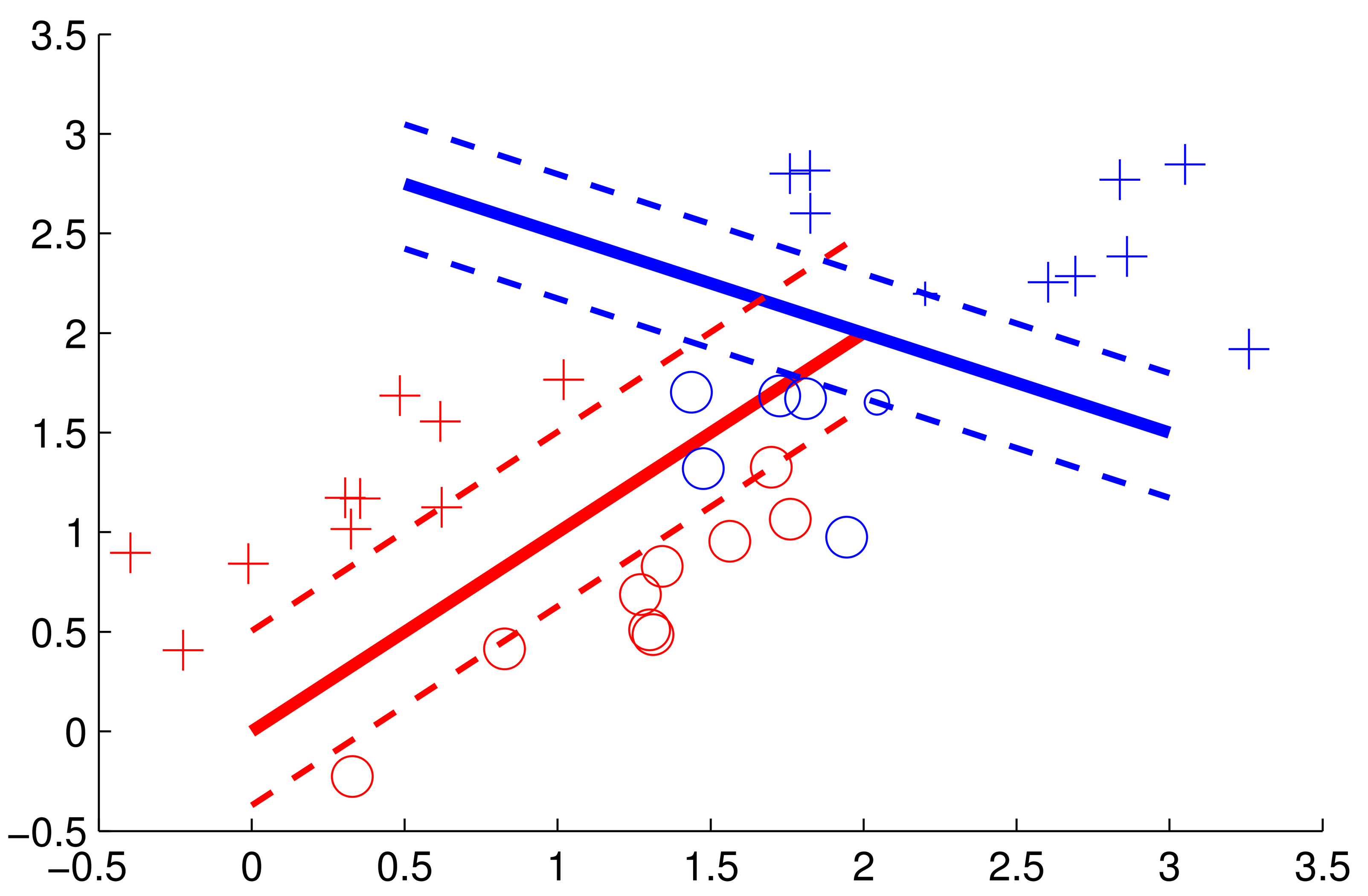
Small-variance Asymptotics for Dirichlet Process Mixtures of SVMs

Yining Wang and Jun Zhu

IIIS and TNList, Tsinghua University

Infininte SVM (*Zhu et al., ICML 2011*)

Infinite SVM: a supervised Bayesian nonparametric model



- Exponential family likelihood model:

$$p(\boldsymbol{x}|\boldsymbol{\mu}) = \exp(-D_\varphi(\boldsymbol{x}, \boldsymbol{\mu}))f_\varphi(\boldsymbol{x}).$$

- Chinese Restaurant Process nonparametric prior:

$$p(z_i = k|\alpha, \boldsymbol{z}_{-i}) = \begin{cases} n_{-i,k}, & \text{if } n_{-i,k} > 0 \\ \alpha, & \text{otherwise} \end{cases}$$

- Max-margin posterior regularization:

$$\min_{q(\Theta)\in\mathcal{P}} \mathrm{KL}(q\|p_0) - \mathbb{E}_q[\log p(\boldsymbol{X}|\boldsymbol{\Theta})] + 2c\cdot\mathcal{R}(q(\Theta), \boldsymbol{X}, \boldsymbol{y}),$$

$$\mathcal{R}(q(\Theta), \boldsymbol{X}, \boldsymbol{y}) = \sum_{i=1}^{n} \max(0, 1 - y_i\mathbb{E}_q[\boldsymbol{\eta}_{z_i}]^\top \boldsymbol{x}_i).$$

