Fast Random Kernelized Features: Extending Support Vector Machine Classification for (High-Dimensional) IDC Classification

Tianshu Huang and Nimit Kalra

The University of Texas at Austin

December 12, 2018

Dataset and Objective

- IDC (Invasive Ductile Carcinoma) Dataset
- Samples are prepared and cropped into 50x50 patches
- 279 patients with 275,222 patches between them; each patient has patches classified as **normal or invasive**
- Binary Classification Problem (with Graded Variants)

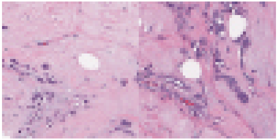


Figure: Normal or Invasive?

Random Fourier Features

- We approximate a shift-invariant kernel (Gaussian, Laplacian, Cauchy) by sampling from its Fourier transform (implemented via Monte-Carlo Rejection Sampling).
- Instead of operating on a kernel space with feature map $K(x, y) = \phi(x)^T \phi(y)$, approximate ϕ with $z : z(x)^T z(y) \approx K(x, y)$
- Feature space is $N \times D$-dimensional, which is an improvement over the original $N \times N$-dimensional kernel space.

$$e^{\frac{-\|x-y\|_2^2}{2}} \mapsto (2\pi)^{-D/2} e^{\frac{-\|z\|_2^2}{2}}$$

$$e^{\|x-y\|} \mapsto \prod_i \frac{1}{\pi(1 + z_i^2)}$$

$$\prod_i \frac{2}{1 + (x_i - y_i)^2} \mapsto e^{\|z\|_1}$$

Random Binning Features

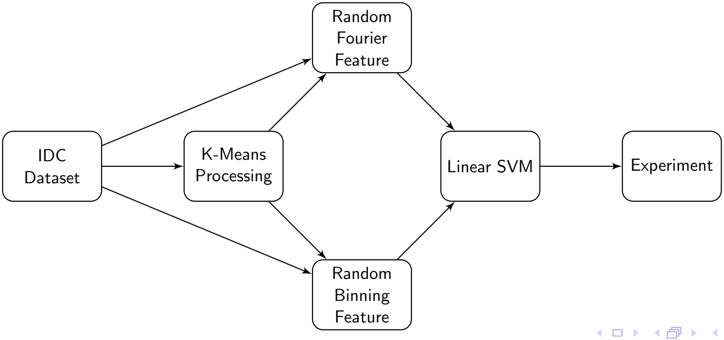
- Partition into bins such that $\mathbb{P}[\text{x and y are in same bin}] = K(x, y)$.
- Sample width from Laplacian. Sample shift uniformly.
- Many partitions will result in the average converging to $K(x, y)$, yielding a better approximation.
- For each width, "hash" the bin placements into a binary bit string.
- Repeat this once for each feature.

Testing Methodology

- Trained classifier on first 254 patients; tested against last 25 patients.
- Experiments run on TACC Lonestar5 (48 Core, 64GB RAM)
- Sanity Check: we evaluate against two minimum baseline criteria
  - Random: 50% error. Classes are randomly assigned.
  - Constant: Always classifying to the same class (minimum of the prior probability). For our testing data, this is 25.5% error.

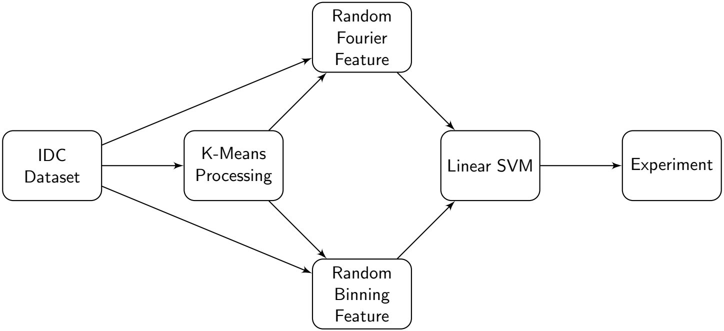
Testing Methodology: Implementation and Performance

- Option to randomly sample an equal proportion of training data from each patient (ex: pTrain = 0.01).
- Wrote multi-threaded implementations of Random Fourier Features and Random Binning Features.
- Random features generated on initialization. Pre-processing executed on image load to save memory.



Testing Methodology: Implementation and Performance

- Parallelized image loading; results piped through MapReduce.
- Processed images are input into a Linear SVM (sklearn.svm.LinearSVC).



Results and Performance

Table: Kernel SVM, Gaussian Kernel

N	2788	25665
False Positive	1466	1440
False Negative	2365	1984
True Positive	3001	3382
True Negative	14201	14227
Training Time	38.51	2012.53
Testing Time	241.25	2513.11
Per Test (ms)	11.47	119.48
Error	18.2%	16.3%

Results and Performance

Table: K-Means Histogram Based Random Feature Linear SVM

Feature Type	Binning	Binning	Binning	Fourier	Fourier
K	7	7	7	4	4
N	51036	254189	254189	254189	254189
Features	300	300	500	500	500
Kernel	Laplacian	Laplacian	Laplacian	Gaussian	Laplacian
False Positive	3024	1416	1279	1789	1648
False Negative	6740	2171	2234	2663	2775
True Positive	8040	3195	3132	2703	2591
True Negative	14362	14251	14388	13878	14019
Training Time	59.13	296.94	330.34	237.74	239.14
Testing Time	43.84	43.43	44.67	27.75	27.82
Per Test (ms)	2.08	2.06	2.12	1.32	1.32
Error	16.7%	17.1%	16.7%	21.2%	21.0%

Results and Performance

Table: RGB Image Based Random Feature Linear SVM

Feature Type	Binning	Fourier
N	254189	51036
Features	2500	7500
Kernel	Laplacian	Gaussian
False Positive	288	1451
False Negative	310	1759
True Positive	240	3607
True Negative	1290	14216
Training Time	1746.44	463.59
Testing Time	1227.95	77.41
Error	28.1%	15.3%

Research Strategies, Observations, and Future Directions

- Curse of dimenstionality: $50 \times 50 \times 3 = 7500$.
- Detecting overfitting/underfitting, etc. and their causes.
- Pre-processing: Histogram of Gradients, Sobel Operators, Edge Detection. Different color spaces (i.e: HSV).
- Capsule Neural Networks by Geoffrey Hinton are designed to include spatial relationships and scale. By manual inspection, size of geometric shapes is crucial to classification. Hypothesis: CapsNet beats CNN's for IDC Classification.
- Rotation of samples to maximize difference/variance; non-aligned patches. Low effect on risk of overfitting, order-of-magnitude increase in dataset size.
- Formulating SVM as Minimum Enclosing Ball: Core Vector Machines.
- Initial attempts had 35% error. We reduced this to 15%!

Citations

- Ali Rahimi and Benjamin Recht. "Random Features for Large-Scale Kernel Machines." Advances in Neural Information Processing Systems. 2008.
- Wu, Lingfei, et al. "Revisiting Random Binning Features: Fast Convergence and Strong Parallelizability." Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016.
- Netzer, Yuval, et al. "Reading Digits in Natural Images with Unsupervised Feature Learning." NIPS Workshop on Deep Learning and Unsupervised Feature Learning. Vol. 2011. No. 2. 2011.
- Janowczyk, Andrew. "Invasive Ductal Carcinoma Identification Dataset." http://www.andrewjanowczyk.com/deep-learning/.