Variational Autoencoder

High Dimensional Regression using VAE Framework

2019.02.11

박 정수

DMIS Data Mining & Information Systems Laboratory

KOREA UNIVERSITY

Auto-Encoding Variational Bayes

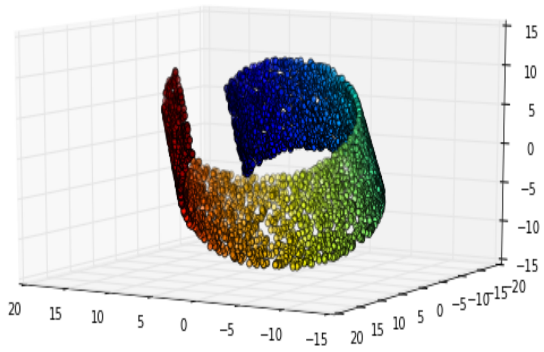
ICLR 2014

Diederik P.Kingma

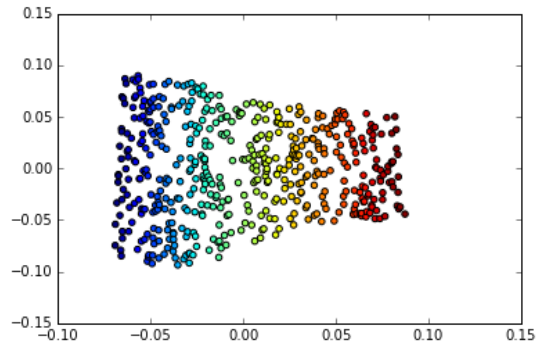
Max Welling

1 Introduction

Manifold



“Swiss Roll” : 3D



Manifold : 2D

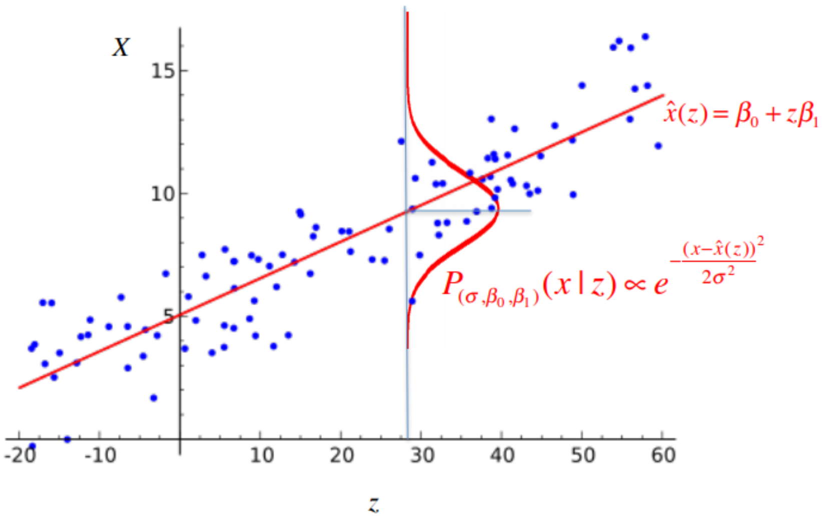
DMIS Data Mining & Information Systems Laboratory

