Determinantal Point Processes as Balancing Priors for Variational Autoencoder

Tian Chen

Departement of Statistics, UCI

Background

Imbalanced Learning Problem

- Imbalanced learning problem: significant or even extreme imbalances, unfavorable accuracies across classes

Imbalanced Learning Problem

- Imbalanced learning problem: significant or even extreme imbalances, unfavorable accuracies across classes
- e.g. Mammography Data Set: 10,923 'negative' samples and 260 'positive' samples;

Imbalanced Learning Problem

- Imbalanced learning problem: significant or even extreme imbalances, unfavorable accuracies across classes
- e.g. Mammography Data Set: 10,923 'negative' samples and 260 'positive' samples; majority class 100% accuracy, minority class 0-10% accuracy

Imbalanced Learning Problem

- Imbalanced learning problem: significant or even extreme imbalances, unfavorable accuracies across classes
- e.g. Mammography Data Set: 10,923 'negative' samples and 260 'positive' samples; majority class 100% accuracy, minority class 0-10% accuracy
- Potential solutions: sampling methods, cost-sensitive learning methods

Determinantal Point Process (DPP)

- Determinantal Point Process (DPP): a point process favors repulsion, which assigns higher probability to more diverse subsets

Determinantal Point Process (DPP)

- Determinantal Point Process (DPP): a point process favors repulsion, which assigns higher probability to more diverse subsets
- In a discrete setting, suppose the ground set is $\mathcal{Y}$, $\mathcal{P}$ is defined to be a determinantal point process, if for every $A \subseteq \mathcal{Y}$,

$$\mathcal{P}(A \subseteq Y) \propto det(L_A)$$

where L is a kernel matrix: $\mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$, and L_A is its submatrix corresponding to all entries in A.

Determinantal Point Process (DPP)

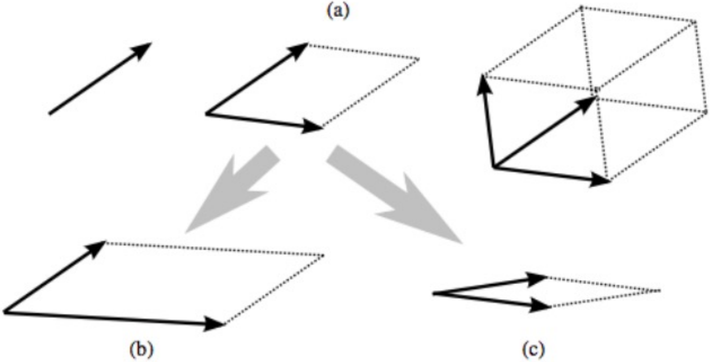
- Determinantal Point Process (DPP): a point process favors repulsion, which assigns higher probability to more diverse subsets
- In a discrete setting, suppose the ground set is $\mathcal{Y}$, $\mathcal{P}$ is defined to be a determinantal point process, if for every $A \subseteq \mathcal{Y}$,

$$\mathcal{P}(A \subseteq Y) \propto det(L_A)$$

  where L is a kernel matrix: $\mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$, and L_A is its submatrix corresponding to all entries in A.
- Example: $A = \{i, j\}$, where $i, j \in \mathcal{Y}$, then:

$$\mathcal{P}(i, j \in Y) \propto det(L_A) = \begin{vmatrix} L_{ii} & L_{ij} \\ L_{ji} & L_{jj} \end{vmatrix}$$

Determinantal Point Process (DPP)



1

[1]Kulesza, Alex, and Ben Taskar. "Determinantal point processes for machine learning." Foundations and Trends in Machine Learning 5.23 (2012): 123-286

Variations of DPP

- **Continuous DPP**: $\Omega \subseteq \mathbb{R}^D$, similarly, we have a positive definite kernel function $L : \Omega \times \Omega \to \mathbb{R}$ and for any point configuration $A \subseteq \Omega$: $P_L(A) \propto \det(L_A)$.

Variations of DPP

- **Continuous DPP**: $\Omega \subseteq \mathbb{R}^D$, similarly, we have a positive definite kernel function $L : \Omega \times \Omega \to \mathbb{R}$ and for any point configuration $A \subseteq \Omega$: $P_L(A) \propto \det(L_A)$.
- **k-DPP**: fix the subset size for every drawn. A k-DPP is a determinantal point process over subsets with cardinality k.
  - For discrete setting, the likelihood is:

$$P_L(A) = \frac{det(L_A)}{\sum_{|B|=k} det(L_B)} = \frac{det(L_A)}{e_k(\lambda_1, \cdots, \lambda_N)}$$

  - For continuous setting:

$$P_L(A) = \frac{det(L_A)}{e_k(\lambda_{1:\infty})}$$

where $e_k(\lambda_{1:\infty})$ is generally difficult to obtain.

