PATTERN RECOGNITION

JIN YOUNG CHOI

ECE, SEOUL NATIONAL UNIVERSITY

Notice

Lecture Notes: Slides

References:

Pattern Classification by Richard O. Duda, et al.

Pattern Recognition and Machine Learning by Cristopher M. Bishop

Assistant: 박슬기, 133(ASRI)-412, seulki.park@snu.ac.kr

Evaluation: Quiz 40%, Midterm 30%, Final 30%,

Video: Every week two videos are uploaded.

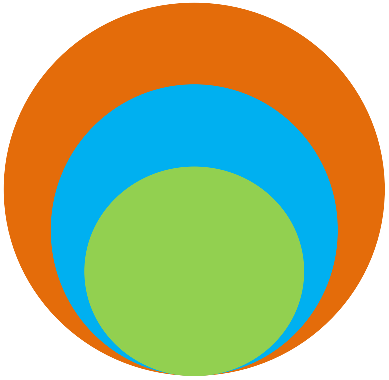
Video 1: **upload**: Sun. 09:00, **Quiz Due**: Tue. 24:00

Video 2: **upload**: Wed. 09:00, **Quiz Due:** Fri. 24:00

Class web: etl.snu.ac.kr

Office: 133(ASRI)-406, jychoi@snu.ac.kr

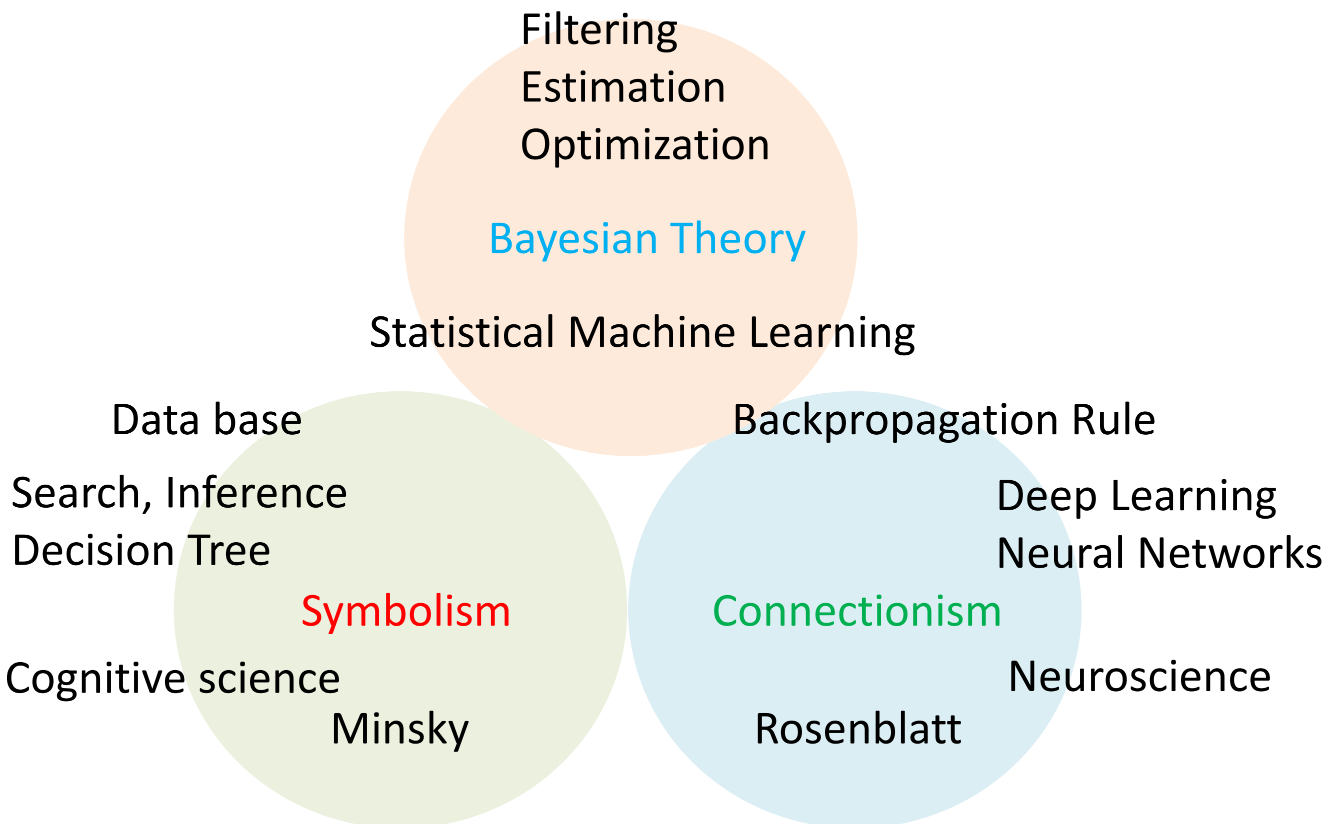
INTRODUCTION TO
AI: ARTIFICIAL INTELLIGENCE
ML: MACHINE LEARNING
DL: DEEP LEARNING



JIN YOUNG CHOI

ECE, SEOUL NATIONAL UNIVERSITY

Artificial Intelligence



Artificial Intelligence

Learning from Experience (Observations, Examples)

?

Inference(Reasoning) for a Question

?

Artificial Intelligence

Learning from Experience (Observations, Examples)

If birds are given, then we can learn their features such as # of legs, shape of mouth, etc.
If cancer patients are given, then we can observe their symptoms via diagnosis
.....

Inference(Reasoning) for a Question

If features of something are given, then we can recognize what is it.
If symptoms of a patient are given, then we can infer what is his decease.
......

