Infinite Restricted Boltzmann Machine

Marc-Alexandre Côté
Université de Sherbrooke

Joint work with Hugo Larochelle

A RESEARCH AGENDA

- Deep learning successes have required a lot of labeled training data
  - collecting and labeling such data requires significant human labor
  - is that really how we'll solve AI ?
- Alternative solution : exploit other sources of data that are imperfect but plentiful
  - unlabeled data (unsupervised learning)
  - multimodal data (multimodal learning)
  - multidomain data (transfer learning, domain adaptation)

A RESEARCH AGENDA

- By far the largest source is unlabeled data
  - effectively requires algorithms for *life-long learning*

- We are currently poorly equipped to deal with this setting
  - how to do online learning for non-convex models, with a changing input distribution?
  - how to have models whose capacity adapts during training?

A RESEARCH AGENDA

- By far the largest source is unlabeled data
  - effectively requires algorithms for *life-long learning*

- We are currently poorly equipped to deal with this setting
  - how to do online learning for non-convex models, with a changing input distribution?
  - how to have models whose capacity adapts during training?

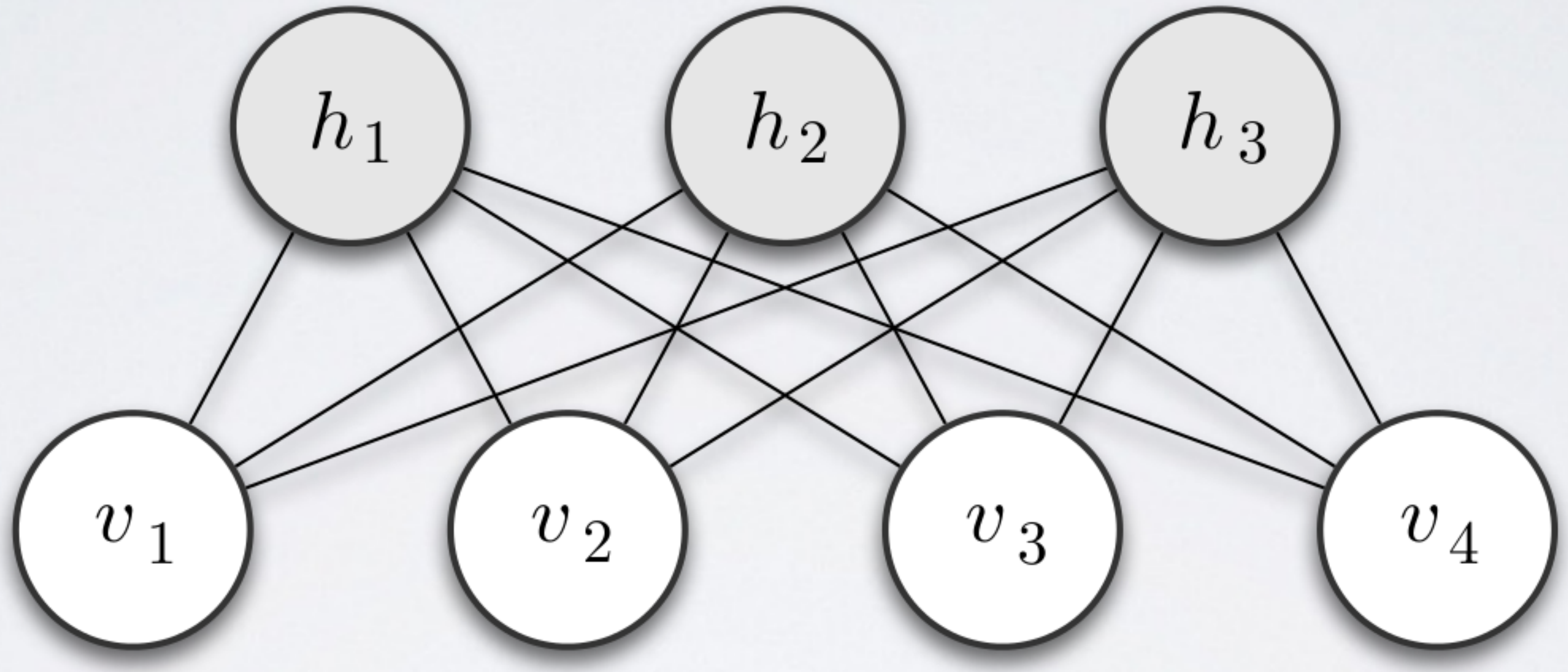
A RESEARCH AGENDA

- By far the largest source is unlabeled data
  - effectively requires algorithms for *life-long learning*