KOREA UNIVERSITY

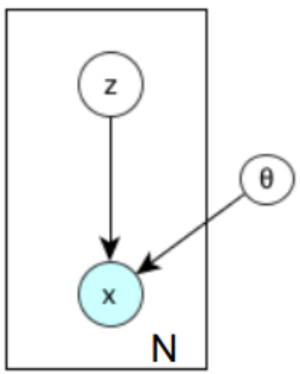
1 Introduction

Mapping from Z to X with Linear Regression

$P_\theta(X \mid Z)$



Graphical Notation

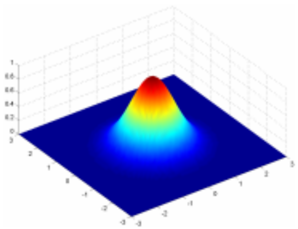


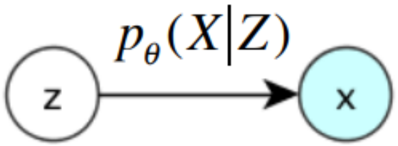
1 Introduction

Generating from MVG

$z \sim p(z)$ multivariate Gaussian

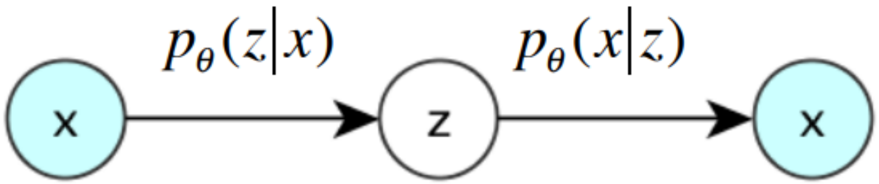
$x|z \sim p_\theta(x|z)$





1 Introduction

Intractability



$$p(z \mid \boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid z)p(z)}{p(\boldsymbol{x})},$$

$p_\theta(z|x)$ Intractable

Variational Bayesian Inference

2 Method

Variational Bayesian Inference

$p_\theta(z|x)$ Intractable

Approximate it by $q_\phi(\mathbf{z}|\mathbf{x})$

Basic Idea: Intractable density estimation transformed into optimization problem!

EX)How to approximate q to p (SGD) // intuitive explanation

$$D_{\mathrm{KL}}\left(q(z)||p(z\mid \boldsymbol{x})\right) = \int q(z)\log\frac{q(z)}{p(z)}dz + \int q(z)\log p(\boldsymbol{x})dz - \int q(z)\log p(\boldsymbol{x}\mid z)dz.$$

$$D_{\mathrm{KL}}\left(q(z)||p(z\mid \boldsymbol{x})\right) = D_{\mathrm{KL}}\left(q(z)||p(z)\right) + \log p(\boldsymbol{x}) - \mathbb{E}_{z\sim q(z)}\left[\log p(\boldsymbol{x}\mid z)\right]$$

$$\boxed{\frac{\partial}{\partial\theta_q}}D_{\mathrm{KL}}\left(q(z)||p(z\mid \boldsymbol{x})\right) = \frac{\partial}{\partial\theta_q}\mathbb{E}_{z\sim q(z)}\left[\log q(z) - \log p(z) - \log p(\boldsymbol{x}\mid z)\right]$$

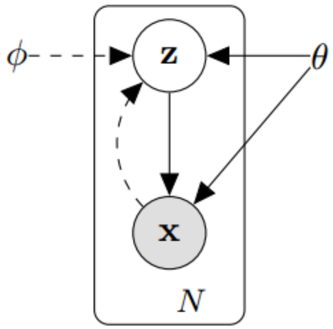
$$\frac{\partial}{\partial\theta_q}D_{\mathrm{KL}}\left(q(z)||p(z\mid \boldsymbol{x})\right) = \frac{\partial}{\partial\theta_q}\mathbb{E}_{\epsilon\sim \mathrm{N}(0,1)}\left[\log q(\mu_q+\sigma_q\epsilon) - \log p(\mu_q+\sigma_q\epsilon) - \log p(\boldsymbol{x}\mid z=\mu_q+\sigma_q\epsilon)\right]$$

$$\frac{1}{K}\sum_{i=0}^{K}\left[\frac{\partial}{\partial\theta_q}\left(\log q(\mu_q+\sigma_q\epsilon_i) - \log p(\mu_q+\sigma_q\epsilon_i) - \log p(\boldsymbol{x}\mid z=\mu_q+\sigma_q\epsilon_i)\right)\right]_{\epsilon_i\sim \mathrm{N}(0,1)}$$