Artificial Intelligence

Learning from Experience (Observations, Examples)

If birds are given, then we can learn their features such as # of legs, shape of mouth, etc.
If cancer patients are given, then we can record their symptoms via diagnosis
If $y = y_1$, then $x = x_1$
If $y = y_2$, then $x = x_2$

y			x			
y_1	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
y_2	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}

DB
Decision Tree

Inference(Reasoning) for a Question

If features of something are given, then we can recognize what is it.
If symptoms of a patient are given, then we can infer what is his decease.
If $x = x_1$, then $y = y_1$
If $x = x_2$, then $y = y_2$

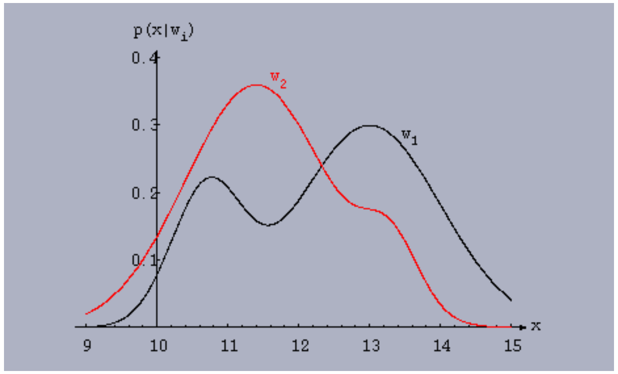
y			x			
y_1	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
y_2	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}

Symbolism

Search-based Inference Engine

Artificial Intelligence

Density Estimation



Learning from Experience (Observations, Examples)

If $y = y_1$, then $x = x_1$

If $y = y_2$, then $x = x_2$

$$p(x = x_i|y = y_j), p(y = y_j)$$

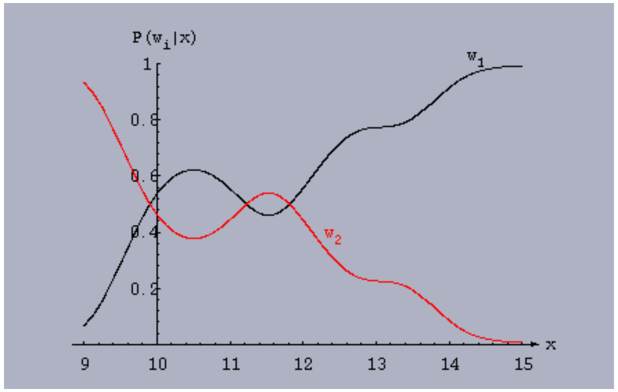
Inference(Reasoning) for a Question

If $x = x_1$, then $y = y_1$

If $x = x_2$, then $y = y_2$

$$p(y = y_j|x = x_i) = \frac{(p(x = x_i|y = y_j)p(y = y_j))}{p(x)},$$

$$p(x) = \sum_i p(x = x_i|y = y_j)p(y = y_j)$$



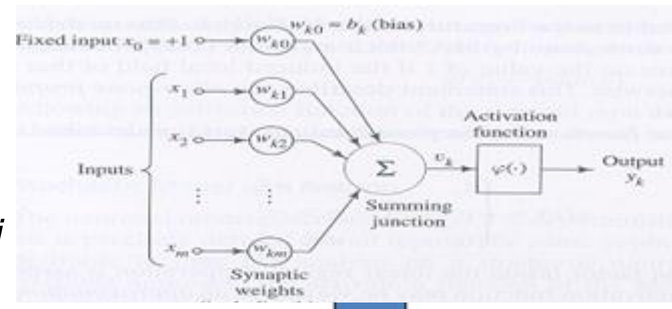
Bayesian Theory

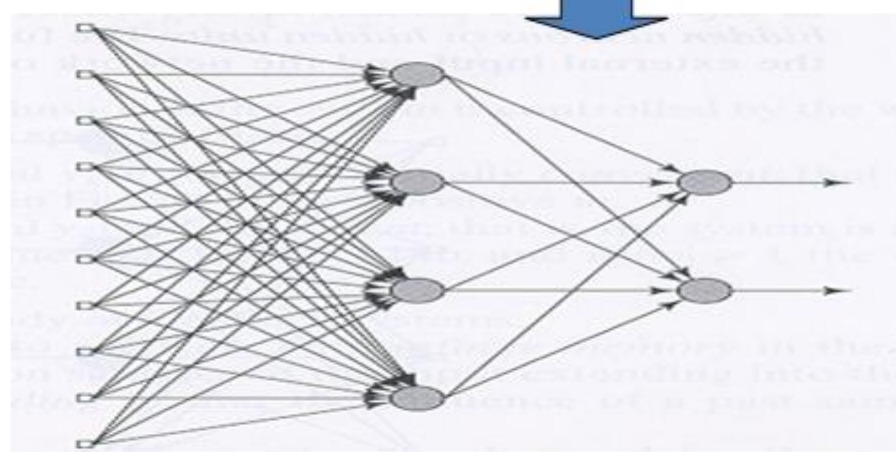
Artificial Intelligence

Deep Neural Networks for Learning and Inference

$$o = f(W, x), \text{ ex, } o_j = p(y = y_j | x = x_i)$$

Network Training
Inference (feedforward)





Training (learning)

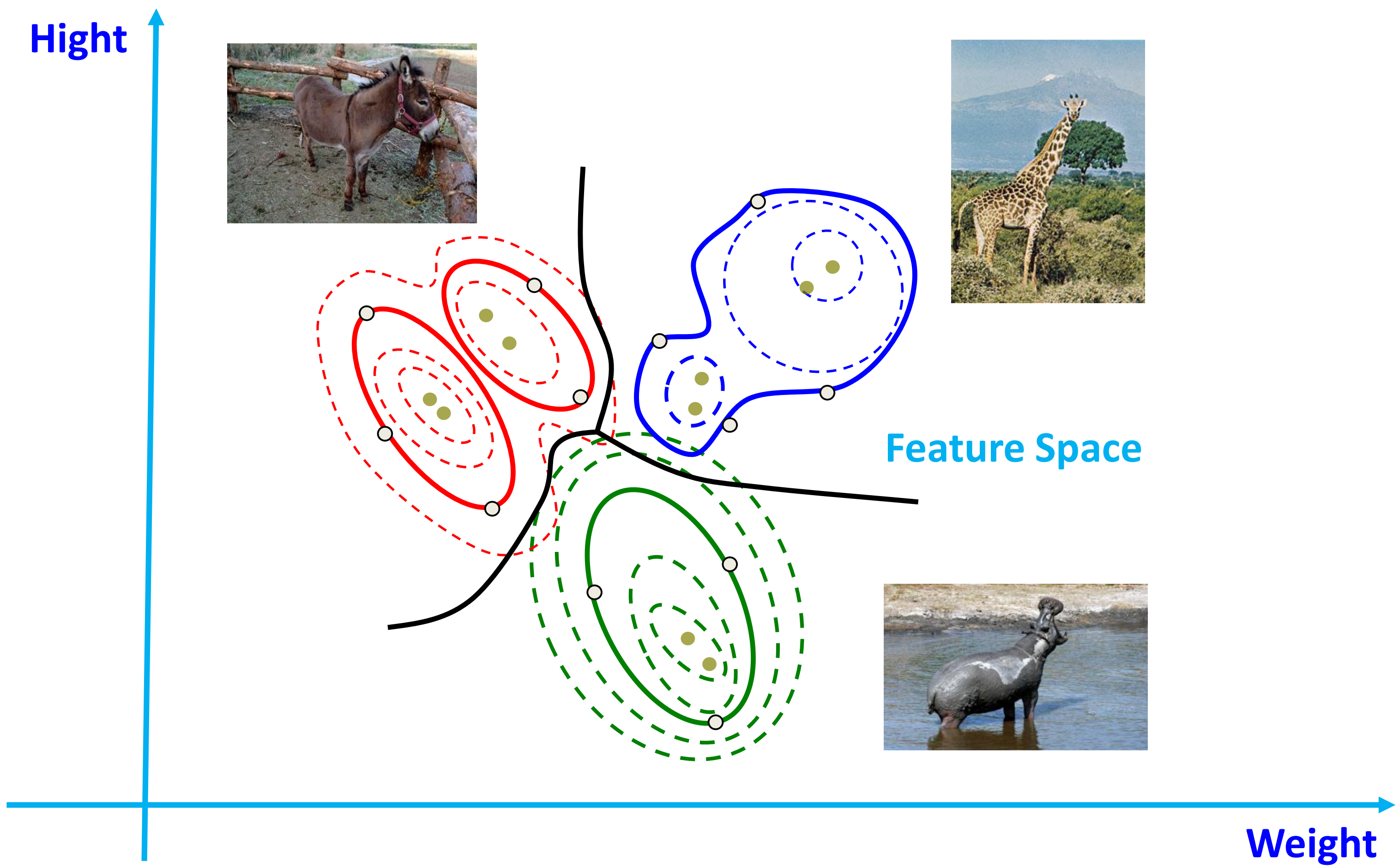
Find W to minimize the errors between o_j and d_j for given training data $\{(x_p, l_p)\}$