- We are currently poorly equipped to deal with this setting
  - how to do online learning for non-convex models, with a changing input distribution?
  - how to have models whose capacity adapts during training?

- In this talk: an infinite restricted Boltzmann machine (iRBM)
  - RBM with capacity that can grow during training
  - growing mechanism is derived naturally from the energy function definition

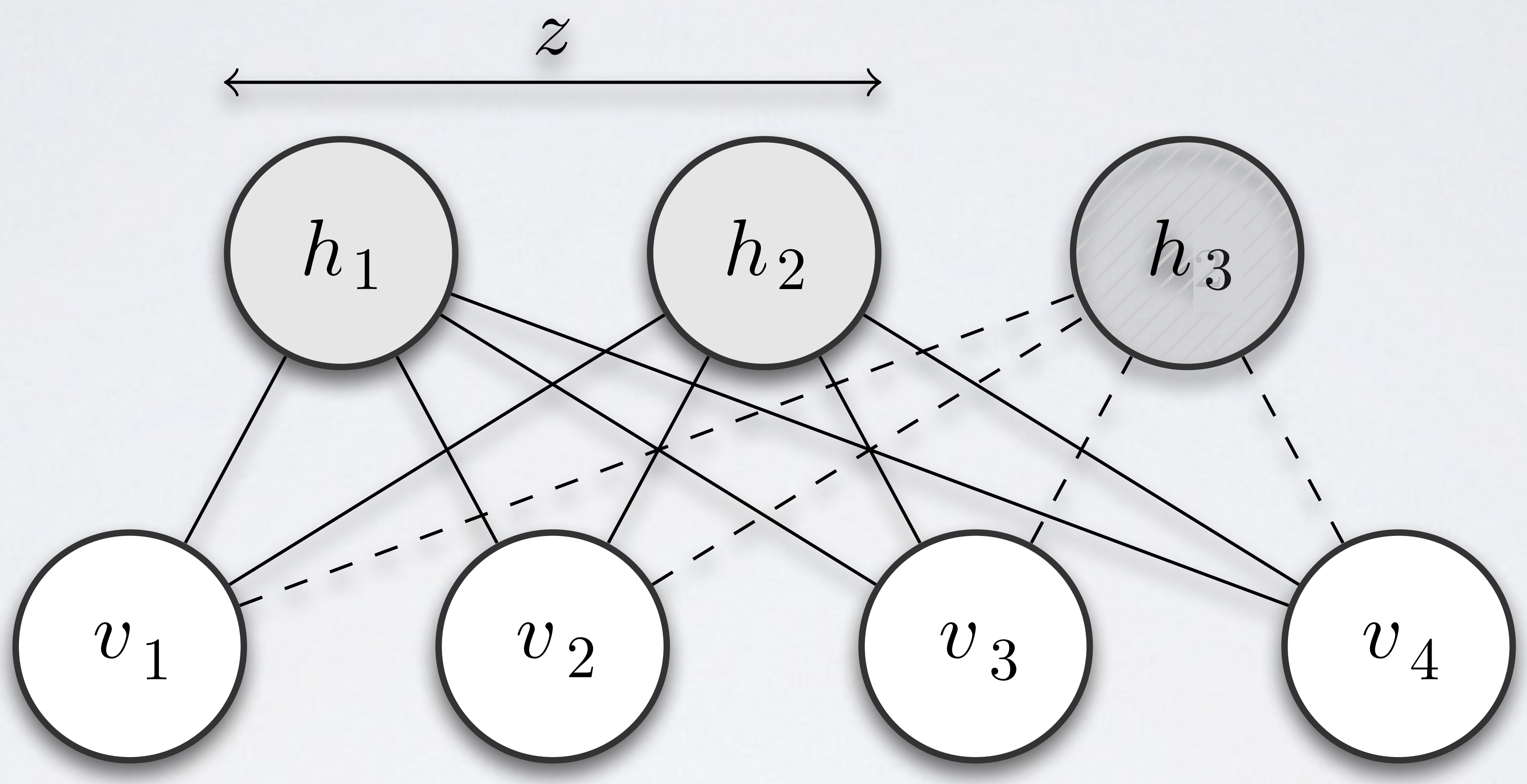
RESTRICTED BOLTZMANN MACHINE



$$\mathrm{E}(\mathbf{v},\mathbf{h}) = -\mathbf{h}^T\mathbf{W}\mathbf{v} - \mathbf{v}^T\mathbf{b}^{\mathrm{v}} - \mathbf{h}^T\mathbf{b}^{\mathrm{h}}$$

