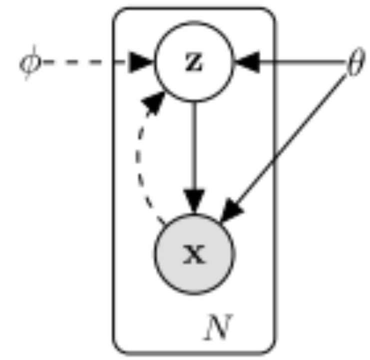
Semi-Supervised Learning with Deep Generative Models

Baruch Tabanpour

04/22/2015

Deep Generative Models

- Neural Net + Generative Model



$$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I}); \qquad p_\theta(\mathbf{x}|\mathbf{z}) = f(\mathbf{x}; \mathbf{z}, \boldsymbol{\theta}),$$

$$p_\theta(\mathbf{z}|\mathbf{x}) = p_\theta(\mathbf{x}|\mathbf{z})p_\theta(\mathbf{z})/p_\theta(\mathbf{x})$$

Deep Generative Models

- (Kingma & Welling, 2014; Rezende et al., 2014) simultaneously proposed:
- Approximative inference method for intractable posterior
- Scalable: make batch updates

Variational Auto-Encoder

- Recognition model $q_\phi(\mathbf{z}|\mathbf{x})$
- Lower bound of marginal likelihood

$$\mathcal{L}(\theta, \phi; \mathbf{x}^{(i)}) \simeq \mathcal{J}(\theta, \phi; \mathbf{x}^{(i)})$$

Algorithm 1 Auto-Encoding Variational Bayes

Initialize: ϕ, θ
repeat
  $\mathbf{X}^M \leftarrow$ Random minibatch of M datapoints
  $\epsilon \leftarrow$ Random sample from $p(\epsilon)$
  $\mathbf{g} \leftarrow \nabla_{\phi,\theta}\mathcal{J}^M(\theta, \phi; \mathbf{X}^M, \epsilon)$ (minibatch estimate to likelihood)
  $\phi, \theta \leftarrow$ Update using $\mathbf{g}$ (e.g. AdaGrad)
until Convergence of parameters (ϕ, θ)
return θ, ϕ