Inference(Reasoning)

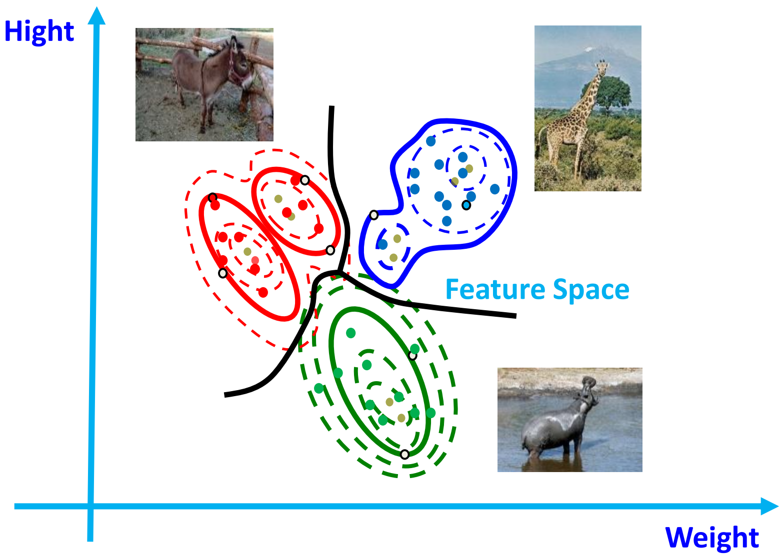
Calculate o_j via the deep network

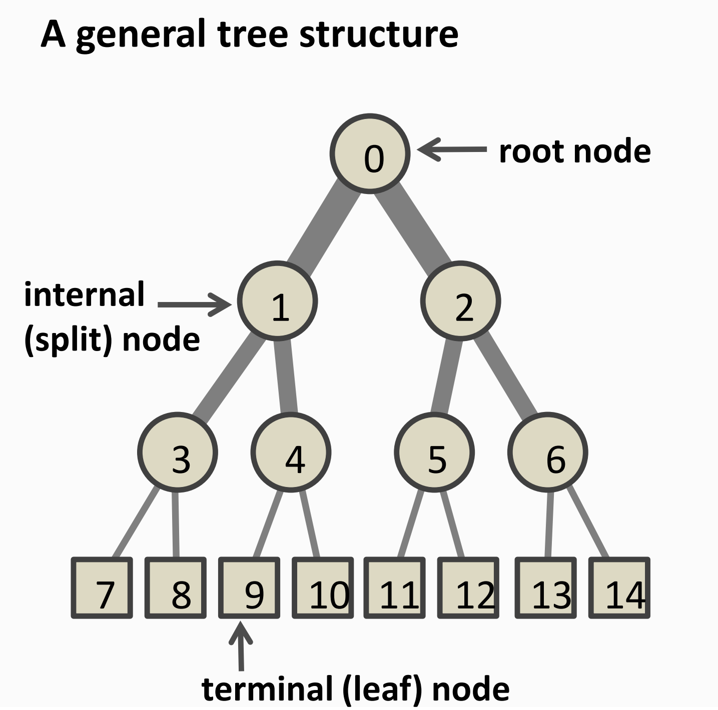
Connectionism

Learning and Inference



Learning and Inference





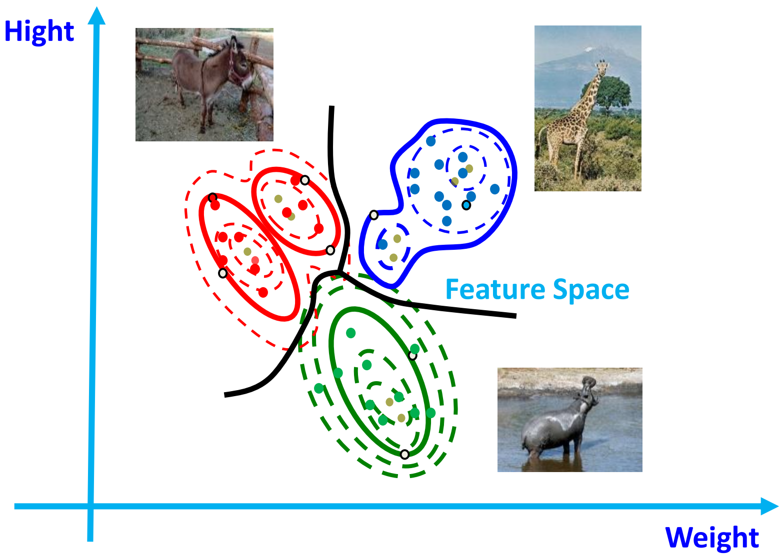
y	x					
y_1	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
y_2	x_{21}	x_{22}	x_{23}	x_{24}	x_{25}	x_{26}