$$P(\mathbf{v},\mathbf{h}) = e^{-\mathrm{E}(\mathbf{v},\mathbf{h})} / Z$$

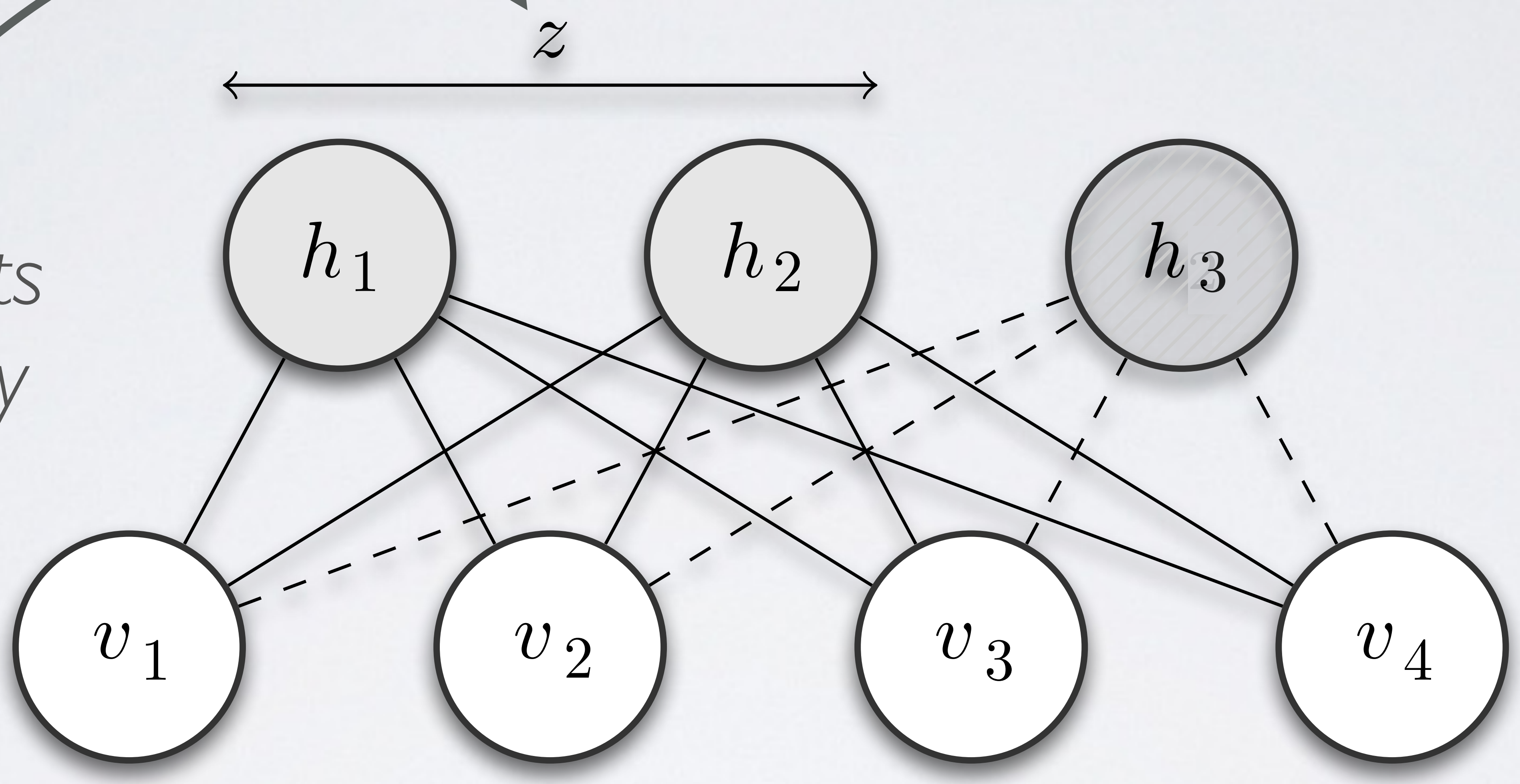
ORDERED RESTRICTED BOLTZMANN MACHINE



$z \in \{1, \ldots, H\}$

$$\mathbf{E}(\mathbf{v}, \mathbf{h}, z) = -\mathbf{v}^T \mathbf{b}^{\text{v}} - \sum_{i=1}^{z} h_i(\mathbf{W}_{i\cdot}\mathbf{v} + b_i^{\text{h}}) - \beta_i$$

