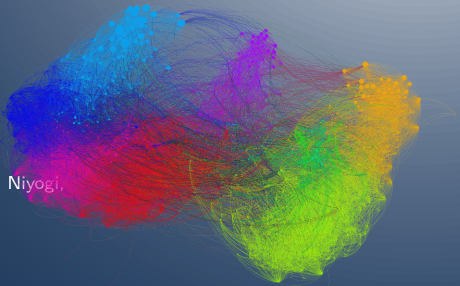
Graphs in Machine Learning

Transductive Generalization Bounds

Stability-Based Bounds for SSL

Michal Valko

Inria & ENS Paris-Saclay, MVA



Partially based on material by: Mikhail Belkin, Partha Niyogi, Olivier Chapelle, Bernhard Schölkopf

Transductive Generalization Bounds

True risk vs. empirical risk

Transductive Generalization Bounds

True risk vs. empirical risk

$$R_P(f) = \frac{1}{N}\sum_i (f_i - y_i)^2$$

$$\widehat{R}_P(f) = \frac{1}{n_l}\sum_{i\in l} (f_i - y_i)^2$$

Transductive Generalization Bounds

True risk vs. empirical risk

$$R_P(f) = \frac{1}{N}\sum_i (f_i - y_i)^2$$

$$\widehat{R}_P(f) = \frac{1}{n_l}\sum_{i \in l} (f_i - y_i)^2$$

We look for **transductive** bounds in the form

$$R_P(f) \leq \widehat{R}_P(f) + \text{errors}$$

Transductive Generalization Bounds

Bounding transductive error using stability analysis

http://www.cs.nyu.edu/~mohri/pub/str.pdf

Transductive Generalization Bounds

Bounding transductive error using stability analysis

http://www.cs.nyu.edu/~mohri/pub/str.pdf

How to bound $(f_i - y_i)^2$?

Transductive Generalization Bounds

Bounding transductive error using stability analysis

http://www.cs.nyu.edu/~mohri/pub/str.pdf

How to bound $(f_i - y_i)^2$?

We bound $(\ell_i^\star - y_i)^2$ with soft HFS (hard HFS is difficult):

Transductive Generalization Bounds

Bounding transductive error using stability analysis

http://www.cs.nyu.edu/~mohri/pub/str.pdf

How to bound $(f_i - y_i)^2$?

We bound $(\ell_i^\star - y_i)^2$ with soft HFS (hard HFS is difficult):

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

Transductive Generalization Bounds

Bounding transductive error using stability analysis

http://www.cs.nyu.edu/~mohri/pub/str.pdf

How to bound $(f_i - y_i)^2$?

We bound $(\ell_i^\star - y_i)^2$ with soft HFS (hard HFS is difficult):

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

Closed form solution

$$\boldsymbol{\ell}^\star = \left(\mathbf{C}^{-1}\mathbf{Q} + \mathbf{I}\right)^{-1} \mathbf{y}$$

Transductive Generalization Bounds

Consider two datasets differing in exactly one *labeled* point.

Transductive Generalization Bounds

Consider two datasets differing in exactly one *labeled* point.
$\boldsymbol{\ell}_1^\star = (\mathbf{C}_1^{-1}\mathbf{Q} + \mathbf{I})^{-1}\mathbf{y}_1$ and $\boldsymbol{\ell}_2^\star = (\mathbf{C}_2^{-1}\mathbf{Q} + \mathbf{I})^{-1}\mathbf{y}_2$

Transductive Generalization Bounds

Consider two datasets differing in exactly one *labeled* point.
$\boldsymbol{\ell}_1^\star = (\mathbf{C}_1^{-1}\mathbf{Q} + \mathbf{I})^{-1}\mathbf{y}_1$ and $\boldsymbol{\ell}_2^\star = (\mathbf{C}_2^{-1}\mathbf{Q} + \mathbf{I})^{-1}\mathbf{y}_2$

What is the maximal difference $\beta \geq \|\boldsymbol{\ell}_2^\star - \boldsymbol{\ell}_1^\star\|_\infty$ in the solutions?

Transductive Generalization Bounds

Consider two datasets differing in exactly one *labeled* point.
$\boldsymbol{\ell}_1^\star = (\mathbf{C}_1^{-1}\mathbf{Q} + \mathbf{I})^{-1}\mathbf{y}_1$ and $\boldsymbol{\ell}_2^\star = (\mathbf{C}_2^{-1}\mathbf{Q} + \mathbf{I})^{-1}\mathbf{y}_2$

What is the maximal difference $\beta \geq \|\boldsymbol{\ell}_2^\star - \boldsymbol{\ell}_1^\star\|_\infty$ in the solutions?

By the generalization bound of Belkin Belkin et al., 2004,
w.p.$1 - \delta$

http://web.cse.ohio-state.edu/~mbelkin/papers/RSS_COLT_04.pdf

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

Transductive Generalization Bounds

Bounding transductive error

$$\|\boldsymbol{\ell}_2^\star - \boldsymbol{\ell}_1^\star\|_\infty \leq \beta \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Transductive Generalization Bounds

Bounding transductive error

$$\|\boldsymbol{\ell}_2^\star - \boldsymbol{\ell}_1^\star\|_\infty \leq \beta \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Now, let us plug in the values for our problem.

Transductive Generalization Bounds

Bounding transductive error

$$\|\boldsymbol{\ell}_2^\star - \boldsymbol{\ell}_1^\star\|_\infty \leq \beta \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Now, let us plug in the values for our problem.

Remember that $\lambda_m(\mathbf{Q}) = \lambda_m(\mathbf{L}) + \gamma_g$ and $\lambda_M(\mathbf{Q}) = \lambda_M(\mathbf{L}) + \gamma_g$.