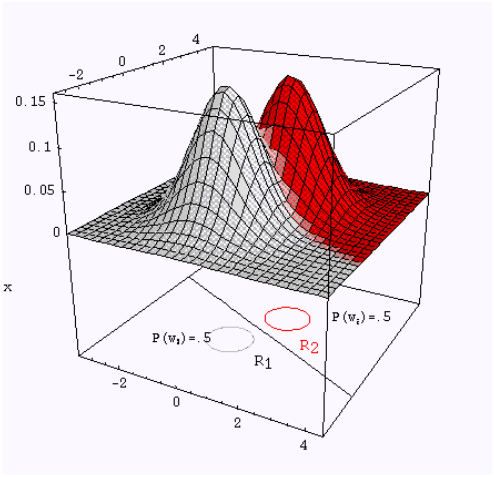
DB

Decision Tree

Symbolism

Learning and Inference

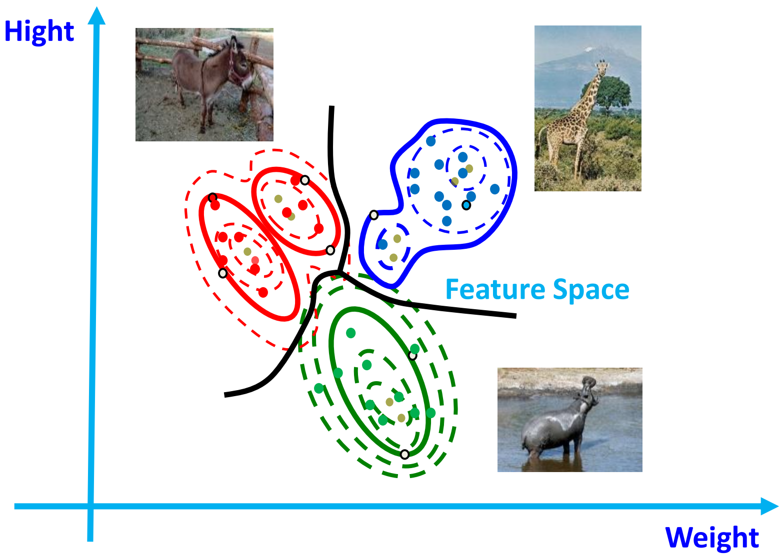




$$p(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x}-\boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu}) \right]$$

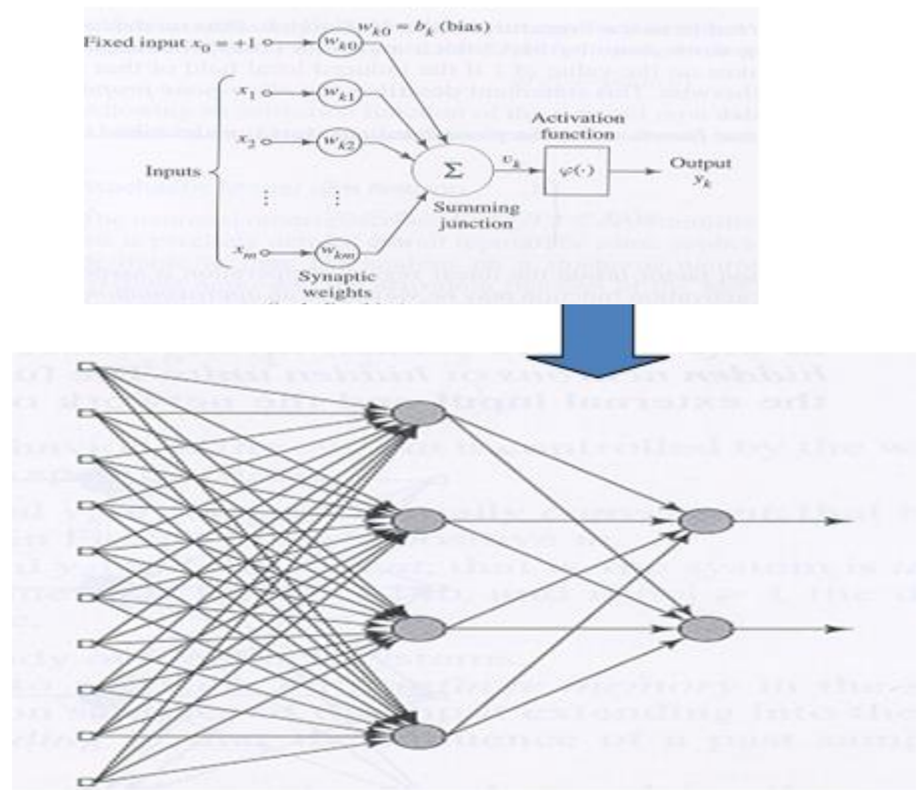
Bayesian Theory

Learning and Inference



Network Training

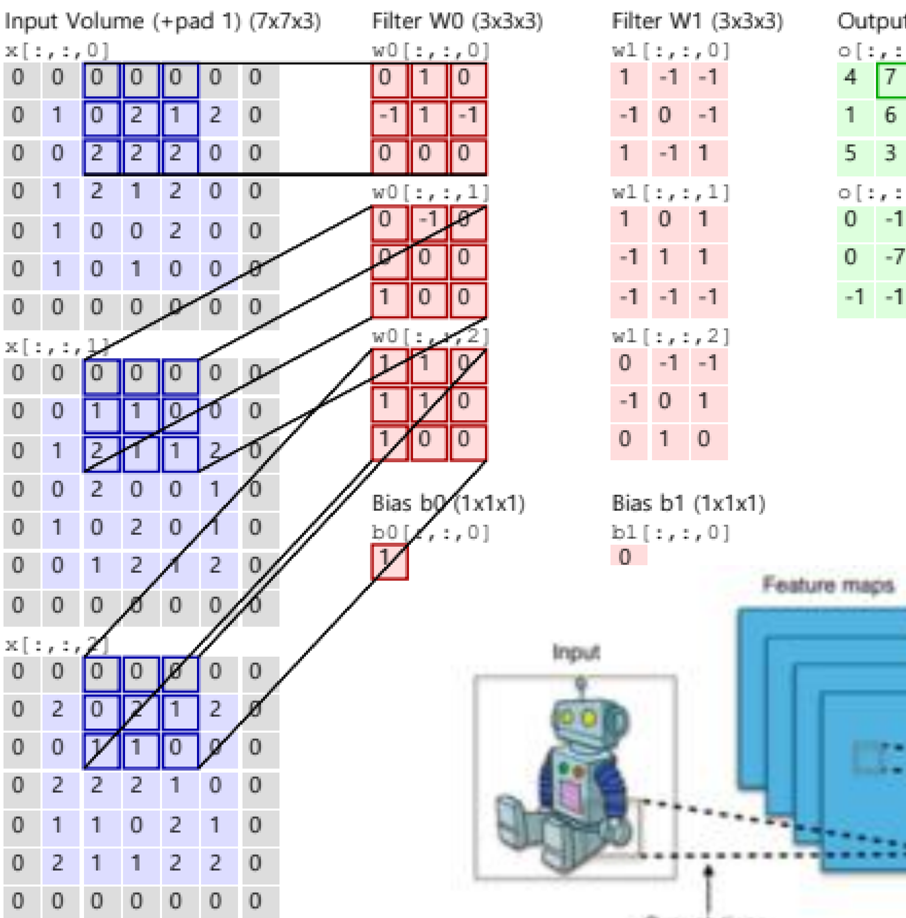
Inference (feedforward)

