Optimal rates for random Fourier features

Bharath Sriperumbudur*
Dept. of Statistics, PSU

Zoltán Szabó*
Gatsby Unit, UCL

#95

Montréal, NIPS
December 8, 2015

The kernel toolkit

- Widely successful and popular.
- Pro:
  - flexible; $f(x_i)$, $\partial^{\mathbf{d}} f(x_i) \xrightarrow{example}$
  - kernel ridge regression, infD exp. family fitting.
- Contra: poor scalability.
- Methods: incomplete Cholesky, Nyström, sketching, RFF.
- RFF [Rahimi & Recht '07]: $(\boldsymbol{\omega}_j)_{j=1}^m \overset{\text{i.i.d.}}{\sim} \Lambda$

$$k(\mathbf{x}, \mathbf{y}) = \int_{\mathbb{R}^d} e^{i\boldsymbol{\omega}^T(\mathbf{x}-\mathbf{y})} \mathrm{d}\Lambda(\boldsymbol{\omega}), \qquad \hat{k}(\mathbf{x}, \mathbf{y}) = \frac{1}{m}\sum_{j=1}^m e^{i\boldsymbol{\omega}_j^T(\mathbf{x}-\mathbf{y})}$$

$$= \langle \phi_m(\mathbf{x}), \phi_m(\mathbf{y}) \rangle_{\mathbb{C}^m}$$

$\Rightarrow$ possibility of fast linear methods (solving in the primal).

Approximation quality

- [Rahimi & Recht '07, Sutherland & Schneider '15]:

$$\|k - \hat{k}\|_{L^\infty(\mathcal{S})} = \mathcal{O}_p\left(|\mathcal{S}|\sqrt{\frac{\log m}{m}}\right).$$

- Contributions:
  1. Finite-sample L^∞-guarantee $\xrightarrow{\text{spec.}}$

$$\left\|k - \hat{k}\right\|_{L^\infty(\mathcal{S})} = \mathcal{O}_{a.s.}\left(\frac{\sqrt{\log|\mathcal{S}|}}{\sqrt{m}}\right).$$

     ECFs [Csörgő & Totik '83]: $|\mathcal{S}_m| = e^{o(m)}$ – optimal rate, asymptotic!
  2. Finite sample L^p guarantees.
  3. Finite sample guarantees in L^p and L^∞: $\partial^{\mathbf{a},\mathbf{b}}k$.

#95