol{\ell}^{\mathrm{q}} - \boldsymbol{\ell}^{\mathrm{o}} = (\mathbf{Z}^{\mathrm{q}})^{-1}\mathbf{y} - (\mathbf{Z}^{\mathrm{o}})^{-1}\mathbf{y} = (\mathbf{Z}^{\mathrm{q}}\mathbf{Z}^{\mathrm{o}})^{-1}(\mathbf{Z}^{\mathrm{o}} - \mathbf{Z}^{\mathrm{q}})\mathbf{y}$$
$$= (\mathbf{Z}^{\mathrm{q}}\mathbf{Z}^{\mathrm{o}})^{-1}\mathbf{C}^{-1}(\mathbf{Q}^{\mathrm{o}} - \mathbf{Q}^{\mathrm{q}})\mathbf{y}$$

$$\|\boldsymbol{\ell}^{\mathrm{q}} - \boldsymbol{\ell}^{\mathrm{o}}\|_2 \leq \frac{\lambda_M(\mathbf{C}^{-1})\|(\mathbf{Q}^{\mathrm{q}} - \mathbf{Q}^{\mathrm{o}})\mathbf{y}\|_2}{\lambda_m(\mathbf{Z}^{\mathrm{q}})\lambda_m(\mathbf{Z}^{\mathrm{o}})}$$

$||\cdot||_F$ and $||\cdot||_2$ are compatible and $\mathbf{y}_i$ is zero when unlabeled:

$$\|(\mathbf{Q}^{\mathrm{q}} - \mathbf{Q}^{\mathrm{o}})\mathbf{y}\|_2 \leq \|\mathbf{Q}^{\mathrm{q}} - \mathbf{Q}^{\mathrm{o}}\|_F \cdot \|\mathbf{y}\|_2 \leq \sqrt{n_l}\|\mathbf{Q}^{\mathrm{q}} - \mathbf{Q}^{\mathrm{o}}\|_F$$

Furthermore, $\lambda_m(\mathbf{Z}^{\mathrm{o}}) \geq \dfrac{\lambda_m(\mathbf{Q}^{\mathrm{o}})}{\lambda_M(\mathbf{C})} + 1 \geq \gamma_g$ and $\lambda_M(\mathbf{C}^{-1}) \leq c_u^{-1}$

We get $\|\boldsymbol{\ell}^{\mathrm{q}} - \boldsymbol{\ell}^{\mathrm{o}}\|_2 \leq \dfrac{\sqrt{n_l}}{c_u\gamma_g^2}\|\mathbf{Q}^{\mathrm{q}} - \mathbf{Q}^{\mathrm{o}}\|_F$

Online SSL with Graphs: Analysis

Bounding **quantization error** $(\ell_t^{\mathrm{q}}[t] - \ell_t^{\mathrm{o}}[t])^2$

The quantization error depends on $\|\mathbf{Q}^{\mathrm{q}} - \mathbf{Q}^{\mathrm{o}}\|_F = \|\mathbf{L}^{\mathrm{q}} - \mathbf{L}^{\mathrm{o}}\|_F$.

When can we keep $\|\mathbf{L}^{\mathrm{q}} - \mathbf{L}^{\mathrm{o}}\|_F$ under control?

Charikar guaranteed **distortion** error of $Rm/(m-1)$

For what kind of data $\{\mathbf{x}_i\}_{i=1,\ldots,n}$ is the distortion small?

Assume manifold $\mathcal{M}$

- all $\{\mathbf{x}_i\}_{i\geq 1}$ lie on a smooth s-dimensional compact $\mathcal{M}$
- with boundary of bounded geometry Def. 11 of Hein [HAL07]
  - should not intersect itself
  - should not fold back onto itself
  - has finite volume V
  - has finite surface area A

Online SSL with Graphs: Analysis

Bounding quantization error $(\ell_t^{\mathrm{q}}[t]-\ell_t^{\mathrm{o}}[t])^2$

Bounding $\|\mathbf{L}^{\mathrm{q}} - \mathbf{L}^{\mathrm{o}}\|_F$ when $\mathbf{x}_i \in \mathcal{M}$

Consider k-sphere packing of radius r with centers contained in $\mathcal{M}$.

What is the maximum volume of this packing?

$kc_s r^s \leq V + Ac_{\mathcal{M}} r$ with $c_s, c_{\mathcal{M}}$ depending on dimension and $\mathcal{M}$.

If k is large $\rightarrow r <$ **injectivity radius** of $\mathcal{M}$ [HAL07] and $r < 1$:

$$r < \left(\left(V + Ac_{\mathcal{M}}\right) / \left(kc_s\right)\right)^{1/s} = \mathcal{O}\left(k^{-1/s}\right)$$

r-packing is a $2r$-covering:

$$Rm/(m-1) = \max_{i=1,\dots,n} \|\mathbf{x}_i - \mathbf{c}\|_2 = 2(1+\varepsilon)\mathcal{O}\left(k^{-1/s}\right) = \mathcal{O}\left(k^{-1/s}\right)$$

But what about $\|\mathbf{L}^{\mathrm{q}} - \mathbf{L}^{\mathrm{o}}\|_F$?