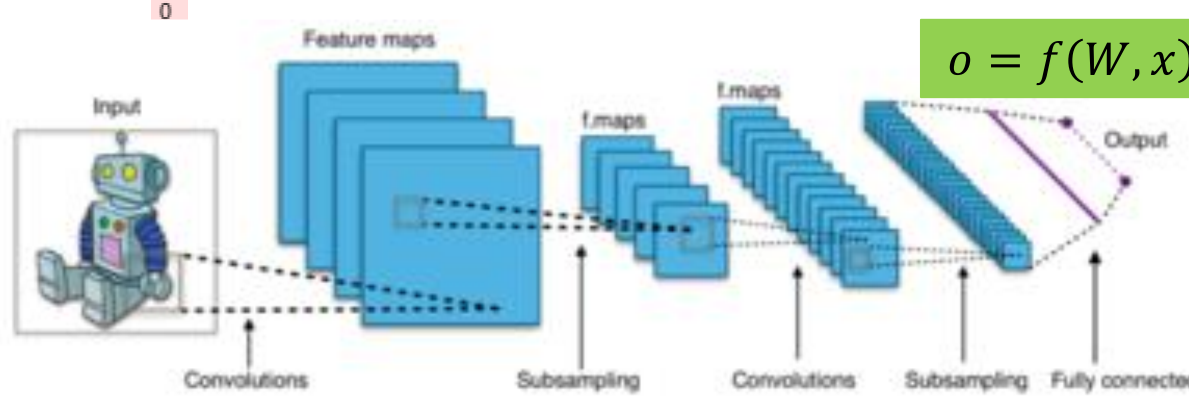


$$g_i(\mathbf{x}) = \mathbf{w}_i^t \mathbf{x} + w_{i0},$$

Connectionism

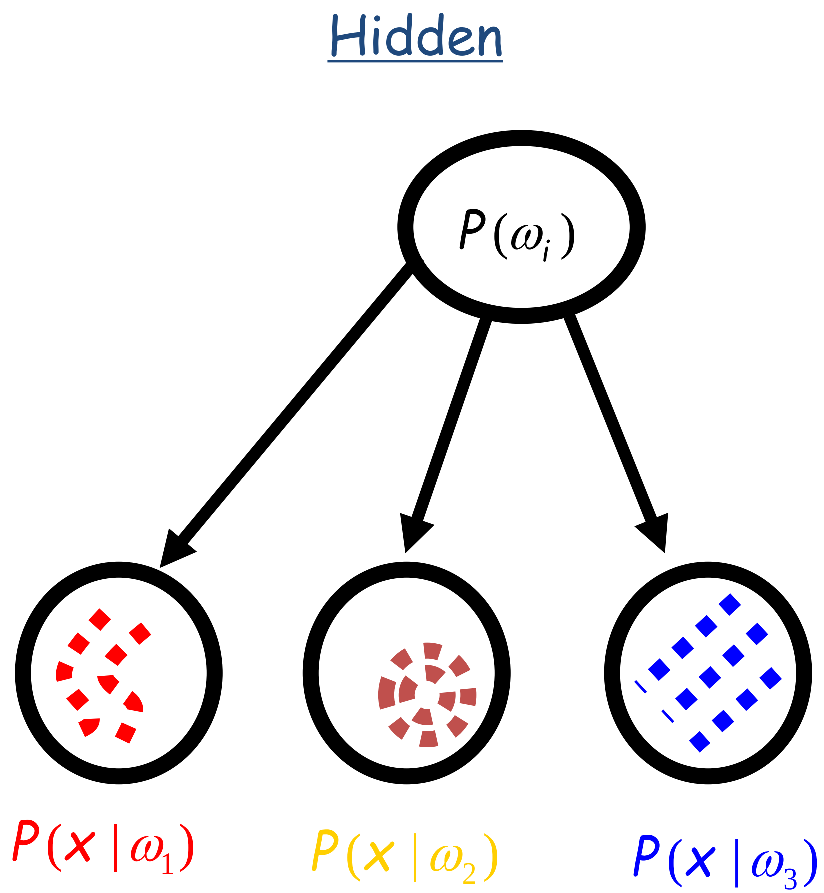
Convolutional Neural Networks

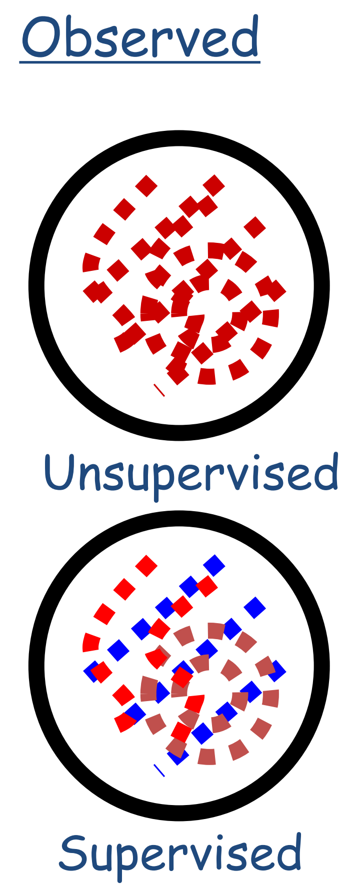




$$o = f(W, x)$$

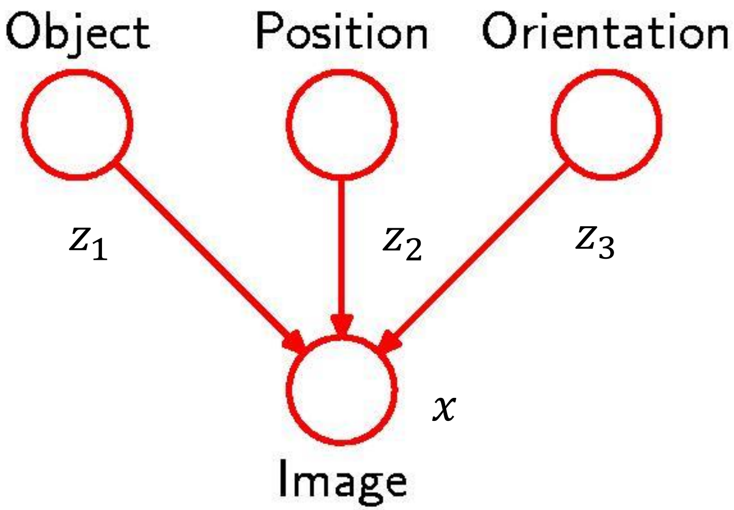
Supervised/Unsupervised Learning,





Generative/Discriminative Model

- Generating images



$$p(z_1, z_2, z_3|x)$$

Generative approach:

