Diversified Mini-Batch Sampling using Repulsive Point Processes

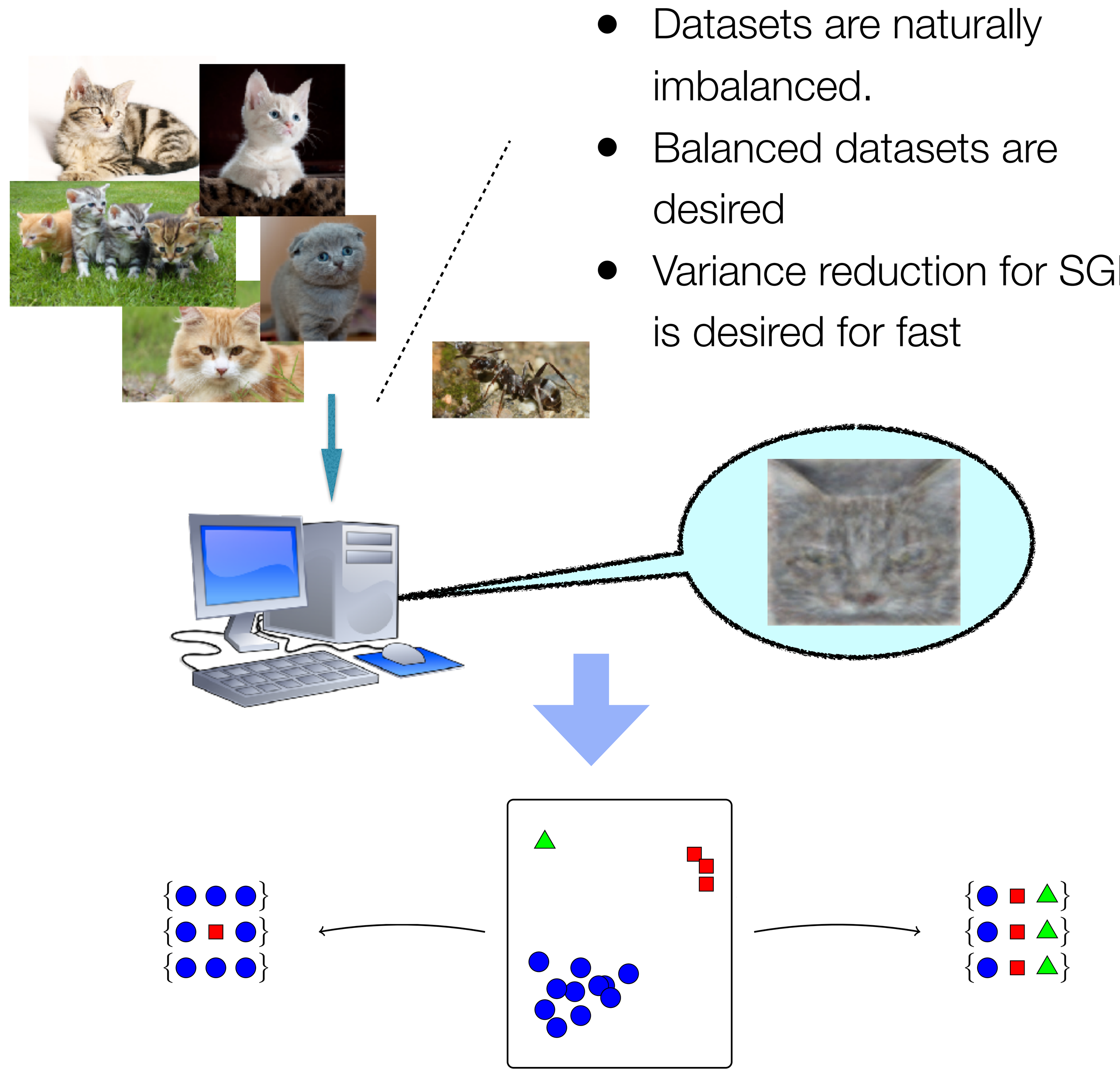
Cheng Zhang
Disney Research
Pittsburgh, USA

