HMMA 308 : Machine Learning

Large Dimension Probabilistic Methods

Cassandre Lepercque

https://github.com/cassandrelepercque/HMMA308_Project

Université de Montpellier

Table of Contents

Table of Contents

The main problem

- What is the problem?
- Multi-faced challenges!
- For who?
- How can we fix the problem ?

Table of Contents

Mapping model

Randomized feature map

$$z : \mathbb{R}^d \to \mathbb{R}^D,$$

$$k(x, y) = \langle \phi(x), \phi(y) \rangle \approx z(x)' z(y).$$

Where,

- x and y are in $\mathbb{R}^d$,
- $k(x, y,)$ defines an inner product,
- ϕ is a lifting,
- z' is the transposed matrix of z and z is low-dimensional.

Random Fourier Features

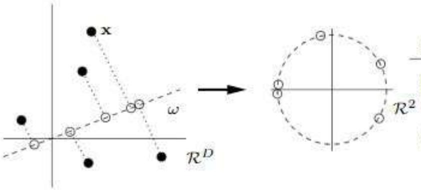
Random Fourier bases

$$cos(\omega' x + b),$$

Where,

- $\omega \in \mathbb{R}^d$,
- $b \in \mathbb{R}$,

are random variables.



Kernel Name	$k(\Delta)$	$p(\omega)$
Gaussian	$e^{-\frac{\|\Delta\|_2^2}{2}}$	$(2\pi)^{-\frac{D}{2}} e^{-\frac{\|\omega\|_2^2}{2}}$
Laplacian	$e^{-\|\Delta\|_1}$	$\prod_d \frac{1}{\pi(1+\omega_d^2)}$
Cauchy	$\prod_d \frac{2}{1+\Delta_d^2}$	$e^{-\|\Delta\|_1}$

Figure: Random Fourier features.

Random Fourier Features

Theorem

Bochner. *A continuous kernel $k(x, y) = k(x - y)$ on $\mathbb{R}^d$ is positive definite if and only if $k(\delta)$ is the Fourier transform of a non-negative measure.*

- Shift-invariant kernel properly scale, define by $\zeta_\omega(x) = e^{j\omega' x}$:
$$k(x - y) = \int_{\mathbb{R}^d} p(\omega) \times e^{j\omega'(x-y)} d\omega$$
$$= \mathbb{E}_\omega[\zeta_\omega(x)\ \zeta_\omega(y)],$$
- So, $\zeta_\omega(x)\ \zeta_\omega(y)$ is an unbiased estimate of $k(x, y)$ when ω is drawn from p.

Random Binning Features

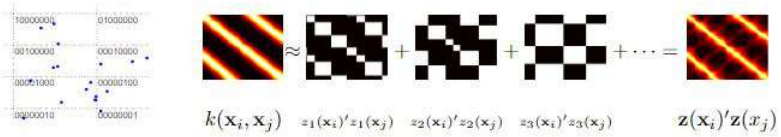


Figure: On the *left* side, you can see what the algorithm does. It repeatedly partitions the input space using a randomly shifted grid at a randomly chosen resolution and assigns to each point x the bit string $z(x)$ associated with the bin to which it is assigned. On the *right* side, you can see the binary adjacency matrix that describes this partitioning has $z(x_i)' z(x_j)$ in its ij-th entry and is an unbiased estimate of kernel matrix.

Random Binning Features

- randomized mapping to approximate the "hat" kernel as,
$$\hat{k}(x, y; \delta) = \max\left(0,\ 1 - \frac{|x-y|}{\delta}\right).$$

- Shift-invariant kernels written as a convex combinations of "hat" kernels on a compact subset of $\mathbb{R} \times \mathbb{R}$:
$$k(x, y) = \int_0^\infty \hat{k}(x, y; \delta)\ p(\delta) d\delta$$

- If the pitch δ of the grid is sampled from p, z gives a random map again for k, because,
$$\begin{aligned}\mathbb{E}_{\delta,u}[z(x)'z(y)] &= \mathbb{E}_\delta[\mathbb{E}_u[z(x)'z(y)|\delta]] \\ &= \mathbb{E}_\delta[\hat{k}(x, y; \delta)] = k(x, y).\end{aligned}$$

Table of Contents

Data Analysis

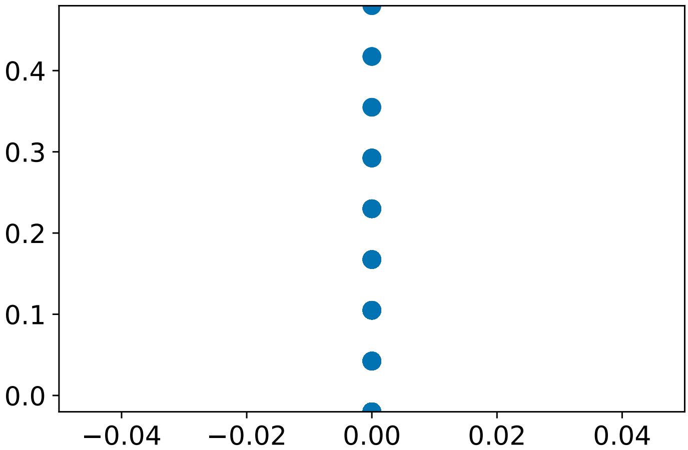


Figure: Visualization of the dataset.

Data Analysis

Figure: Classification accuracy of the dataset.

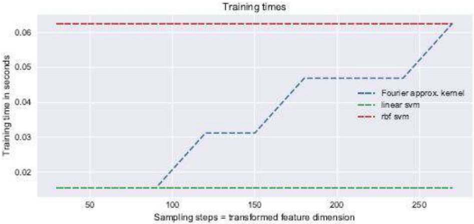


Figure: Training times of the dataset.

Data Analysis

Methods	RBF Kernel	Linear Kernel	Fourier approx. kernel
Score	0.972%	0.934%	0.954%

Table: Scores of the different methods.

Data Analysis

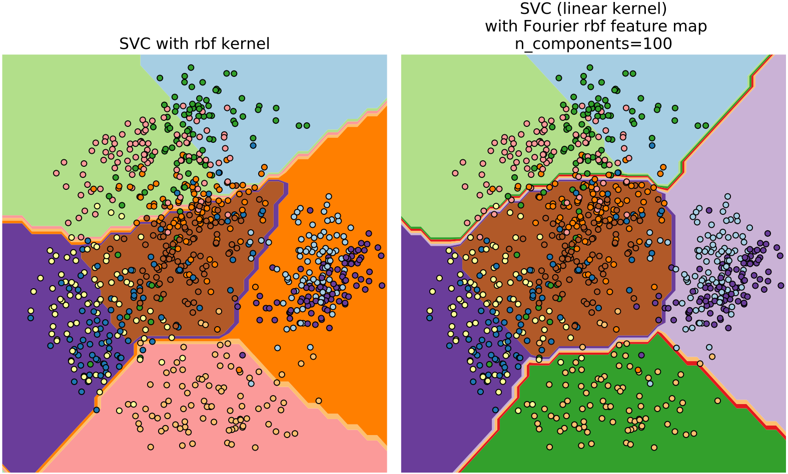


Figure: SVC (*Support Vector Classification*) with two different methods.

Table of Contents

Conclusion

- We have presented randomized features whose inner products uniformly approximate many popular kernels,and demonstrated that these features are a powerful and economical tool for large-scale supervised learning.
- We can note that any mixture of these features (like combining partitioning with Fourier features or sampling frequencies from mixture models) can be readily computed and applied to learning problems.

Conclusion

Thanks for your listening!