Proposed Method

Balance Latent Space with DPP Prior

- For a latent variable model, the latent space will be redundant and dominated by the major class in the presence of imbalanced data

Balance Latent Space with DPP Prior

- For a latent variable model, the latent space will be redundant and dominated by the major class in the presence of imbalanced data
- DPP as a 'diversity encouraging prior' for the latent variables; also regarded as a regularizer

Balance Latent Space with DPP Prior

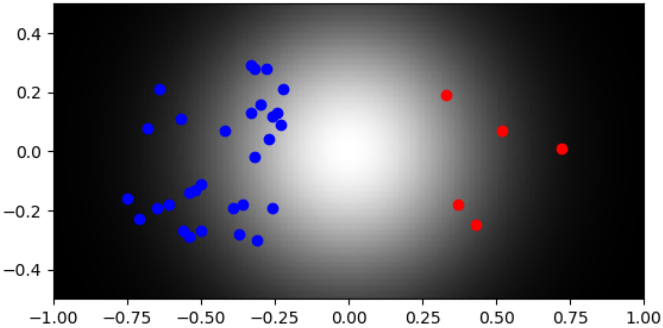
- For a latent variable model, the latent space will be redundant and dominated by the major class in the presence of imbalanced data
- DPP as a 'diversity encouraging prior' for the latent variables; also regarded as a regularizer
- Resultant prior is in favor of the minor class

Balance Latent Space with DPP Prior

- For a latent variable model, the latent space will be redundant and dominated by the major class in the presence of imbalanced data
- DPP as a 'diversity encouraging prior' for the latent variables; also regarded as a regularizer
- Resultant prior is in favor of the minor class
- Assumption: samples from the same class/cluster are more similar

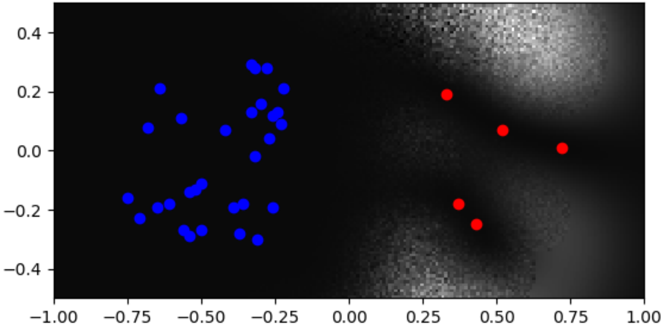
Balance Latent Space with DPP Prior

- Simulation: balanced intrinsic distribution, imbalanced samples
  - $p(z'|z)$ given z from two clusters with imbalanced ratio with independent standard normal prior: $p(z) = \prod_{n=1}^{N} N(z_n|0, I)$

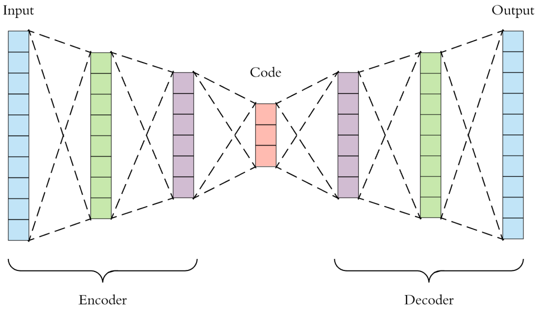


Balance Latent Space with DPP Prior

- Simulation: balanced intrinsic distribution, imbalanced samples
  - $p(z'|z)$ given z from two clusters with imbalanced ratio with continuous k-DPP prior: $\frac{det(L_Z)}{e_k(\lambda_{1:\infty})}$



Autoencoder



- Encoder: $\phi : \mathcal{X} \rightarrow \mathcal{F}$
- Decoder: $\psi : \mathcal{F} \rightarrow \mathcal{X}$
- $\phi, \psi = \arg\min_{\phi,\psi} \|X - (\psi \circ \phi)X\|^2$