Gibbs Sampling

- Averaging classifier *vs.* Gibbs classifier:

$$\mathcal{R}'(q(\Theta), \boldsymbol{X}, \boldsymbol{y}) = \sum_{i=1}^{n} \mathbb{E}_q[\max(0, 1 - y_i\boldsymbol{\eta}_{z_i}^\top \boldsymbol{x}_i)].$$

- Data augmentation (*Polson and Scott, 2011*):

$$\phi(y_i|z_i, \boldsymbol{\eta}) = \int_0^\infty \frac{1}{\sqrt{2\pi\omega_i}} \exp\left(-\frac{(\omega_i + c\zeta_i^{z_i})^2}{2\omega_i}\right) d\omega_i.$$

- Posterior distribution:

$$q(\Theta, \boldsymbol{\omega}) = \frac{p_0(\Theta)\prod_{i=1}^n p(\boldsymbol{x}_i|z_i, \boldsymbol{\mu})\phi(y_i, \omega_i|z_i, \boldsymbol{\eta})}{Z(D)}.$$

Small-variance Asymptotic (SVA) Analysis

1. For exponential family distributions (taking $\beta \to \infty$):

$$p(\boldsymbol{x}|\tilde{\theta}(\boldsymbol{\mu})) = \exp(-\beta D_\varphi(\boldsymbol{x}, \boldsymbol{\mu}))f_{\tilde{\varphi}}(\boldsymbol{x}),$$

2. For existing clusters:

$$q(z_i = k|\boldsymbol{z}_{-i}, \beta) \propto n_{-i,k}f_{\tilde{\varphi}}(\boldsymbol{x}_i)\exp(-\beta D_\varphi(\boldsymbol{x}_i, \boldsymbol{\mu}_k) - 2c\beta'(\zeta_i^k)_+) \Rightarrow Q_i(k) = s\cdot D_\varphi(\boldsymbol{x}_i, \boldsymbol{\mu}_k) + 2c(\zeta_i^k)_+$$

3. For creating a new cluster:

$$q(z_i = z_{new}|\alpha, \beta) \propto \alpha\cdot p(\boldsymbol{x}_i|\beta)\cdot\int \phi(y_i|\boldsymbol{\eta}, \beta')dp_0(\boldsymbol{\eta}) \Rightarrow Q_i(new) = \lambda + 2c(\zeta_i^*)_+ + \frac{\|\boldsymbol{\eta}^*\|}{\nu^2}$$

Gibbs-iSVM

Repeat until convergence:

- $q(\boldsymbol{\mu}_k) \propto \exp(-\kappa' D_\varphi((\tau + \bar{\boldsymbol{x}}_k)/\kappa', \boldsymbol{\mu}_k)),$
- $q(\boldsymbol{\eta}_k) = \mathcal{N}(\boldsymbol{\eta}_k; \boldsymbol{\lambda}_k, \boldsymbol{\Lambda}_k),$
- $q(\omega_i^{-1}) = \mathcal{IG}(\omega_i^{-1}; 1/(c|\zeta_i^{z_i}|), 1),$
- $q(z_i = k) \propto \begin{cases} n_{-i,k}p(\boldsymbol{x}_i|\boldsymbol{\mu}_k)\phi(y_i|\boldsymbol{x}_i, \boldsymbol{\eta}_k), \text{ if } n_{-i,k} > 0 \\ \alpha\cdot(\boldsymbol{x}_i)\int\phi(y_i|\boldsymbol{\eta})dp_0(\boldsymbol{\eta}), \text{ otherwise} \end{cases}$

Max-Margin DP-means

Repeat until convergence:

- $\boldsymbol{\mu}_k \leftarrow \bar{\boldsymbol{x}}_k = \sum_{i\in\mathcal{N}_k}\boldsymbol{x}_i,$
- $\boldsymbol{\eta}_k \leftarrow \boldsymbol{\lambda}_k,$
- $\omega_i \leftarrow c\cdot|\zeta_i^{z_i}| = c\max(0, 1 - y_i\boldsymbol{\eta}_{z_i}^\top\boldsymbol{x}_i),$
- $z_i \leftarrow \mathrm{argmin}_k Q_i(k).$

Experiments

1. Classification performance

		MNL	L-SVM	RBF-SVM	dpMNL	DP+SVM	Gibbs-iSVM	M^2DPM
Synth. I	Acc (%)	66.4	69.5	66.2	68.8	70.8	70.9	71.1
	Time (s)	0.1	0.1	0.2	29.8	0.1	4.2	0.1
Synth. II	Acc(%)	54.4	65.5	54.6	54.6	58.9	62.9	64.4
	Time (s)	11.1	14.3	1.2	126.2	2.4	40.7	3.6
Protein	F1 (%)	41.2	47.3	49.5	49.5	47.9	50.1	49.9
	Times (s)	2.9	0.5	1.6	98.2	0.2	223.4	8.1
Parkinson	Acc. (%)	85.6	87.2	87.2	87.7	86.2	88.9	88.7
	Time (s)	0.1	0.1	0.1	22.2	0.1	1.8	0.1

2. Nonparametric clustering performance on synthetic datasets

n_0	100	300	1000	3000	10000
K_0	8	9	11	13	14
K	8	8	11	12	14

3. Clustering results on the Parkinson dataset

Group	I	II	III	IV	V
Avg. age	65.9	67.0	65.3	77.0	65.4
Avg. stage (0-4)	1.7	1.7	1.3	2.3	1.5

ynwang.yining@gmail.com dcszj@mail.tsinghua.edu.cn