Data-dependent prior PAC-Bayes Bounds: Empirical study

Emilio Parrado-Hernandez
Signal Processing and Communications
University Carlos III of Madrid

Foundations and new trends on PAC-Bayesian Learning

UCL, March 22-23 2010

Joint work with John Shawe-Taylor, Amiran Ambroladze and Shiliang Sun

Outline

- Experimental setup
- Algorithms
- Tightness of bounds
- Model selection
- Conclusions

Experimental setup

- Take 80% for training and 20% for testing
- In Prior SVM bounds use 50% of training set to learn the prior

- Model selection: use the bound to pick *C* and the Gaussian kernel width from a 5x5 grid [C={1,10,100,1000,10000}, s={.25, .5, 1, 2, 4} · sqrt(dim)]
- Report bound on the Gibbs classifier and error on the test set

Data-dependent priors

- Gaussian distributions for priors (Kullback-Leibler divergence easy to compute)
- Mean and Covariance matrix of the prior
- Common setting of PAC-Bayes bound considers zero-mean and identity covariance matrix as prior classifier distribution

Mean vector

- Prior-SVM: Use part of the training data to learn prior direction [mean] (*w*)
- Multiple Prior: consider *J* scalings η_j to place [spherical] Gaussians along (*w*)
- Expectation Prior: $\mathbb{E}\{y\phi(\mathbf{x})\}$
- MiniSVM: $\mathbb{E}\{\mathbf{w}_m\}$, where $\mathbf{w}_m$ is an SVM trained with *m* samples.

Covariance Matrix

- Identity Matrix: default choice
- τ Prior: Stretch the covariance along the prior direction

Algorithms

- SVM $\min_{\mathbf{w},\xi_i} \left[\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{m} \xi_i \right]$

subject to $y_i \mathbf{w}^T \phi(\mathbf{x}_i) \geq 1 - \xi_i$

$i = 1, \ldots, m$

- η Prior SVM $\min_{\mathbf{v},\eta,\xi_i} \left[\frac{1}{2}\|\mathbf{v}\|^2 + C\sum_{i=1}^{m-r} \xi_i \right]$

subject to

$$y_i(\mathbf{v} + \eta\bar{\mathbf{w}}_r)^T \boldsymbol{\phi}(\mathbf{x}_i) \geq 1 - \xi_i \; i = 1, \ldots, m - r$$

$$\xi_i \geq 0 \; i = 1, \ldots, m - r$$

Data sets description

Problem	**# samples**	**dim.**	**Test Error (10FCV)**
Waveform	5000	21	0.086 ± 0.008
Ringnorm	7400	20	0.014 ± 0.003
Pima	768	8	0.236 ± 0.031
Digits	5620	64	0.006 ± 0.002
Spam	4601	57	0.062 ± 0.008

SVM: Bound tightness

Problem	**PAC-Bayes**	**Prior PAC-B**	**τ PriorPB**	**Expec**	**τ-Expec**
han	0.175	0.107	0.108	0.157	0.176
wav	0.203	0.185	0.184	0.202	0.205
pim	0.424	0.420	0.423	0.428	0.433
rin	0.203	0.110	0.110	0.201	0.204
spa	0.254	0.198	0.198	0.249	0.255

	M=4		M=16		M=64		M=128	
Problem	Expe	τ Expe	Expe	τ Expe	Expe	τ Expe	Expe	τ Expe
han	0.163	0.176	0.172	0.176	0.176	0.176	0.177	0.176
wav	0.203	0.205	0.203	0.205	0.204	0.205	0.206	0.206
pim	0.432	0.435	0.435	0.436	0.446	0.440	0.458	0.445
rin	0.203	0.204	0.197	0.204	0.200	0.204	0.204	0.204
spa	0.251	0.255	0.250	0.255	0.255	0.255	0.256	0.256

ηPrior SVM: Bounds

				τ Expe MiniSVM			
Problem	Prior-PB	τ Prior-PB	τ Expe-PB	M=4	M=16	M=64	M=128
han	0.050	0.047	0.174	0.227	0.225	0.219	0.217
wav	0.178	0.176	0.215	0.246	0.243	0.241	0.242
pim	0.428	0.416	0.444	0.487	0.488	0.495	0.505
rin	0.053	0.050	0.222	0.290	0.289	0.288	0.289
spa	0.186	0.178	0.269	0.313	0.312	0.313	0.314

Quality of Priors

Classification error rate of the prior classifiers on the test sets

			MiniSVM			
Problem	Prior-SVM	Expec	M=4	M=16	M=64	M=128
han	0.0196	0.0890	0.0883	0.0765	0.0418	0.0285
wav	0.0921	0.3304	0.3001	0.1314	0.1057	0.0950
pim	0.2412	0.3568	0.3487	0.3071	0.2461	0.2403
rin	0.0203	0.4935	0.4944	0.4697	0.2214	0.1497
spa	0.0863	0.3900	0.3991	0.3528	0.1561	0.1143

SVM model selection

Problem	**2FCV**	**PAC-Bayes**	**Prior PB**	τ **Prior PB**	**Expec**	τ **Expec**
han	0.007	0.007	0.014	0.014	0.007	0.007
wav	0.090	0.084	0.088	0.087	0.084	0.084
pim	0.244	0.229	0.229	0.231	0.229	0.229
rin	0.016	0.018	0.018	0.018	0.018	0.018
spa	0.066	0.067	0.077	0.077	0.067	0.067

	M=4		M=16		M=64		M=128	
Problem	Expe	τ Expe	Expe	τ Expe	Expe	τ Expe	Expe	τ Expe
han	0.008	0.008	0.008	0.008	0.008	0.008	0.008	0.008
wav	0.082	0.082	0.082	0.082	0.082	0.082	0.082	0.082
pim	0.238	0.238	0.243	0.238	0.240	0.238	0.240	0.238
rin	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018
spa	0.066	0.066	0.066	0.066	0.066	0.066	0.066	0.066

Model Selection PSVM

				τ Expe MiniSVM			
Problem	Prior-PB	τ Prior-PB	τ Expe-PB	M=4	M=16	M=64	M=128
han	0.010	0.009	0.010	0.015	0.016	0.016	0.017
wav	0.087	0.086	0.085	0.085	0.085	0.085	0.085
pim	0.233	0.233	0.231	0.256	0.247	0.242	0.236
rin	0.016	0.016	0.018	0.020	0.020	0.019	0.018
spa	0.070	0.072	0.067	0.073	0.072	0.069	0.069

Summary

- Data dependent priors yield tighter values of the PAC-Bayes bound for SVM
- The better the prior, the tighter the bound
- Bound driven model selection results comparable to 2fold crossvalidation
- To optimise the bound as a regularisation term leads to tightest bounds and acceptable classification error rates