2

Method

Auto-Encoding Variational Bayes



Directed graphical model

- Using the idea of variational inference, we can transform the problem of calculating posterior distribution into optimization problem
- Posterior and Likelihood distribution can be modeled as encoder and decoder in the form of auto encoder
- Then we can use gradient descent to optimize both posterior and likelihood

2 Method

The Variational Bound(ELBO)

$$
\begin{aligned}
\boxed{L = \log(p(x))} \\
&= \sum_z q(z|x)\ \log(p(x)) \\
&= \sum_z q(z|x)\ \log\left(\frac{p(z,x)}{p(z|x)}\right) \\
&= \sum_z q(z|x)\ \log\left(\frac{p(z,x)}{q(z|x)}\frac{q(z|x)}{p(z|x)}\right) \\
&= \sum_z q(z|x)\ \log\left(\frac{p(z,x)}{q(z|x)}\right) + \sum_z q(z|x)\ \log\left(\frac{q(z|x)}{p(z|x)}\right) \\
&= L^v + D_{KL}\left(q(z|x) \| p(z|x)\right) \\
&\geq L^v
\end{aligned}
$$

Can't be optimized

$$
\begin{aligned}
L^v &= \sum_z q(z|x)\ \log\left(\frac{p(z,x)}{q(z|x)}\right) \\
&= \sum_z q(z|x)\ \log\left(\frac{p(x|z)p(z)}{q(z|x)}\right) \\
&= \sum_z q(z|x)\ \log\left(\frac{p(z)}{q(z|x)}\right) + \sum_z q(z|x)\ \log(p(x|z)) \\
&= -D_{KL}\left(q(z|x) \| p(z)\right) + \mathbb{E}_{q(z|x)}\left(\log(p(x|z))\right) \\
&= -D_{KL}\left(q(z|x^{(i)}) \| p(z)\right) + \mathbb{E}_{q(z|x^{(i)})}\left(\log\left(p(x^{(i)}|z)\right)\right)
\end{aligned}
$$