Using VAE for SSL

MNIST: 50k train/ 20k test

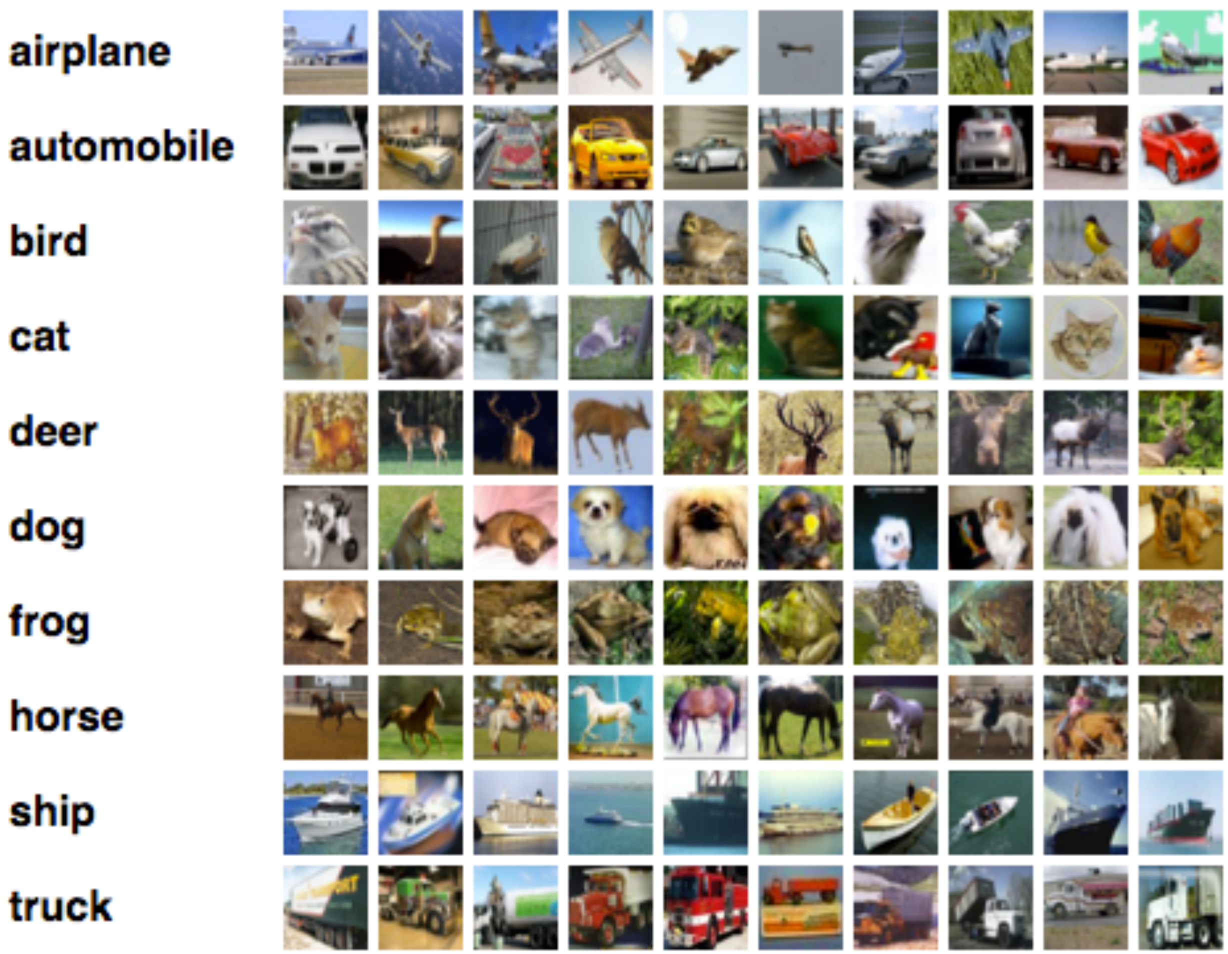
Table 1: Benchmark results of semi-supervised classification on MNIST with few labels.

N	NN	CNN	TSVM	CAE	MTC	AtlasRBF	M1+TSVM	M2	M1+M2
100	25.81	22.98	16.81	13.47	12.03	8.10 ($\pm$ 0.95)	11.82 ($\pm$ 0.25)	11.97 ($\pm$ 1.71)	**3.33** ($\pm$ 0.14)
600	11.44	7.68	6.16	6.3	5.13	–	5.72 ($\pm$ 0.049)	4.94 ($\pm$ 0.13)	**2.59** ($\pm$ 0.05)
1000	10.7	6.45	5.38	4.77	3.64	3.68 ($\pm$ 0.12)	4.24 ($\pm$ 0.07)	3.60 ($\pm$ 0.56)	**2.40** ($\pm$ 0.02)
3000	6.04	3.35	3.45	3.22	2.57	–	3.49 ($\pm$ 0.04)	3.92 ($\pm$ 0.63)	**2.18** ($\pm$ 0.04)

N	KNN	SVM	TSVM	M1+KNN	M1+SVM	M1+TSVM	M1+M2
100	35.3 ± 2.9	20.9 ± 1.6	16.845	40.7 ± 1.3	21.94 ± 1.35	12.38 ± 0.78	4.49 ± 0.1
600	16.0 ± 0.6	9.6 ± 0.5	8.06	16.2 ± 1.9	7.33 ± 0.60	5.93 ± 0.29	2.58 ± 0.06
1000	12.6 ± 0.9	8.30 ± 0.43	6.97	14.3 ± 0.8	5.87 ± 0.41	4.60 ± 0.12	2.45 ± 0.04
3000	8.1 ± 0.3	5.16 ± 0.21	4.7	8.7 ± 0.2	3.11 ± 0.12	2.95 ± 0.07	2.36 ± 0.05

Table 1. MNIST benchmark results

CIFAR-10


airplane
automobile
bird
cat
deer
dog
frog
horse
ship
truck

SSL on another dataset

- CIFAR-10 – 32 X 32 color images (50k train, 10k test)
- http://www.cs.toronto.edu/~kriz/cifar.html
- http://github.com/dpkingma/nips14-ssl