Cengiz Öztireli
Disney Research
Zurich, Switzerland

Stephan Mandt
Disney Research
Pittsburgh, USA

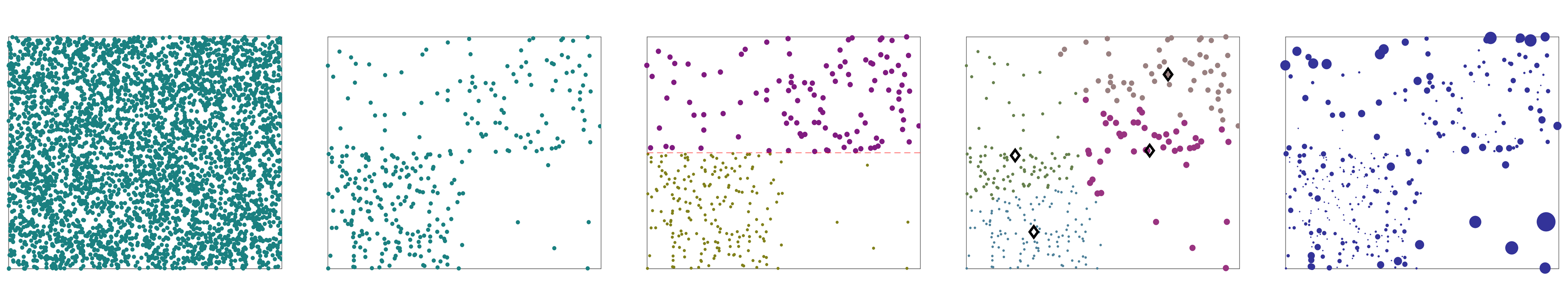
Introduction

- Datasets are naturally imbalanced.
- Balanced datasets are desired
- Variance reduction for SGD is desired for fast



Sample randomly　　Sample repulsively

Related Work



(a) True population (b) Dataset (c) Stratified Sampling (d) Pre-clustering (K-means) (e) k-DPP sampling

Goal: Point wise data re-balancing.

(c) P.L. Zhao and T. Zhang. Accelerating minibatch stochastic gradient descent using stratified sampling. *arXiv preprint arXiv:1405.3080*, 2014.
(d) T.F Fu and Z.H. Zhang. CPSG-MCMC: Clustering-based preprocessing method for stochastic gradient MCMC. In *AISTATS*, 2017.
(e) C. Zhang, H. Kjellström, and S. Mandt. Determinantal Point Processes for Mini-Batch Diversification. UAI, 2017.

- Determinantal Point Processes (DPP):

$$\mathscr{P}(Y) = \frac{det(L_Y)}{det(L+I)} \propto det(L_Y)$$

$$\begin{vmatrix} L_{ii} & L_{ij} \\ L_{ji} & L_{jj} \end{vmatrix} = L_{ii}L_{jj} - L_{ij}L_{ji}$$

- k-DPP:

$$\mathscr{P}_L^k(Y) = \frac{det(L_Y)}{\sum_{|Y'|=k} det(L_{Y'})}$$

Diversified risk: $J^*(\theta) = \frac{1}{k} \underset{\vec{x}\sim \text{k-DPP}}{\mathbb{E}} [\ell(\vec{x};\theta)]$　　$J^*(\theta) = \frac{1}{k}\sum_{i=1}^{N} b_i \ell(x_i,\theta)$

Main contribution

- **Generalzation of our previous working using DPP (d).** We generalize the proof of variance reduction and show that all we need is a point process with repulsive correlations.
- **Accelerate the sampling using Poisson Disk Sampling with dart throwing.**

$$\mathcal{O}(Nk^3) \longrightarrow \mathcal{O}(k^2)$$

Mini-Batch Sampling with Stochastic Point Processes

Define stochastic point processes with product densities [1,2]:

1st order product densities　$\lambda(\mathbf{x}) := \varrho^{(1)}(\mathbf{x}) = \lambda$