Model $p(z, x) = p(x|z)p(z)$

Use Bayes' theorem

$$p(z|x) = \frac{p(x|z)p(z)}{p(x)}$$

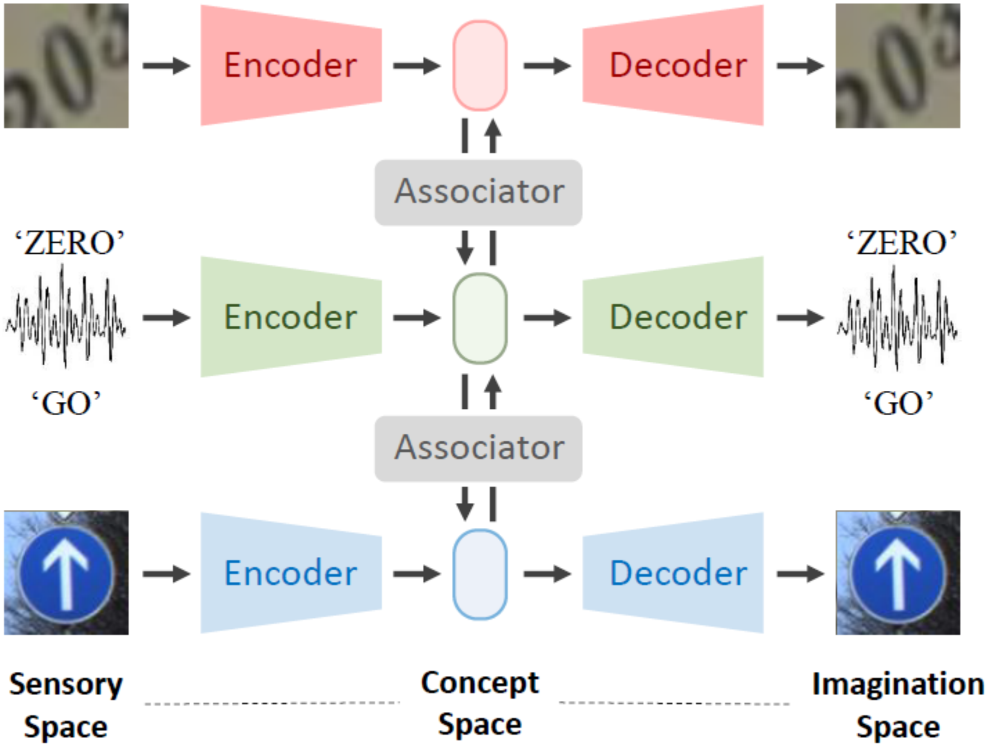
Discriminative approach:

Model $p(z|x)$ directly

Unsupervised Learning,

Clustering: K-means, etc.

Variational Auto-Encoder



Statistical Learning

L_2 Loss

$$L(d, f(W, x)) = \|d - f(W, x)\|_2^2$$

Total Loss

$$\mathcal{L}(W) = \int L(d, f(W, x))dp(x, d)$$

where $p(x, d)$ is a joint PDF of x and d, but unknown

Empirical Total Loss

$$\mathcal{L}_i(W) = \frac{1}{N}\sum_{n=1}^{N} L(d_n, f(W, x_n))$$

Statistical Learning

$$I(x_k) = \log(\frac{1}{p_k}) \rightarrow \begin{cases} base\ 2 \rightarrow bits \\ base\ e \rightarrow nats \end{cases}$$

32 bits $\rightarrow p_k = 1/2^{32}$ for uniform distribution $\rightarrow I_k = 32$ bits

① $I(x_k) = 0$ for $p_k = 1$

② $I(x_k) \geq 0$ for $0 \leq p_k \leq 1$

③ $I(x_k) \geq I(x_j)$ for $p_k \leq p_j$

Entropy : a measure of the *average amount of information conveyed per message*, i.e., expectation of Information

$$H(x) = E[I(x_k)] = \sum_k p_k I(x_k) = -\sum_k p_k \log p_k$$

Statistical Learning

Entropy becomes maximum when p_k is equiprobable

32 bits $\rightarrow p_k = 1/2^{32}$ for uniform distribution $\rightarrow I_k = 32$ bits

$$\rightarrow 0 \leq H(X) \leq -\sum_{k=1}^{2^{32}} \frac{1}{2^{32}} \log \frac{1}{2^{32}} = 32$$

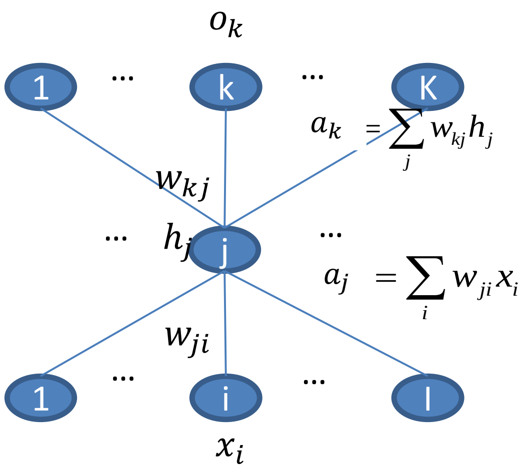
$$\rightarrow H(X) = 0 \quad \text{for an event } p_k = 1 \text{ and } p_{j \neq k} = 0$$

Cross Entropy Loss:

$$\mathcal{L}(\mathrm{W}) = -\sum_k^K t_k \log f_k(W, x)$$

$$f_k(W, x) = \frac{e^{a_k}}{\sum_j e^{a_j}} \text{ (softmax)}$$

t_k : target label (one hot: 0000100)