Variational Autoencoder

- Variational autoencoder (VAE): prior $p(z)$ instead of deterministic z
  - Encoder: parameters ϕ of the approximating $q_\phi(z|x)$; sample z from $q_\phi(z|x)$
  - Decoder: parameters ψ of $p_\theta(x|z)$; reconstruct x by sampling from $p_\theta(x|z)$

Variational Autoencoder

- Variational autoencoder (VAE): prior $p(z)$ instead of deterministic z
  - Encoder: parameters ϕ of the approximating $q_\phi(z|x)$; sample z from $q_\phi(z|x)$
  - Decoder: parameters ψ of $p_\theta(x|z)$; reconstruct x by sampling from $p_\theta(x|z)$
- Variational approach for learning z: maximizing variational lower bound

$$\begin{aligned}\mathcal{L}(\theta, \phi; x) &= \log p_\theta(X) - KL(q_\phi(z|x) \| p_\theta(z|x)) \\ &= E_{q_\phi(z|x)}(\log p_\theta(x|z)) - KL(q_\phi(z|x) \| p_\theta(z))\end{aligned}$$

Variational Autoencoder

- Variational autoencoder (VAE): prior $p(z)$ instead of deterministic z
  - Encoder: parameters ϕ of the approximating $q_\phi(z|x)$; sample z from $q_\phi(z|x)$
  - Decoder: parameters ψ of $p_\theta(x|z)$; reconstruct x by sampling from $p_\theta(x|z)$
- Variational approach for learning z: maximizing variational lower bound

$$\begin{aligned}\mathcal{L}(\theta, \phi; x) &= \log p_\theta(X) - KL(q_\phi(z|x) \| p_\theta(z|x)) \\ &= E_{q_\phi(z|x)}(\log p_\theta(x|z)) - KL(q_\phi(z|x) \| p_\theta(z))\end{aligned}$$

  - $-E_{q_\phi(z|x)}(\log p_\theta(x|z))$: reconstruction loss
  - $KL(q_\phi(z|x) \| p_\theta(z))$: additional KL loss

DPP as prior for VAE

- Standard VAE: independent standard normal prior $p_\theta(z)$
- Modified VAE: continuous k-DPP prior $p_\theta(z)$

DPP as prior for VAE

- Standard VAE: independent standard normal prior $p_\theta(z)$
- Modified VAE: continuous k-DPP prior $p_\theta(z)$
- KL-Divergence Loss:

$$KL(q_\phi(z|x) \| p_\theta(z)) = \sum_{n=1}^{N} (-\log|\Sigma| - D_Z) - \ln\det(L_Z) + \ln(e_k(\lambda_{1:\infty}))$$

$e_k(\lambda_{1:\infty})$ not explicit but constant relative to z.

Experiments

Choosing a kernel function

- Here we use a positive definite kernel function $L(X)_{nm} = q(\mathbf{x}_n)k(\mathbf{x}_n, \mathbf{x}_m)q(\mathbf{x}_m)$ where

$$q(x) = \sqrt{\alpha} \prod_{d=1}^{D} \frac{1}{\sqrt{\pi \rho_d}} \exp(-\frac{x_d^2}{2\rho_d})$$

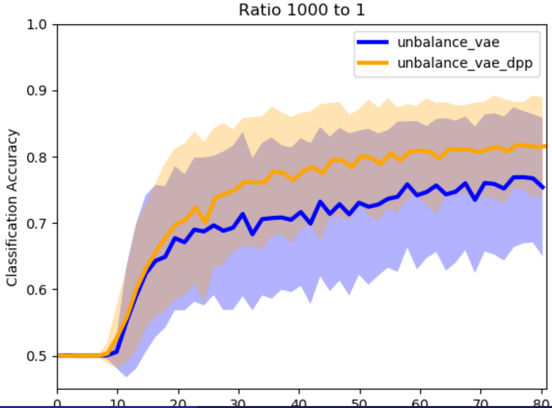
$$k(x, y) = \prod_{d=1}^{D} \exp(-\frac{(x_d - y_d)^2}{2\sigma_d})$$

Experiment 1

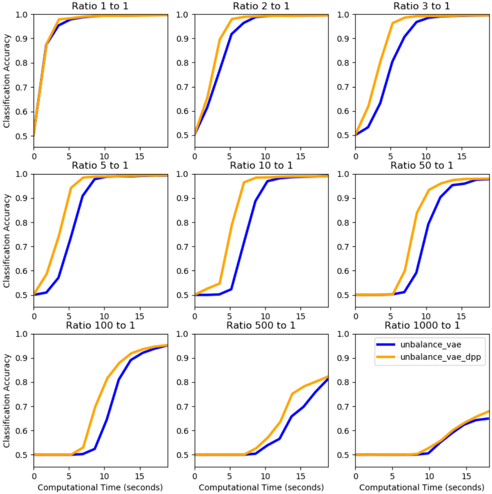
- Two-class MNIST data classification; Latent features are used for classification using logistic regression

Experiment 1

- Data: 5000 MNIST '0', '1' handwritten digits data (minor class: digit '1'). The test data is a balanced dataset with 500 class 0 and 500 class 1.