Regularization   Reconstruction Error

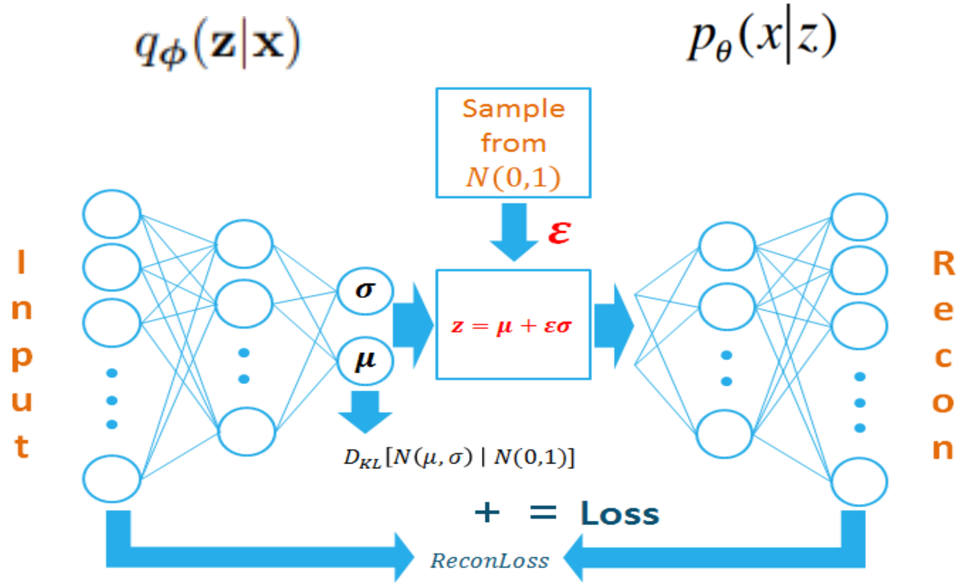
⇨ Variational Bound

DMIS Data Mining & Information Systems Laboratory

KOREA UNIVERSITY

2 Method

Variational Autoencoder

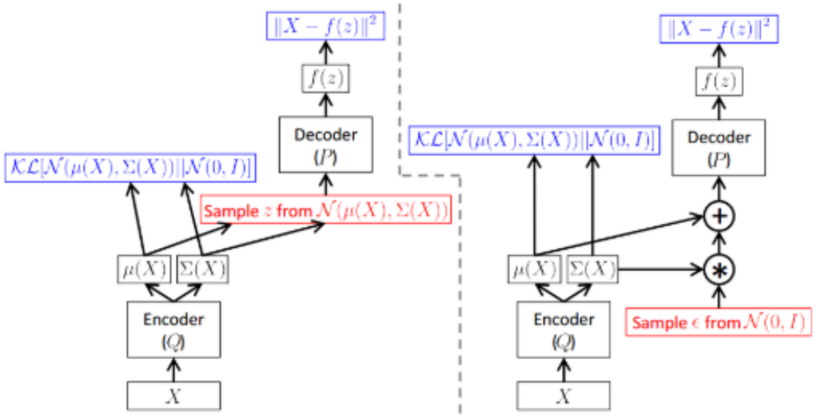


$$L^{V} \approx \underbrace{-D_{KL}\left(q(z|x^{(i)})||p(z)\right)}_{\text{Deterministic}} + \underbrace{\frac{1}{L}\sum_{i=1}^{L}\log\left(p(x^{(i)}|z^{(i,l)})\right)}_{\text{Via Sampling}}$$

Occasionally, L is set as 1

2 Method

Reparameterization Trick



$$z^{(i,l)} \sim N(\mu^{(i)}, \sigma^{2(i)})$$

$$z^{(i,l)} = \mu^{(i)} + \sigma^{(i)} \odot \varepsilon_i \quad \varepsilon_i \sim N(0,1)$$

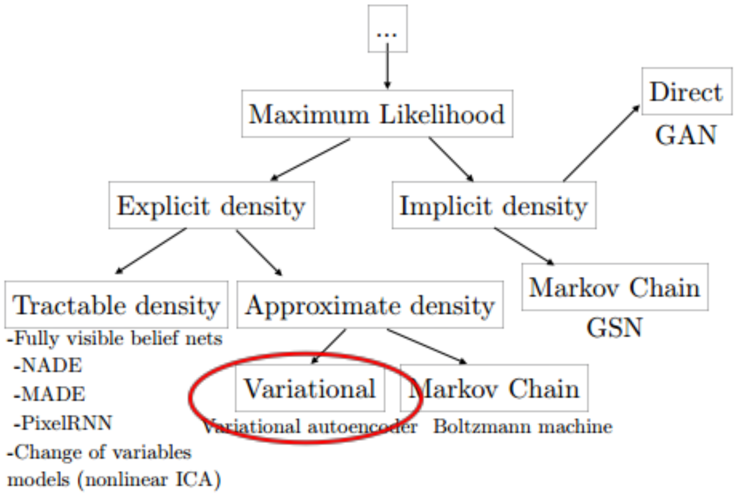
Location - Scale

Differentiable Transformation

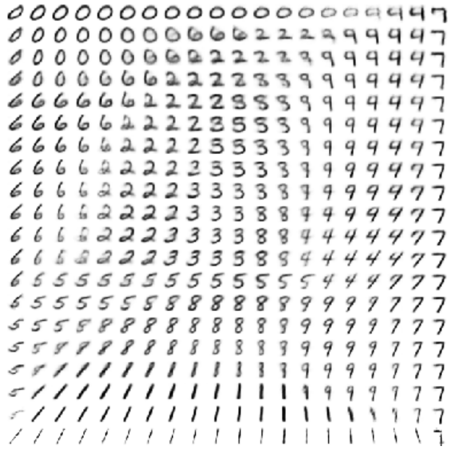
1. **Tractable Inverse CDF**
2. **Location - Scale**
3. **Composition**

3 Beyond VAE

Limitation and Beyond



Taxonomy of Generative Models



Images generated from VAE

Noisy

Why?