Online SSL with Graphs: Analysis

Bounding quantization error $(\ell_t^{\mathrm{q}}[t] - \ell_t^{\mathrm{o}}[t])^2$

If similarity is M-Lipschitz, $\mathbf{L}$ is normalized, $c_{ij}^{\mathrm{o}} = \sqrt{\mathbf{D}_{ii}^{\mathrm{o}}\mathbf{D}_{jj}^{\mathrm{o}}} > c_{min}n$:

$$
\begin{aligned}
\mathbf{L}_{ij}^{\mathrm{q}} - \mathbf{L}_{ij}^{\mathrm{o}} &= \frac{\mathbf{W}_{ij}^{\mathrm{q}}}{c_{ij}^{\mathrm{q}}} - \frac{\mathbf{W}_{ij}^{\mathrm{o}}}{c_{ij}^{\mathrm{o}}} \\
&\leq \frac{\mathbf{W}_{ij}^{\mathrm{q}} - \mathbf{W}_{ij}^{\mathrm{o}}}{c_{ij}^{\mathrm{q}}} + \frac{\mathbf{W}_{ij}^{\mathrm{q}}(c_{ij}^{\mathrm{q}} - c_{ij}^{\mathrm{o}})}{c_{ij}^{\mathrm{o}} c_{ij}^{\mathrm{q}}} \\
&\leq \frac{4MRm}{(m-1)c_{min}n} + \frac{4M(nMRm)}{((m-1)c_{min}n)^2} \\
&= O\left(\frac{R}{n}\right)
\end{aligned}
$$

Finally, $\|\mathbf{L}^{\mathrm{q}} - \mathbf{L}^{\mathrm{o}}\|_F^2 \leq n^2\mathcal{O}(R^2/n^2) = \mathcal{O}(k^{-2/s})$.

Are the assumptions reasonable?

Online SSL with Graphs: Analysis

Bounding quantization error $(\ell_t^{\mathrm{q}}[t] - \ell_t^{\mathrm{o}}[t])^2$

We showed $\|\mathbf{L}^{\mathrm{q}} - \mathbf{L}^{\mathrm{o}}\|_F^2 \leq n^2\mathcal{O}(R^2/n^2) = \mathcal{O}(k^{-2/s}) = \mathcal{O}(1)$.

$$\frac{1}{n}\sum_{t=1}^{n}(\ell_t^{\mathrm{q}}[t] - \ell_t^{\mathrm{o}}[t])^2 \leq \frac{n_l}{c_u^2\gamma_g^4}\|\mathbf{L}^{\mathrm{q}} - \mathbf{L}^{\mathrm{o}}\|_F^2 \leq \frac{n_l}{c_u^2\gamma_g^4}$$

This converges to zero at the rate of $\mathcal{O}(n^{-1/2})$ with $\gamma_g = \Omega(n^{1/8})$.

With properly setting γ_g, e.g., $\gamma_g = \Omega(n^{1/8})$, we can have:

$$\frac{1}{n}\sum_{t=1}^{n}(\ell_t^{\mathrm{q}}[t] - y_t)^2 = \mathcal{O}\left(n^{-1/2}\right)$$

What does that mean?

SSL with Graphs: What is behind it?

Why and when it helps?

Can we guarantee benefit of SSL over SSL?

Are there cases when **manifold** SSL is provably helpful?

Say $\mathcal{X}$ is supported on manifold $\mathcal{M}$. Compare two cases:

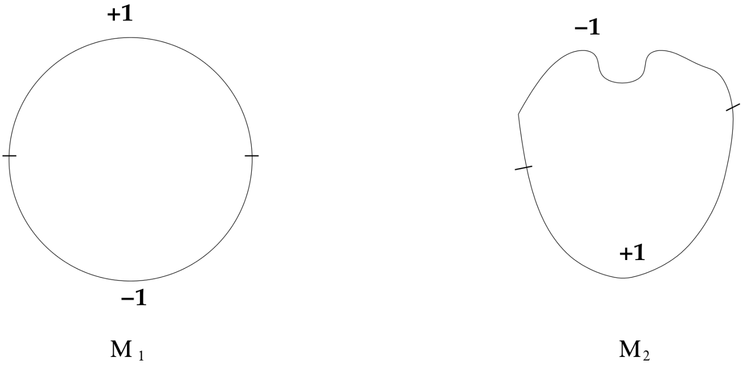
- SL: does not know about $\mathcal{M}$ and only knows $(\mathbf{x}_i, y_i)$
- SSL: perfect knowledge of $\mathcal{M} \equiv$ humongous amounts of $\mathbf{x}_i$

http://people.cs.uchicago.edu/~niyogi/papersps/ssminimax2.pdf

SSL with Graphs: What is behind it?

Set of learning problems:

$$\mathcal{P} = \cup_{\mathcal{M}} \mathcal{P}_{\mathcal{M}} = \{p \in \mathcal{P} | p_{\mathcal{X}} \text{ is uniform on } \mathcal{M}\}$$



SSL with Graphs: What is behind it?

