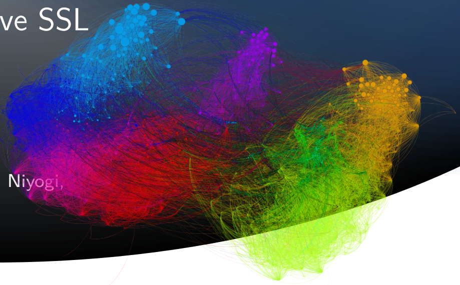
Graphs in Machine Learning

Inductive Generalization Bounds

Theoretical Guarantees for Inductive SSL



Michal Valko

Inria & ENS Paris-Saclay, MVA

Partially based on material by: Mikhail Belkin, Partha Niyogi, [illegible] Chapelle, Bernhard Schölkopf

Inductive Generalization Bounds

We may want to bound the **risk**

$$R_P(f) = \mathbb{E}_{P(\mathbf{x})}\left[\mathcal{L}\left(f\left(\mathbf{x}\right), y\left(\mathbf{x}\right)\right)\right]$$

Inductive Generalization Bounds

We may want to bound the **risk**

$$R_P(f) = \mathbb{E}_{P(\mathbf{x})}\left[\mathcal{L}\left(f\left(\mathbf{x}\right), y\left(\mathbf{x}\right)\right)\right]$$

for some **loss**, e.g., 0/1 loss

$$\mathcal{L}(y', y) = \mathbb{1}\{\mathrm{sgn}(y') \neq y\}$$

Inductive Generalization Bounds

We may want to bound the **risk**

$$R_P(f) = \mathbb{E}_{P(\mathbf{x})}\left[\mathcal{L}\left(f\left(\mathbf{x}\right), y\left(\mathbf{x}\right)\right)\right]$$

for some **loss**, e.g., 0/1 loss

$$\mathcal{L}(y', y) = \mathbb{1}\{\mathrm{sgn}(y') \neq y\}$$

$$R_P(f) \leq \hat{R}_P(f) + \textbf{error terms}$$

Inductive Generalization Bounds

Using classical SLT tools (Equations 3.15 and 3.24 **vapnik1995nature**), with probability $1 - \eta$

$$R_P(f) \leq \frac{1}{N} \sum_i \mathcal{L}(f(\mathbf{x}_i), y_i) + \Delta_I(h, N, \eta).$$

Inductive Generalization Bounds

Using classical SLT tools (Equations 3.15 and 3.24 **vapnik1995nature**), with probability $1 - \eta$

$$R_P(f) \leq \frac{1}{N} \sum_i \mathcal{L}(f(\mathbf{x}_i), y_i) + \Delta_I(h, N, \eta).$$

$N \equiv$ number of samples

Inductive Generalization Bounds

Using classical SLT tools (Equations 3.15 and 3.24 **vapnik1995nature**), with probability $1 - \eta$

$$R_P(f) \leq \frac{1}{N} \sum_i \mathcal{L}(f(\mathbf{x}_i), y_i) + \Delta_I(h, N, \eta).$$

$N \equiv$ number of samples , $h \equiv$ VC dimension of the class

Inductive Generalization Bounds

Using classical SLT tools (Equations 3.15 and 3.24 **vapnik1995nature**), with probability $1 - \eta$

$$R_P(f) \leq \frac{1}{N} \sum_i \mathcal{L}(f(\mathbf{x}_i), y_i) + \Delta_I(h, N, \eta).$$

$N \equiv$ number of samples , $h \equiv$ VC dimension of the class

$$\Delta_I(h, N, \eta) = \sqrt{\frac{h(\ln(2N/h) + 1) - \ln(\eta/4)}{N}}$$

Inductive Generalization Bounds

Using classical SLT tools (Equations 3.15 and 3.24 **vapnik1995nature**), with probability $1 - \eta$

$$R_P(f) \leq \frac{1}{N} \sum_i \mathcal{L}(f(\mathbf{x}_i), y_i) + \Delta_I(h, N, \eta).$$

$N \equiv$ number of samples , $h \equiv$ VC dimension of the class

$$\Delta_I(h, N, \eta) = \sqrt{\frac{h(\ln(2N/h) + 1) - \ln(\eta/4)}{N}}$$

How to bound $\mathcal{L}(f(\mathbf{x}_i), y_i)$?

Inductive Generalization Bounds

For any $y_i \in \{-1, 1\}$ and $\ell_i^\star$

$$\frac{1}{N}\sum_i \mathcal{L}(f(\mathbf{x}_i), y_i) \leq \frac{1}{N}\sum_i \mathcal{L}(f(\mathbf{x}_i), \mathrm{sgn}(\ell_i^\star)) + \frac{1}{N}\sum_i (\ell_i^\star - y_i)^2$$

Inductive Generalization Bounds

For any $y_i \in \{-1, 1\}$ and $\ell_i^\star$

$$\frac{1}{N}\sum_i \mathcal{L}(f(\mathbf{x}_i), y_i) \le \frac{1}{N}\sum_i \mathcal{L}(f(\mathbf{x}_i), \operatorname{sgn}(\ell_i^\star)) + \frac{1}{N}\sum_i (\ell_i^\star - y_i)^2$$

$$\le \left(\frac{1}{N}\sum_i \mathcal{L}(f(\mathbf{x}_i), \operatorname{sgn}(\ell_i^\star))\right) + \widehat{R}_P(\boldsymbol{\ell}^\star) + \Delta_T(\beta, n_l, \delta)$$