L2 loss, KL term, not able to learn true posterior probability…

DMIS Data Mining & Information Systems Laboratory

KOREA UNIVERSITY

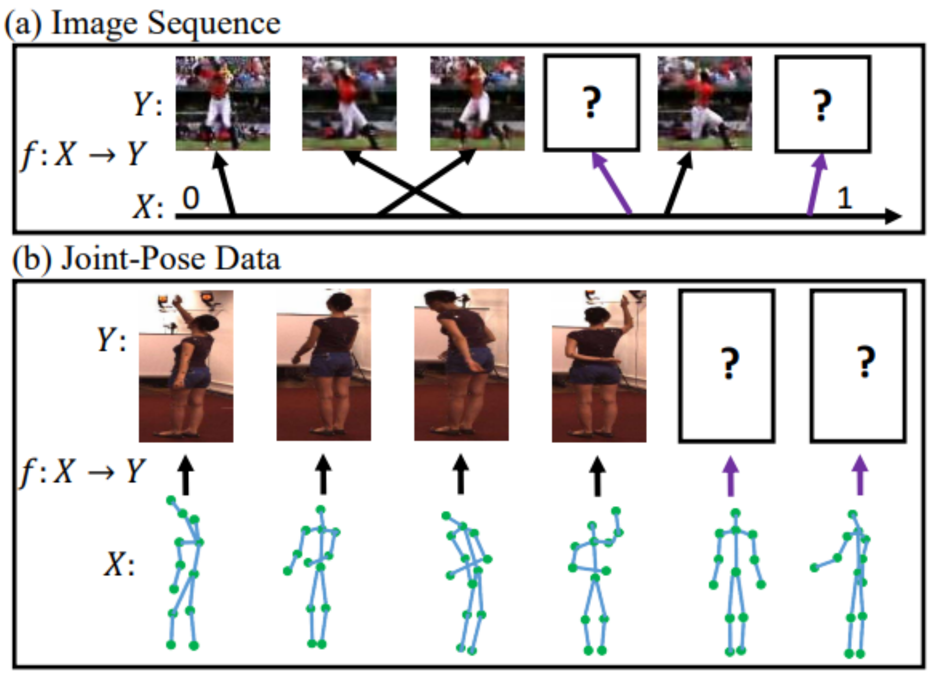
Variational Auto-encoded Regression: High Dimensional Regression of Visual Data on Complex Manifold