2nd order product densities　$\varrho(\mathbf{x},\mathbf{y}) := \varrho^{(2)}(\mathbf{x},\mathbf{y}) = \lambda^2 g(\mathbf{x}-\mathbf{y})$

pair correlation function

Diversified risk: $\hat{J}_{\text{data}'}(\theta) = \sum w_i l(\mathbf{x}_i,\theta)$　Gradient: $\mathbf{k}(\mathbf{x}_i,\theta) := \nabla l(\mathbf{x}_i,\theta)$

$$\text{var}_{\mathcal{P}}(\hat{\mathbf{K}}) = \int_{\mathcal{V}\times\mathcal{V}} w(\mathbf{x})\lambda(\mathbf{x})w(\mathbf{y})\lambda(\mathbf{y})\mathbf{k}(\mathbf{x},\theta)^T\mathbf{k}(\mathbf{y},\theta)\left[\frac{\varrho(\mathbf{x},\mathbf{y})}{\lambda(\mathbf{x})\lambda(\mathbf{y})} - 1\right] d\mathbf{x}d\mathbf{y} + \int_{\mathcal{V}} w^2(\mathbf{x})\|\mathbf{k}(\mathbf{x},\theta)\|^2\lambda(\mathbf{x})d\mathbf{x},$$

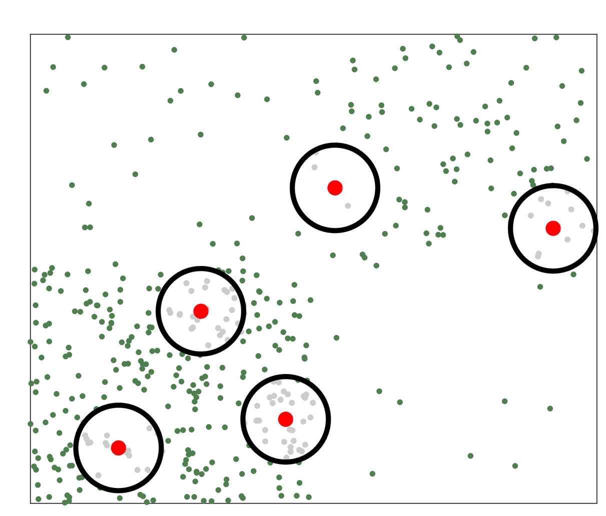
Variance for standard SGD

$\mathbf{k}(\mathbf{x},\theta)^T\mathbf{k}(\mathbf{y},\theta)\left[\frac{\varrho(\mathbf{x},\mathbf{y})}{\lambda(\mathbf{x})\lambda(\mathbf{y})} - 1\right] < 0 \longrightarrow$ Variance reduction

Any repulsive point processes

Poisson Disk Sampling with Dart Throwing

We can simulate all stochastic point processes with generalized dart throwing and statistics fitting as in [1].



The smallest distance between each pair of sample points should be at least r:

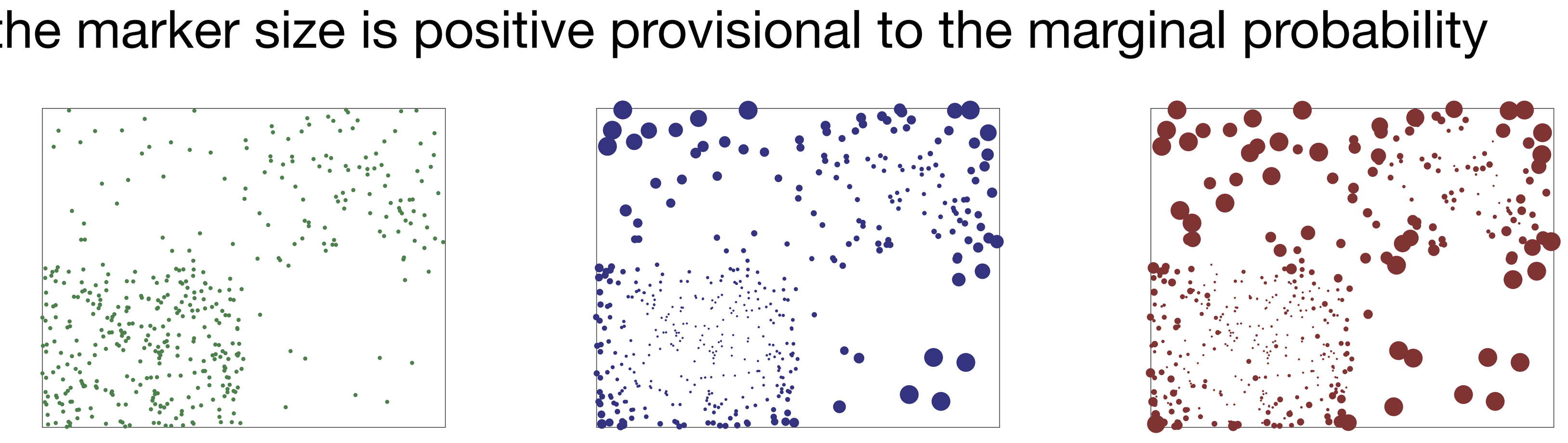
1.) randomly sample a data point;