Transductive Generalization Bounds

Bounding transductive error

$$\|\boldsymbol{\ell}_2^\star - \boldsymbol{\ell}_1^\star\|_\infty \leq \beta \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Now, let us plug in the values for our problem.

Remember that $\lambda_m(\mathbf{Q}) = \lambda_m(\mathbf{L}) + \gamma_g$ and $\lambda_M(\mathbf{Q}) = \lambda_M(\mathbf{L}) + \gamma_g$.
Take $c_l = 1$

Transductive Generalization Bounds

Bounding transductive error

$$\|\boldsymbol{\ell}_2^\star - \boldsymbol{\ell}_1^\star\|_\infty \le \beta \le \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Now, let us plug in the values for our problem.

Remember that $\lambda_m(\mathbf{Q}) = \lambda_m(\mathbf{L}) + \gamma_g$ and $\lambda_M(\mathbf{Q}) = \lambda_M(\mathbf{L}) + \gamma_g$.
Take $c_l = 1$ and $c_l > c_u$.

Transductive Generalization Bounds

Bounding transductive error

$$\|\boldsymbol{\ell}_2^\star - \boldsymbol{\ell}_1^\star\|_\infty \leq \beta \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Now, let us plug in the values for our problem.

Remember that $\lambda_m(\mathbf{Q}) = \lambda_m(\mathbf{L}) + \gamma_g$ and $\lambda_M(\mathbf{Q}) = \lambda_M(\mathbf{L}) + \gamma_g$.
Take $c_l = 1$ and $c_l > c_u$. We have $|y_i| \leq 1$

Transductive Generalization Bounds

Bounding transductive error

$$\|\boldsymbol{\ell}_2^\star - \boldsymbol{\ell}_1^\star\|_\infty \leq \beta \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Now, let us plug in the values for our problem.

Remember that $\lambda_m(\mathbf{Q}) = \lambda_m(\mathbf{L}) + \gamma_g$ and $\lambda_M(\mathbf{Q}) = \lambda_M(\mathbf{L}) + \gamma_g$.
Take $c_l = 1$ and $c_l > c_u$. We have $|y_i| \leq 1$ and $|\ell_i^\star| \leq 1$.

Transductive Generalization Bounds

Bounding transductive error

$$\|\boldsymbol{\ell}_2^\star - \boldsymbol{\ell}_1^\star\|_\infty \leq \beta \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Now, let us plug in the values for our problem.

Remember that $\lambda_m(\mathbf{Q}) = \lambda_m(\mathbf{L}) + \gamma_g$ and $\lambda_M(\mathbf{Q}) = \lambda_M(\mathbf{L}) + \gamma_g$.
Take $c_l = 1$ and $c_l > c_u$. We have $|y_i| \leq 1$ and $|\ell_i^\star| \leq 1$.

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

Transductive Generalization Bounds

Bounding transductive error

$$\|\boldsymbol{\ell}_2^\star - \boldsymbol{\ell}_1^\star\|_\infty \leq \beta \leq \frac{\|\mathbf{y}_2 - \mathbf{y}_1\|_2}{\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1} + \frac{\lambda_M(\mathbf{Q})\|\mathbf{C}_1^{-1} - \mathbf{C}_2^{-1}\|_2 \cdot \|\mathbf{y}_1\|_2}{\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_2)} + 1\right)\left(\frac{\lambda_m(\mathbf{Q})}{\lambda_M(\mathbf{C}_1)} + 1\right)}$$

Now, let us plug in the values for our problem.

Remember that $\lambda_m(\mathbf{Q}) = \lambda_m(\mathbf{L}) + \gamma_g$ and $\lambda_M(\mathbf{Q}) = \lambda_M(\mathbf{L}) + \gamma_g$.
Take $c_l = 1$ and $c_l > c_u$. We have $|y_i| \leq 1$ and $|\ell_i^\star| \leq 1$.

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

This algorithm is β-stable!

Transductive Generalization Bounds

Bounding transductive error

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

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

Does the bound say anything useful?

Transductive Generalization Bounds

Bounding transductive error

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

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

Does the bound say anything useful?

1) The error is controlled.

Transductive Generalization Bounds

Bounding transductive error

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

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

Does the bound say anything useful?

1) The error is controlled.

2) Practical when error $\Delta_T(\beta, n_l, \delta)$ decreases at rate $O(n_l^{-\frac{1}{2}})$.

Transductive Generalization Bounds

Bounding transductive error

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

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

Does the bound say anything useful?

1) The error is controlled.

2) Practical when error $\Delta_T(\beta, n_l, \delta)$ decreases at rate $O(n_l^{-\frac{1}{2}})$.

Achieved when $\beta = O(1/n_l)$.

Transductive Generalization Bounds

Bounding transductive error

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

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

Does the bound say anything useful?

1) The error is controlled.

2) Practical when error $\Delta_T(\beta, n_l, \delta)$ decreases at rate $O(n_l^{-\frac{1}{2}})$.
Achieved when $\beta = O(1/n_l)$. That is, $\gamma_g = \Omega(n_l^{\frac{3}{2}})$.

Transductive Generalization Bounds

Bounding transductive error

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

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

Does the bound say anything useful?

1) The error is controlled.

2) Practical when error $\Delta_T(\beta, n_l, \delta)$ decreases at rate $O(n_l^{-\frac{1}{2}})$.
Achieved when $\beta = O(1/n_l)$. That is, $\gamma_g = \Omega(n_l^{\frac{3}{2}})$.

We have an idea how to set γ_g!

Michal Valko

michal.valko@inria.fr

Inria & ENS Paris-Saclay, MVA

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