Brief Summary

- **Kernel methods** [4].
- **RFF approximation for kernel derivatives + analysis.**
- Randomized algorithms:
  - Low-D feature representation → fast linear methods.
  - Random Fourier features (RFF) [3];
- **Contra:** computationally intensive, poor scalability.

RFF Idea (Kernel Approximation)

\[ k(x, y) = \int f^T(\omega) d\Lambda(\omega) = \int f^T(\omega(x - y)) d\Lambda(\omega) \]

\[ \hat{k}(x, y) = \frac{1}{m} \sum_{j=1}^{m} \cos(\omega_j^T(x-y)) \]

Existing RFF Guarantees

- **[3]:** \( \hat{k} \) is consistent (compact convergence).
- **[6]:** 3 RFF variants, better constants, same rate.

Theorem-1: k approximation, \( L^\infty(S \times S) \)

\[ \|k - \hat{k}\|_{L^\infty(S \times S)} \leq \sup \{ \|k(x, y) - \hat{k}(x, y)\| \} \]

\[ \leq O_{\lambda} \left( \sqrt{m^{-1} \log m} \right) \]

Theorem-2: k approximation, \( L^r(S \times S) \), \( 1 \leq r < \infty \)

For any \( \tau > 0, \) compact \( S \subset \mathbb{R}^d \)

\[ \Lambda^m \left[ \|k - \hat{k}\|_{L^r(S \times S)} \right] \geq \left( \frac{\sqrt{r}}{2^{r+1}4^r + 1} \right) \left( \frac{h(d, |S|, \sigma) + \sqrt{2\tau}}{\sqrt{m}} \right) \leq e^{-\tau}. \]

Remark-2:

- Consequence of Theorem-1.
- \( L^r(S \times S) \)-consistency: if \( \|k - \hat{k}\|_{L^r(S \times S)} = O_{\lambda} \left( \frac{m^{-1/2} |S|^{d/2}}{d \log(m)} \right) \).
- Growing diameter: \( \frac{\sqrt{m}}{\log(m)} \rightarrow 0 \Rightarrow \| |S|_n = o(m^{1/2}) \); \( L^\infty \)-case: \( |S|_n = e^{o(1)}. \)

Theorem-3: k approximation, \( L^r(S \times S) \), \( 1 < r < \infty \)

Applying a direct reasoning: for any \( \tau > 0, \) compact \( S \subset \mathbb{R}^d \)

\[ \Lambda^m \left[ \|k - \hat{k}\|_{L^r(S \times S)} \right] \leq \left( \frac{\sqrt{r}}{2^{r+1}4^r + 1} \right) \left( \frac{C_r}{m^{r-1} \log(m)} + \sqrt{2\tau} \right) \leq e^{-\tau}. \]

Remark-3:

- \( C_r = 0 \) (\( \tau \)), universal constant.
- \( L^r(S \times S) \)-consistency: if \( 2 \leq r \), then \( \|k - \hat{k}\|_{L^r(S \times S)} = O_{\lambda} \left( \frac{m^{-1/2} |S|^{d/2}}{d \log(m)} \right) \).

Kernel Derivative Approximation

\[ \frac{\partial}{\partial k} (x, y) \approx \frac{1}{m} \sum_{j=1}^{m} \left( \langle \omega_j^T \rangle \langle \omega_j(x-y) \rangle \right) \]

\[ \approx \left( \frac{1}{m} \sum_{j=1}^{m} \langle \omega_j \rangle \langle \omega_j(x-y) \rangle \right) \frac{1}{m} \]

Theorem-4: \( \partial^q \partial^q k(x, y) \) approx., \( supp(\Lambda) \) bounded, \( L^\infty(S \times S) \)

Let \( p, q < \infty \)

\[ \partial^p \partial^q k(x, y) = \partial^p \partial^q \langle \omega_j \rangle \langle \omega_j(x-y) \rangle \]

\[ \approx \left( \frac{1}{m} \sum_{j=1}^{m} \langle \omega_j \rangle \langle \omega_j(x-y) \rangle \right) \frac{1}{m} \]

Future Research Directions

- Kernel derivatives: tighter guarantees.
- Prediction using kernel (derivative) estimates.
- Analysis of smart RFF approximations [2].

References


\( ^{\dagger} \) Contributed equally. \( ^{\star} \) Gatsby Unit Qimera Review Symposium, London, UK, September 23rd, 2015. Acknowledgements: This work was supported by the Gatsby Charitable Foundation.