2) if it is within a distance r of any already accepted data point, reject; otherwise, accept the new point.

$$\mathcal{O}(Nk^3) \longrightarrow \mathcal{O}(k^2)$$

Experiments

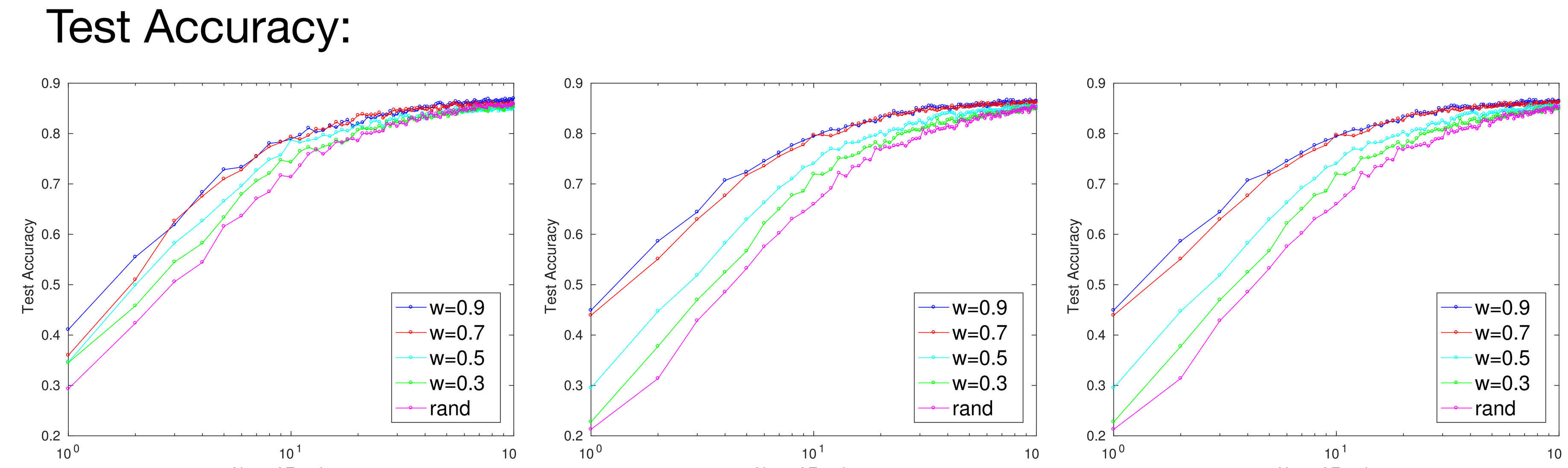
Toy example:
the marker size is positive provisional to the marginal probability



(a) data　(b) DPP　(c) Poisson Disk

Oxford Flower Classification:

Test Accuracy:



Sampling time:

k	50	80	102	150	200
k-DPP time	7.1680	29.687	58.4086	189.0303	436.4746
Fast k-DPP	0.1032	0.3312	0.6512	1.8745	4.1964
Poisson Disk	0.0461	0.0795	0.1048	0.1657	0.2391

References

[1] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan, editors. *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley and Sons, Ltd., 2008
[2] C.A. Öztireli and M. Gross. Analysis and synthesis of point distributions based on pair correlation. *ACM Trans. Graph.*,2012.