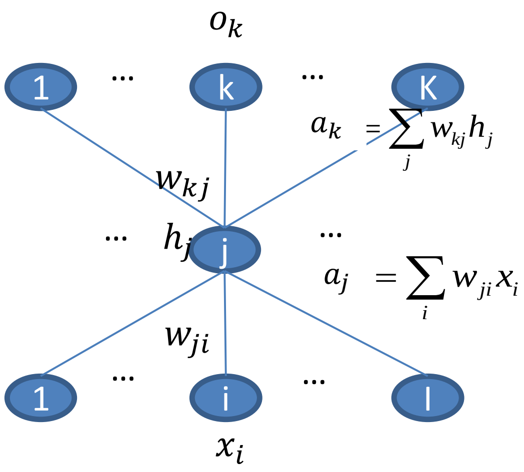
Statistical Learning

Cross Entropy Loss:

$$\mathcal{L}(\mathrm{W}) = -\sum_{k}^{K}[t_k \log f_k(W,x) + (1-t_k)\log(1-f_k(W,x))]$$

$$f_k(W,x) = \frac{1}{1+e^{-a_k}}\text{(sigmoid)}$$

t_k : target label (multi-hot: 00110100)



Statistical Learning

Theorem (Gray 1990)

$$\sum_k p_k \log \frac{p_k}{q_k} \geq 0$$

Relative entropy (or Kullback – Leibler divergence)

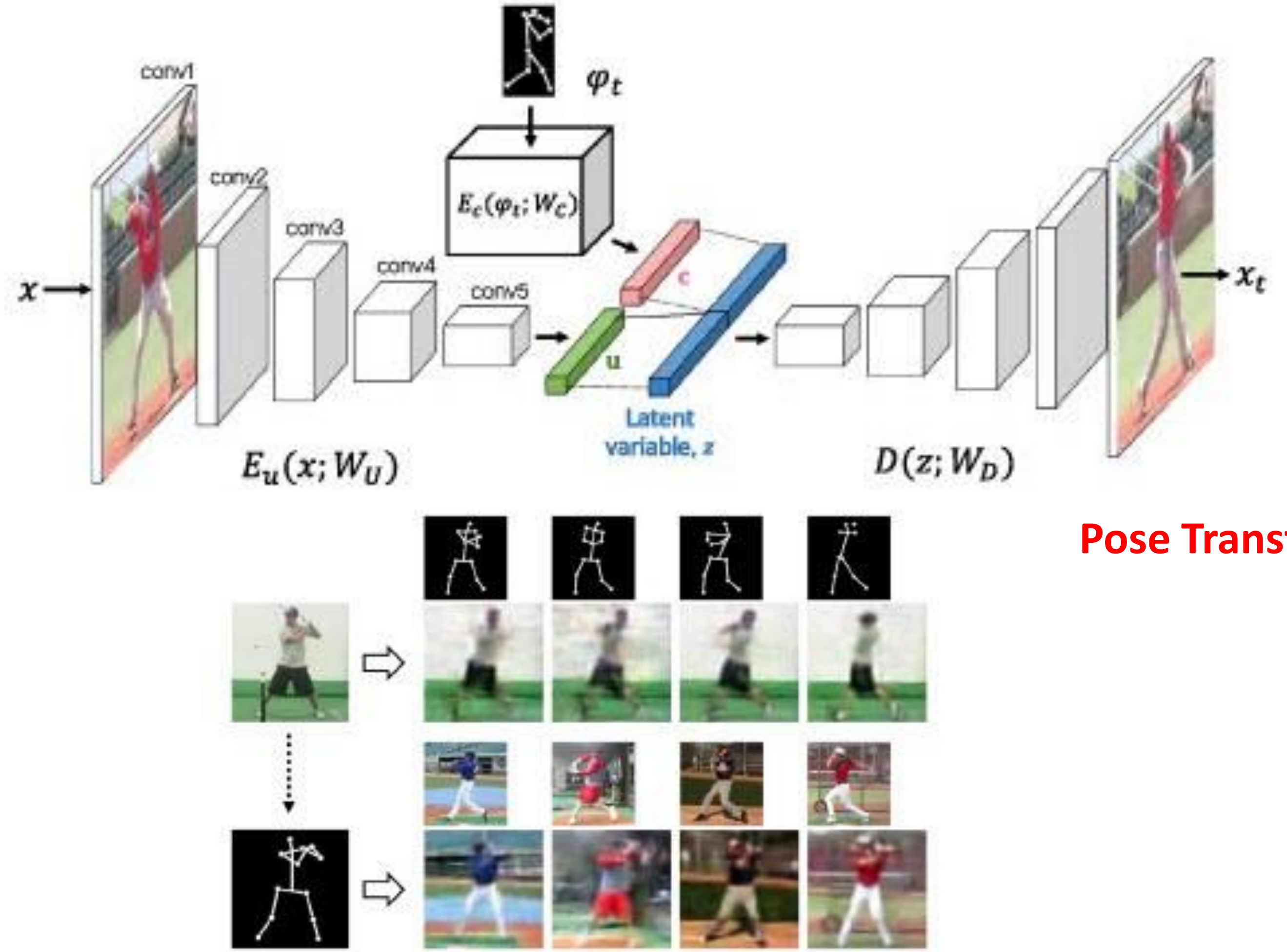
$$D_{KL}(p \parallel q) = \sum_k p_k \log \frac{p_k}{q_k}$$