ORDERED RESTRICTED BOLTZMANN MACHINE

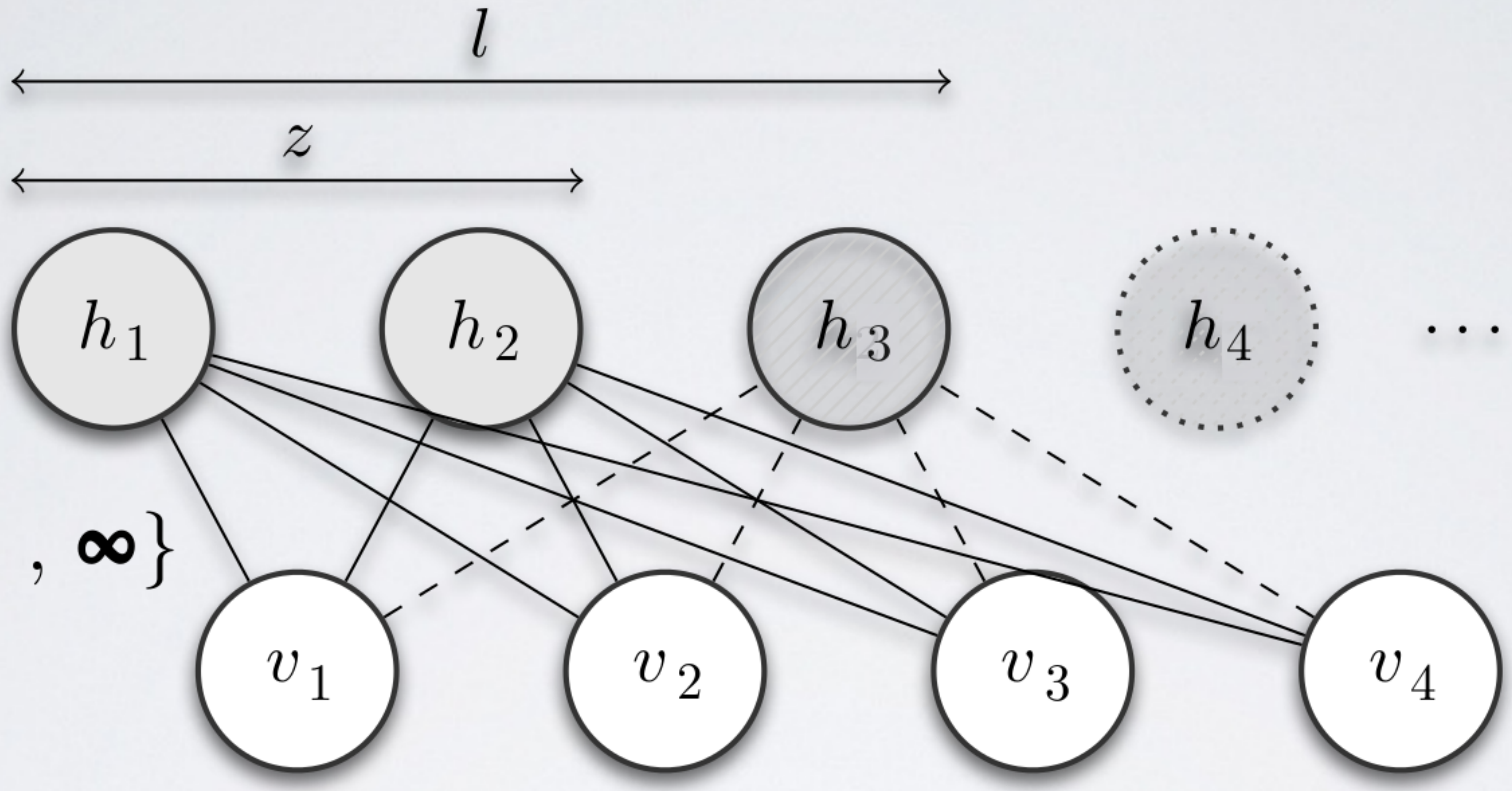


Nb of hidden units involved in energy

$z \in \{1, \dots, H\}$

$$\mathbf{E}(\mathbf{v}, \mathbf{h}, z) = -\mathbf{v}^T \mathbf{b}^{\mathrm{v}} - \sum_{i=1}^{z} h_i(\mathbf{W}_{i\cdot}\mathbf{v} + b_i^{\mathrm{h}}) - \beta_i$$