CVPR 2017

YoungJoon Yoo Sangdoo Yun

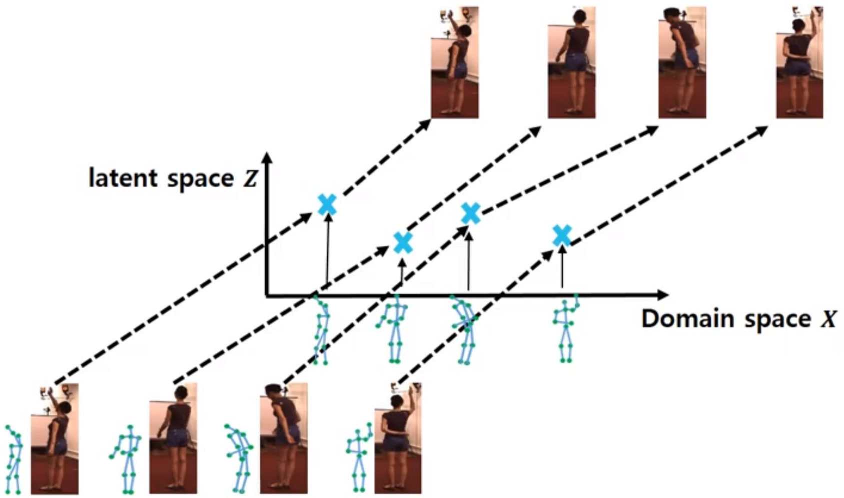
1 Introduction

Motive



2 Method

Regression in Latent Space

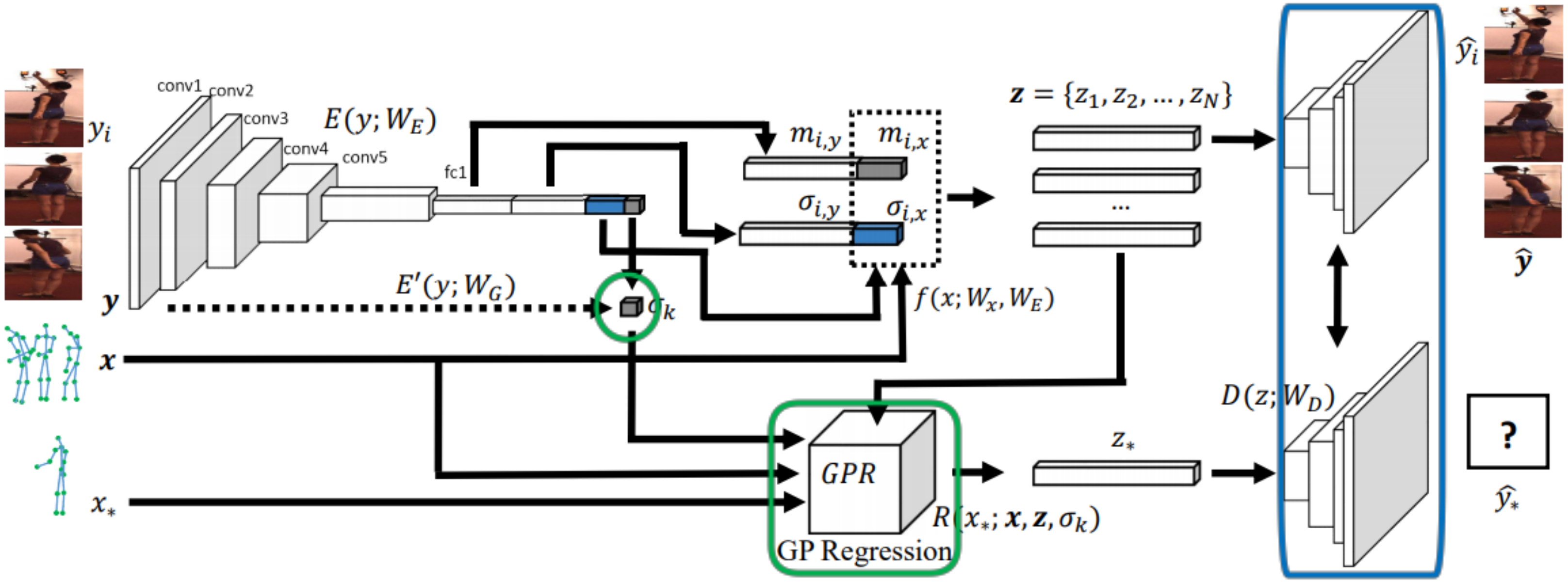


Map the high dimensional visual data into low dimensional latent space

Regress in latent space

2 Method

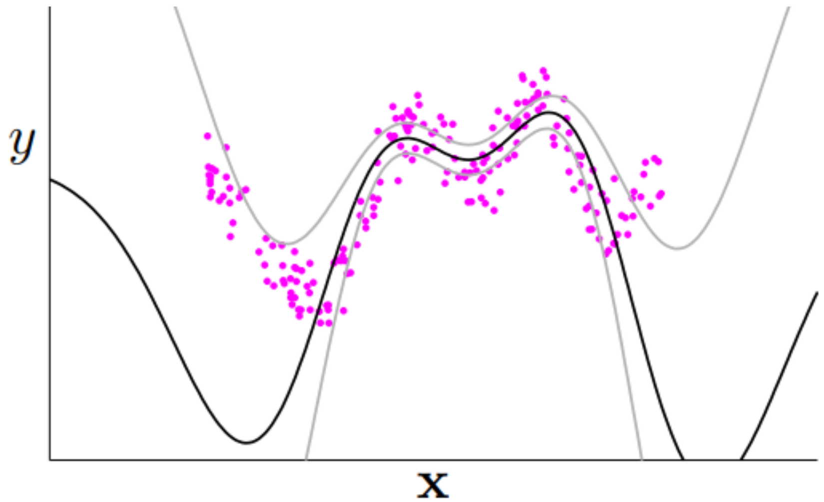
Pipeline



$$L(\theta,\phi) = -D_{KL}(q_\phi(z|x,y)||p_\theta(z)) + \sum_{i=1}^{N}\log p_\theta(y_i|z_i) + \sum_{j=1}^{M}\log p_\theta(\hat{y}_{*j}|z_{*j}).$$

2 Method

Gaussian Process Regression



Gaussian posterior distribution over functions
(Confidence Level)

$$m_{*,j} = K_{*,j}K^{-1}\mathbf{Z},\ \sigma_G = (K_{**,j} - K_{*,j}K^{-1}K_{*,j}^T)I.$$