$D_{KL}(p \parallel q) = 0$ for $p \equiv q$

p_k probability mass function

q_k reference probability mass function

Scene and Object Generation



Pose Transformer

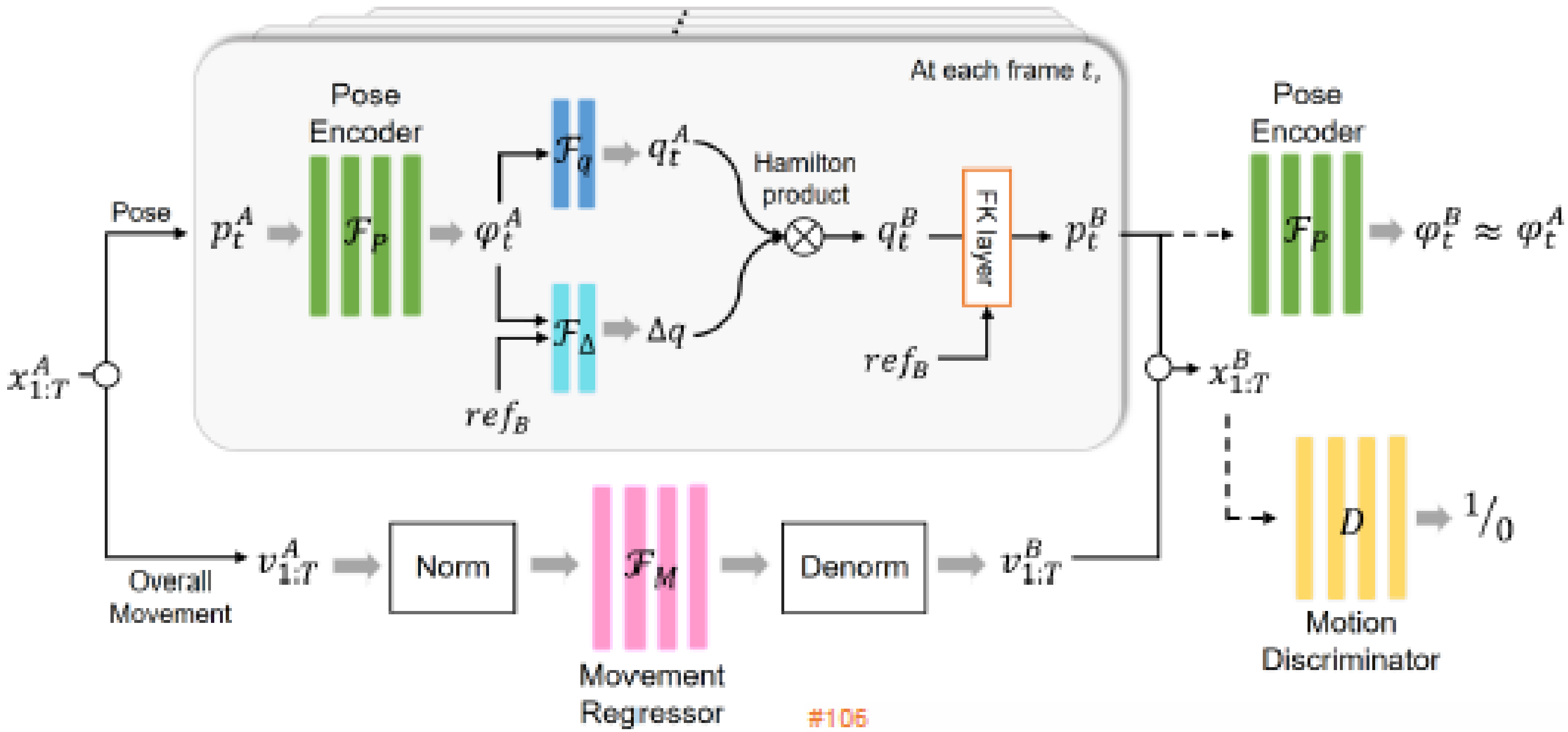
$$\mathcal{L}(\theta, \phi) = \mathcal{L}_{ref} + \mathcal{L}_{pose} + \mathcal{L}_{id}.$$

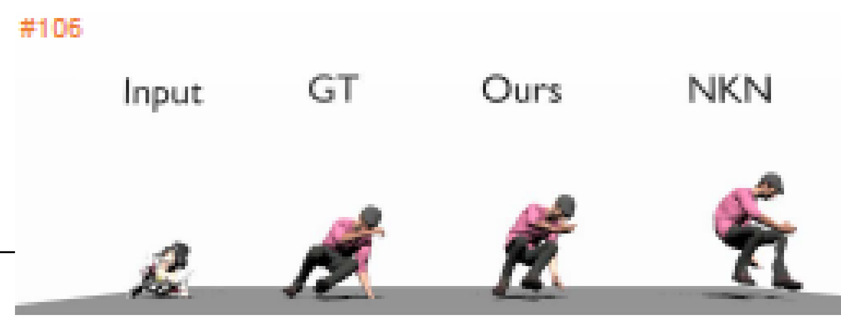
$$\mathcal{L}_{ref} = \mathbb{E}_{q_\phi(z|x_a^k, \varphi_a^k)}[\log p_\theta(x_a^k|z)] - D_{KL}\left(q_\phi(z|x_a^k, \varphi_a^k) \parallel p_\theta(z)\right).$$

$$\mathcal{L}_{pose} = \mathbb{E}_{q_\phi(z|x_a^k, \varphi_a^l)}[\log p_\theta(x_a^l|z)] - D_{KL}\left(q_\phi(z|x_a^k, \varphi_a^l) \parallel p_\theta(z)\right) - \lambda_u \cdot D_{KL}\left(q_\phi(u|x_a^l) \parallel q_\phi(u|x_a^k)\right).$$

$$\mathcal{L}_{id} = \mathbb{E}_{q_\phi(z|x_b^{k'}, \varphi_a^k)}[\log p_\theta(x_b^{k'}|z)] - D_{KL}\left(q_\phi(z|x_b^{k'}, \varphi_a^k) \parallel p_\theta(z)\right) - \lambda_c \cdot D_{KL}\left(q_\phi(c|\varphi_b^{k'}) \parallel q_\phi(c|\varphi_a^k)\right)$$

Motion Retargeting





Outline of ML techniques

Bayes Rule | Likelihood | Posteriori | Priori | Bayes Decision

Learning		Generative Model		Discriminative Model		Learning			
ML(P)E	Bayes. L	GM		Lin. Classifier	SVM	LS	Convex O.		
Histogram		K-NN	Parzen W.	Random Forest		Entropy			
EM, MCMC		GMM		K-SVM	K-SVDD	Convex O.			
MCMC, VI		Bayesian Net		Deep NN		SA	GA	BP(GD)	NM
MLE	VI	Boltzm. Machine		ICA		K-L Divergence			
MCMC, VI		Latent DA		Linear DA		Max. Separa.			
EM		HMM		PCA		Max. Scatter			

Course Outline

Intro. AI, ML, and DL
Intro. Linear Algebra
Intro. Prob. & Information
Bayesian Decision Theory
Dim reduction PCA & LDA
Learning Rules
Support Vector Machine
Deep Convolutional Networks
Bayesian Networks
Parametric pdf Estimation
Non-Parametric pdf Estimation

Boltzman Machine
Markov Chain Monte Carlo
Inference of Bayesian Net, MCMC
Inference of Bayesian Net, VI
Traffic Pattern Analysis, VI
Recent Papers
- Active Learning
- Imbalanced Data Learning
- Out of Distribution
- Weakly Supervised Learning
- Etc.

Questions

1. Describe the common things and differences among symbolism, connectionism, and Bayesian approach.
2. Explain supervised/weekly-supervised/unsupervised learning.
3. What is the difference between discriminative and generative model?