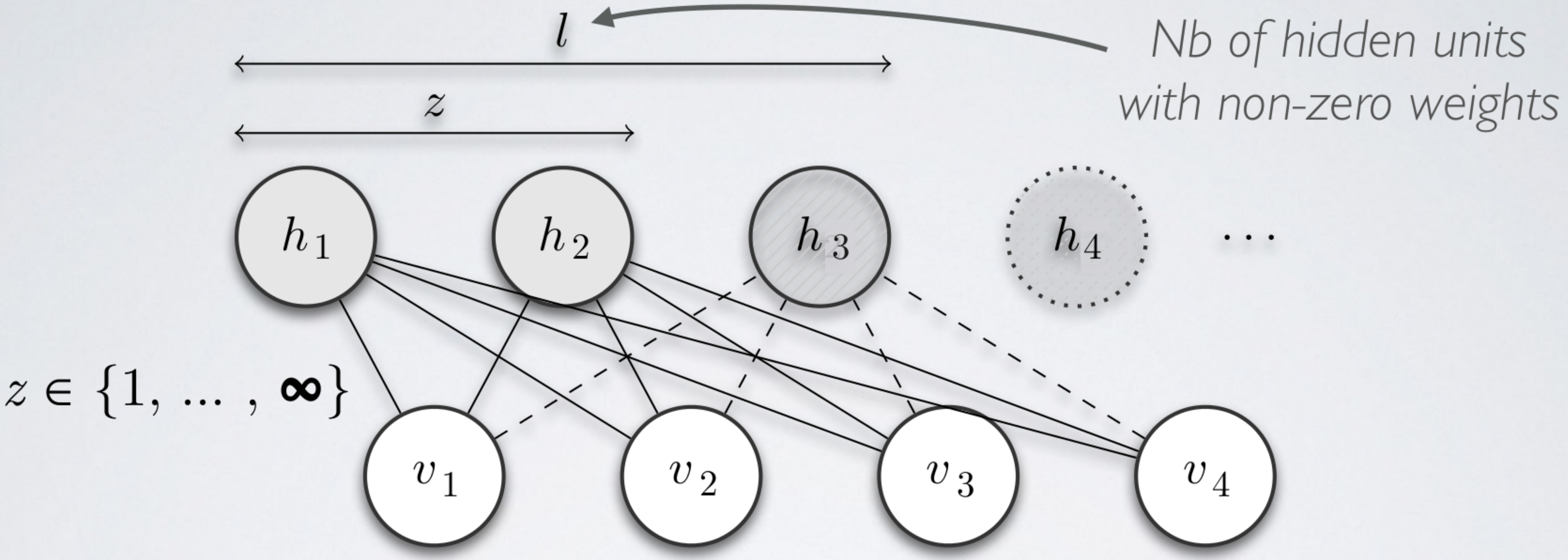
INFINITE RESTRICTED BOLTZMANN MACHINE



$z \in \{1, \ldots, \boldsymbol{\infty}\}$

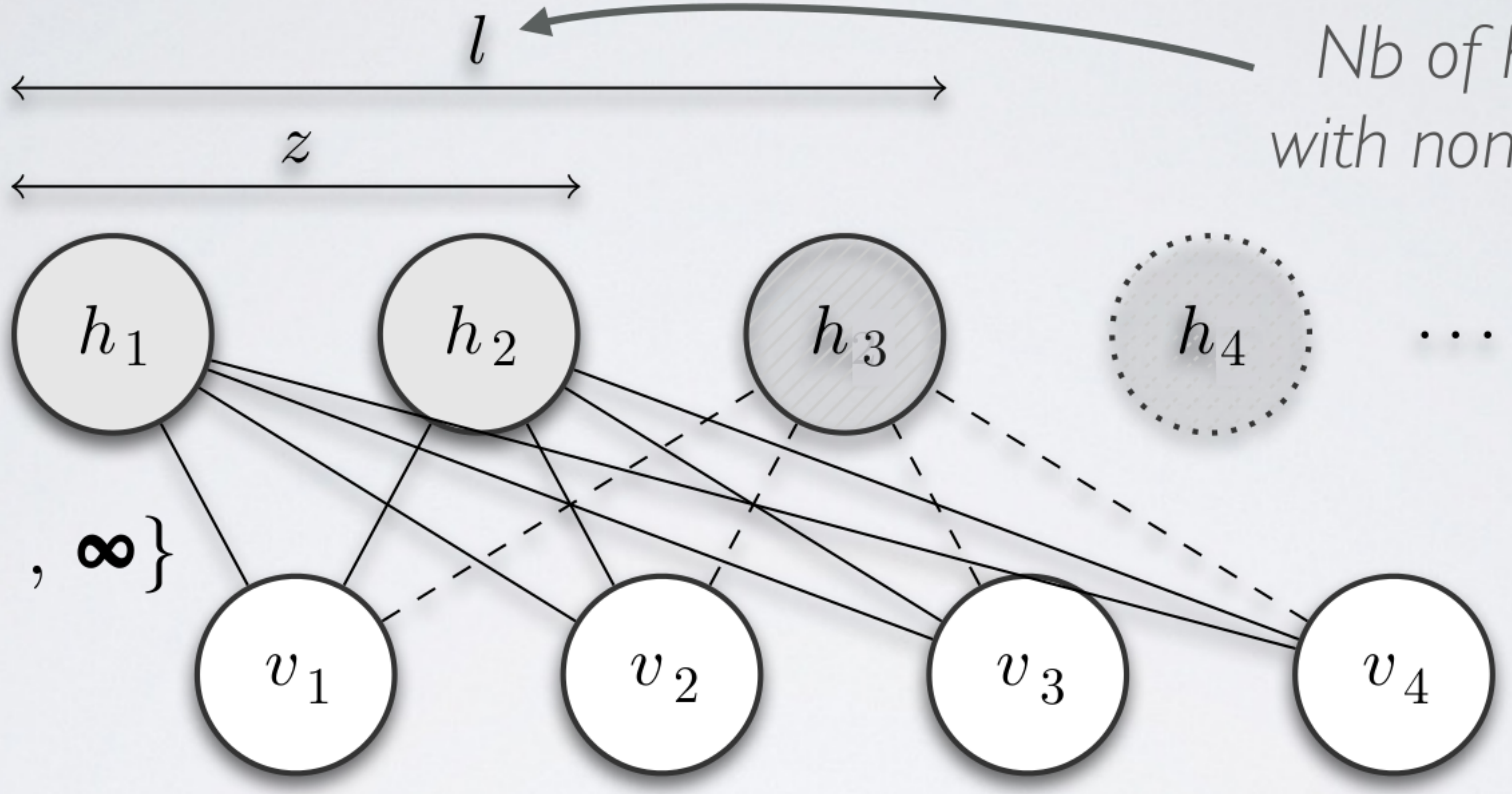
$$\mathbf{E}(\mathbf{v}, \mathbf{h}, z) = -\mathbf{v}^T \mathbf{b}^{\mathrm{v}} - \sum_{i=1}^{z} h_i(\mathbf{W}_{i\cdot}\mathbf{v} + b_i^{\mathrm{h}}) - \beta_i$$