$$K = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_N) \\ \vdots & \ddots & \vdots \\ k(x_N, x_1) & \cdots & k(x_N, x_N) \end{bmatrix},$$

$$K_{**,j} = k(x_{*,j}, x_{*,j}),$$

$$K_{*,j} = [k(x_{*,j}, x_1), k(x_{*,j}, x_2), \cdots, k(x_{*,j}, x_N)].$$

$$k(x_i, x_j) = \sqrt{\sigma_i \sigma_j} \exp ||x_i - x_j||^2$$

$$\mathcal{L} = -\log p(\mathbf{y}|\boldsymbol{\theta})$$

Gradient based optimization of K

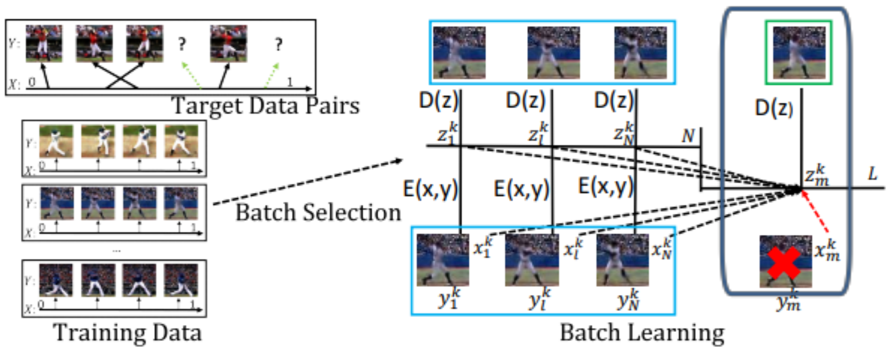
2

Method

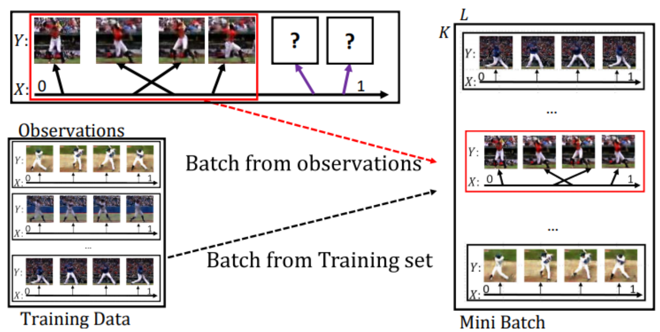
Training

n / n+m

m / n+m



Batch Construction



Finetuning

⇧

Note that the regression part is not trained

3 Experiments

Video Sequence & Pose Estimation

DMIS Data Mining & Information Systems Laboratory

KOREA UNIVERSITY

Thank you.