Set of problems $\mathcal{P} = \cup_{\mathcal{M}} \mathcal{P}_{\mathcal{M}} = \{p \in \mathcal{P} | p_{\mathcal{X}}$ is uniform on $\mathcal{M}\}$
Regression function $m_p = \mathbb{E}[y \,|x]$ when $x \in \mathcal{M}$
Algorithm A and **labeled examples** $\bar{z} = \{z_i\}_{i=1}^n = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$
Minimax rate

$$R(n, \mathcal{P}) = \inf_A \sup_{p \in \mathcal{P}} \mathbb{E}_{\bar{z}} \left[\|A(\bar{z}) - m_p\|_{L^2(p_{\mathbf{x}})} \right]$$

Since $\mathcal{P} = \cup_{\mathcal{M}} \mathcal{P}_{\mathcal{M}}$

$$R(n, \mathcal{P}) = \inf_A \sup_{\mathcal{M}} \sup_{p \in \mathcal{P}_{\mathcal{M}}} \mathbb{E}_{\bar{z}} \left[\|A(\bar{z}) - m_p\|_{L^2(p_{\mathbf{x}})} \right]$$

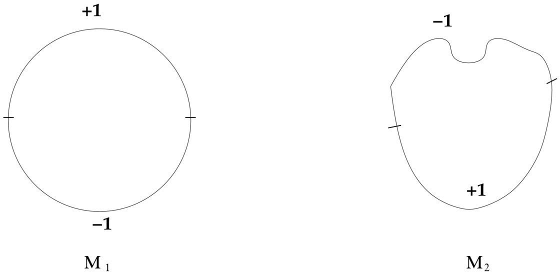
(SSL) When A is allowed to know $\mathcal{M}$

$$Q(n, \mathcal{P}) = \sup_{\mathcal{M}} \inf_A \sup_{p \in \mathcal{P}_{\mathcal{M}}} \mathbb{E}_{\bar{z}} \left[\|A(\bar{z}) - m_p\|_{L^2(p_{\mathbf{x}})} \right]$$

In which cases there is a gap between $Q(n, \mathcal{P})$ and $R(n, \mathcal{P})$?

SSL with Graphs: What is behind it?

Hypothesis space $\mathcal{H}$: half of the circle as $+1$ and the rest as -1



Case 1: $\mathcal{M}$ is known to the learner $(\mathcal{H}_{\mathcal{M}})$

What is a VC dimension of $\mathcal{H}_{\mathcal{M}}$? 2

$$\text{Optimal rate } Q(n, \mathcal{P}) \leq 2\sqrt{\frac{3 \log n}{n}}$$

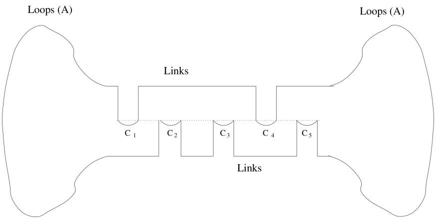
SSL with Graphs: What is behind it?

Case 2: $\mathcal{M}$ is **unknown** to the learner

$$R(n, \mathcal{P}) = \inf_{A} \sup_{p \in \mathcal{P}} \mathbb{E}_{\bar{z}} \left[\| A(\bar{z}) - m_p \|_{L^2(p_{\mathbf{X}})} \right] = \Omega(1)$$

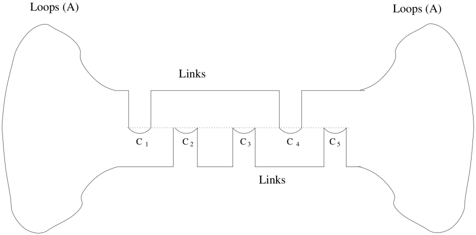
We consider 2^d manifolds of the form

$$\mathcal{M} = \text{Loops} \cup \text{Links} \cup C \text{ where } C = \cup_{i=1}^{d} C_i$$



Main idea: d segments in C, $d - l$ with no data, 2^l possible choices for labels, which helps us to lower bound $\|A(\bar{z}) - m_p\|_{L^2(p_{\mathbf{X}})}$

SSL with Graphs: What is behind it?



Knowing the manifold helps

- C_1 and C_4 are close
- C_1 and C_3 are far
- we also need: **target function varies smoothly**
- altogether: **closeness on manifold** $\rightarrow$ **similarity in labels**

SSL with Graphs: What is behind it?

What does it mean to **know** $\mathcal{M}$?

Different degrees of knowing $\mathcal{M}$

- set membership oracle: $\mathbf{x} \stackrel{?}{\in} \mathcal{M}$
- approximate oracle
- knowing the harmonic functions on $\mathcal{M}$
- knowing the Laplacian $\mathcal{L}_{\mathcal{M}}$
- knowing eigenvalues and eigen*functions*
- topological invariants, e.g., dimension
- metric information: geodesic distance

Michal Valko
michal.valko@inria.fr
sequel.lille.inria.fr
SequeL – INRIA Lille

MVA 2014/2015