Experiment 1



Experiment 2

- Neural decoding: an application to multi-class imbalance learning problem
- 58 trials for odor A, 41 trials for odor B, 37 trials for odor C, 32 trials for odor D and 26 trials for odor E

Experiment 2

- Neural decoding: an application to multi-class imbalance learning problem
- 58 trials for odor A, 41 trials for odor B, 37 trials for odor C, 32 trials for odor D and 26 trials for odor E
- VAE and DPP-VAE comparison: Cross-validation performance

VAE

	precision	recall	f1-score
A	0.706	0.875	0.776
B	0.636	0.625	0.621
C	0.233	0.417	0.294
D	0.215	0.167	0.172
E	0.139	0.083	0.103
ave	0.386	0.433	0.393

DPP VAE

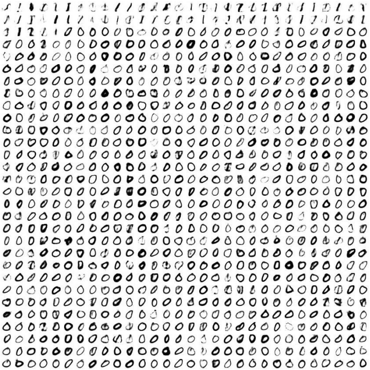
	precision	recall	f1-score
A	0.751	0.917	0.809
B	0.497	0.708	0.575
C	0.361	0.458	**0.377**
D	0.333	0.250	**0.278**
E	0.333	0.083	**0.133**
ave	0.455	0.483	**0.434**

Experiment 3

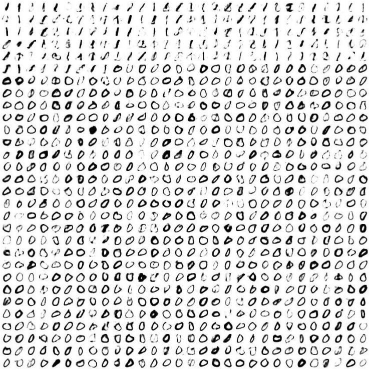
- Balancing data generation: random latent vectors are generated and passed into the trained decoder to generate handwritten ‘0”s and ‘1”s

Experiment 3

- Visualize 900 synthetic data with training ratio 10 to 1



Standard VAE



DPP VAE

Experiment 3

- Balancing data generation: 3 different imbalance ratios: 10 to 1, 100 to 1 and 1000 to 1.

Generated minor class (digit '1') percentage

Class ratio	Training (%)	VAE (%)	DPP-VAE (%)
10:1	9.1%	7.2%	**17.7%**
100:1	0.99%	1.21%	**3.68%**
1000:1	0.0999%	0.0562%	**0.9469%**

Discussion

- We proposed to use Determinantal Point Process as a diversity encouraging prior for latent variable models to alleviate imbalance learning problem

Discussion

- We proposed to use Determinantal Point Process as a diversity encouraging prior for latent variable models to alleviate imbalance learning problem
- Particular application: we modified variational autoencoder by using continuous k-DPP as latent prior, and developed the inference algorithm

Discussion

- We proposed to use Determinantal Point Process as a diversity encouraging prior for latent variable models to alleviate imbalance learning problem
- Particular application: we modified variational autoencoder by using continuous k-DPP as latent prior, and developed the inference algorithm
- Our proposed method improved the minority class performance compared to standard VAE in a classification task as well as generating more balanced synthetic data

Discussion

- We proposed to use Determinantal Point Process as a diversity encouraging prior for latent variable models to alleviate imbalance learning problem
- Particular application: we modified variational autoencoder by using continuous k-DPP as latent prior, and developed the inference algorithm
- Our proposed method improved the minority class performance compared to standard VAE in a classification task as well as generating more balanced synthetic data

Thanks!