Chromatic PAC-Bayes Bounds for non-IID Data
in NIPS 08 Workshop New Challenges in Theoretical Machine Learning:
Learning with Data-dependent Concept Spaces

Liva Ralaivola and Marie Szafranski and Guillaume Stempfel

Laboratoire d’Informatique Fondamentale de Marseille
Université Aix-Marseille I
39, rue F. Joliot Curie, F-13013 Marseille, France
Problem

New Results
Problem

Learning from non-IID data

- Bipartite ranking from pairwise data
- Classification of sequence data using sliding windows
- Classification of webpages
- ...

Goal: a Generalization Bound for non-IID Data

- Genericity
- Tightness
- Easily computable
- Motivates new algorithms
Example: bipartite ranking

Setting \((\mathcal{X}, \mathcal{Y} = \{-1, +1\}, \mathcal{Z} = \mathcal{X} \times \mathcal{Y})\)

- Training set: \(S = \{(X_i, Y_i)\}_{i=1}^\ell \in \mathcal{Z}^\ell\), i.i.d according to \(D\)
- Empirical and true ranking risks – \(n (p)\) number of negative (positive) data in \(S\)

\[
\hat{R}^{\text{rank}}(f, S) = \frac{1}{np} \sum_{i: Y_i=1, j: Y_j=-1} \mathbb{I}_{f(X_i) < f(X_j)} \quad (= 1 - \text{AUC}(f, S))
\]

\[
R^{\text{rank}}(f) = \mathbb{E}_S \hat{R}^{\text{rank}}(f, S) = \mathbb{P}_{X^\pm \sim D^\pm} (f(X^+) < f(X^-))
\]

Non-IIDness

\(\hat{R}^{\text{rank}}(f, S)\) involves non-independent r.v. \(Z_{ij} := (X_i^+, X_j^-) \in \mathcal{Z} := \mathcal{X} \times \mathcal{X} \times \mathcal{Y}\)
Problem

New Results
Property [4]
Let $\Gamma = (V, E)$ be a graph. Let $c(\Gamma)$ be the clique number of $\Gamma$. Let $\Delta(\Gamma)$ be the maximum degree of a vertex in $\Gamma$. The following holds

$$1 \leq c(\Gamma) \leq \chi^*(\Gamma) \leq \chi(\Gamma) \leq \Delta(\Gamma) + 1.$$ 

In addition, $1 = c(\Gamma) = \chi^*(\Gamma) = \chi(\Gamma) = \Delta(\Gamma) + 1$ if and only if $\Gamma$ is totally disconnected.

Example: Bipartite Ranking

$$\chi^*(\Gamma) = \chi(\Gamma) = \max(n, p)$$
Chromatic Pac-Bayes Bounds

Theorem (Chromatic PAC-Bayes Bound)
∀m, ∀D_m, ∀H, ∀δ ∈ (0, 1], ∀P, with probability at least 1 − δ over the random draw of Z ∼ D_m, the following holds

∀Q, kl(\hat{e}_Q || e_Q) ≤ \chi^* \left[ KL(Q||P) + \ln \frac{m + \chi^*}{\delta \chi^*} \right],

(1)

where \chi^* is the fractional chromatic number of Γ(D_m).

Ingredients of the proof.
Usual route to prove the IID Pac Bayes bound, properties of proper exact fractional covers, and convexity (e.g. Jensen)
Chromatic Pac-Bayes Bound for Bipartite Ranking

Theorem (Linear Ranking Pac-Bayes Bound [1])

∀ ℓ, ∀ D over \( \mathcal{X} \times \mathcal{Y} \), ∀ \( \delta \in (0, 1] \), the following holds with probability at least \( 1 - \delta \) over the draw of \( S \sim D^\ell \):

\[
\forall w, \mu > 0, \quad kl(\hat{R}_{Qw,\mu}^{\text{rank}} \parallel R_{Qw,\mu}^{\text{rank}}) \leq \frac{1}{\ell_{\text{min}}} \left[ \frac{\mu^2}{2} + \ln \frac{\ell_{\text{min}} + 1}{\delta} \right].
\]

Where the prior \( P \) is \( \mathcal{N}(0, I) \), the posterior \( Q_{w,\mu} \) is \( \mathcal{N}(\mu, 1) \) in the direction of \( w \) and \( \mathcal{N}(0, 1) \) in all other directions.

Ranking experiments, preliminary results (UCI datasets, [3])

Gaussian kernel: \( k(x, x') = \exp(-\frac{\|x-x\|^2}{2\sigma^2}) \). Homemade ranker.

Diabetes (268 / 500) Flare Solar (367 / 300) German (211 / 489)