Inductive Generalization Bounds

Combining inductive + transductive error

With probability $1 - (\eta + \delta)$.

$$R_P(f) \;\leq\; \frac{1}{N}\sum_i \mathcal{L}(f(\mathbf{x}_i), \mathrm{sgn}(\ell_i^\star)) + \widehat{R}_P(\boldsymbol{\ell}^\star) + \Delta_T(\beta, n_l, \delta) + \Delta_I(h, N, \eta)$$

Inductive Generalization Bounds

Combining inductive + transductive error

With probability $1-(\eta+\delta)$.

$$R_P(f) \;\leq\; \frac{1}{N}\sum_i \mathcal{L}(f(\mathbf{x}_i), \operatorname{sgn}(\ell_i^\star)) + \widehat{R}_P(\boldsymbol{\ell}^\star) + \Delta_T(\beta, n_l, \delta) + \Delta_I(h, N, \eta)$$

We need to account for ε. With probability $1-(\eta+\delta)$.

$$R_P(f) \leq \frac{1}{N}\sum_{i:|\ell_i^\star|\geq\varepsilon} \mathcal{L}(f(\mathbf{x}_i), \operatorname{sgn}(\ell_i^\star)) + \frac{2\varepsilon n_\varepsilon}{N} + \widehat{R}_P(\boldsymbol{\ell}^\star) + \Delta_T(\beta, n_l, \delta) + \Delta_I(h, N, \eta)$$

Inductive Generalization Bounds

Combining inductive + transductive error

With probability $1-(\eta+\delta)$.

$$R_P(f) \le \frac{1}{N}\sum_i \mathcal{L}(f(\mathbf{x}_i), \operatorname{sgn}(\ell_i^\star)) + \widehat{R}_P(\boldsymbol{\ell}^\star) + \Delta_T(\beta, n_l, \delta) + \Delta_I(h, N, \eta)$$

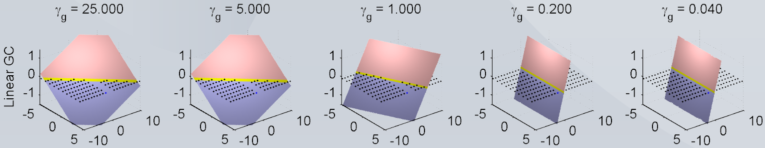
We need to account for ε. With probability $1-(\eta+\delta)$.

$$R_P(f) \le \frac{1}{N}\sum_{i:|\ell_i^\star|\ge\varepsilon} \mathcal{L}(f(\mathbf{x}_i), \operatorname{sgn}(\ell_i^\star)) + \frac{2\varepsilon n_\varepsilon}{N} + \widehat{R}_P(\boldsymbol{\ell}^\star) + \Delta_T(\beta, n_l, \delta) + \Delta_I(h, N, \eta)$$

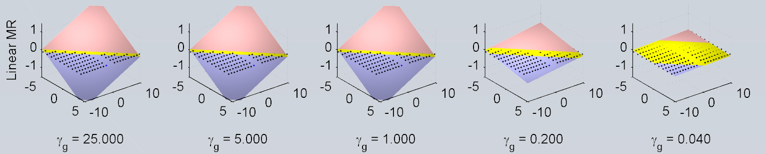
We should have $\varepsilon \le n_l^{-1/2}$!

SSL with Graphs: LapSVMs and MM Graph Cuts

MMGC for 2D data and **linear** $\mathcal{K}$ works as we want

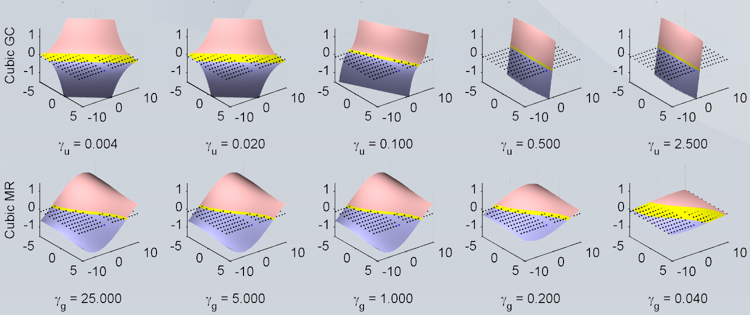


LSVM for 2D data and **linear** $\mathcal{K}$ only changes the slope



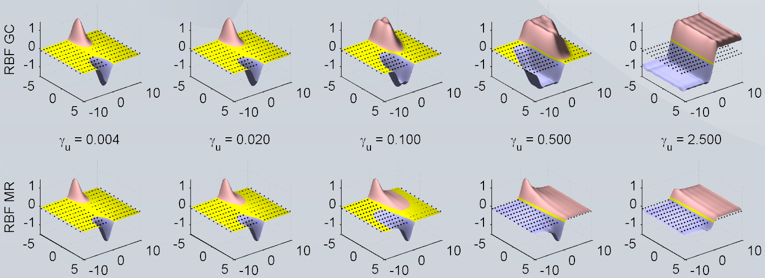
SSL with Graphs: LapSVMs and MM Graph Cuts

LSVM for 2D data and **cubic** $\mathcal{K}$ is also not so good



SSL with Graphs: LapSVMs and MM Graph Cuts

MMGC and LSVM for 2D data and RBF $\mathcal{K}$



Michal Valko

michal.valko@inria.fr

Inria & ENS Paris-Saclay, MVA

https://misovalko.github.io/mva-ml-graphs.html