INFINITE RESTRICTED BOLTZMANN MACHINE



$$\mathbf{E}(\mathbf{v}, \mathbf{h}, z) = -\mathbf{v}^T \mathbf{b}^{\mathrm{v}} - \sum_{i=1}^{z} h_i(\mathbf{W}_{i\cdot}\mathbf{v} + b_i^{\mathrm{h}}) - \beta_i$$

INFINITE RESTRICTED BOLTZMANN MACHINE



Nb of hidden units with non-zero weights

$z \in \{1, \ldots, \boldsymbol{\infty}\}$

Chosen such that infinite sum converges

$$\mathbf{E}(\mathbf{v}, \mathbf{h}, z) = -\mathbf{v}^T \mathbf{b}^{\mathrm{v}} - \sum_{i=1}^{z} h_i(\mathbf{W}_{i \cdot} \mathbf{v} + b_i^{\mathrm{h}}) - \beta_i$$

INFINITE RESTRICTED BOLTZMANN MACHINE

- Free energy, given a certain value of z

$$F(\mathbf{v}, z) = -\mathbf{v}^T \mathbf{b}^{\mathrm{v}} - \sum_{i=1}^{z} \mathrm{soft}_+(\mathbf{W}_{i\cdot}\mathbf{v} + b_i^{\mathrm{h}}) - \beta_i$$

- For the iRBM to be practical, we must be able to compute

$$P(z|\mathbf{v}) = \frac{\exp(-F(\mathbf{v}, z))}{Z(\mathbf{v})} = \frac{\exp(-F(\mathbf{v}, z))}{\sum_{z'}^{\infty} \exp(-F(\mathbf{v}, z'))}$$

INFINITE RESTRICTED BOLTZMANN MACHINE

$$Z(\mathbf{v}) = \sum_{z=1}^{l} \exp(-F(\mathbf{v}, z)) + \sum_{z=l+1}^{\infty} \exp(-F(\mathbf{v}, z))$$

$$= \sum_{z=1}^{l} \exp(-F(\mathbf{v}, z)) + \sum_{z=l+1}^{\infty} \exp\left(-F(\mathbf{v}, l) + \sum_{i=l+1}^{z} \text{soft}_+(\mathbf{W}_{i\cdot}\mathbf{v} + b_i^{\mathbf{h}}) - \beta_i\right)$$

INFINITE RESTRICTED BOLTZMANN MACHINE

$$\beta_i = \beta \text{soft}_+(b_i^{\text{h}})$$

$$\beta > 1$$

$$Z(\mathbf{v}) = \sum_{z=1}^{l} \exp(-F(\mathbf{v}, z)) + \sum_{z=l+1}^{\infty} \exp(-F(\mathbf{v}, z))$$

$$= \sum_{z=1}^{l} \exp(-F(\mathbf{v}, z)) + \sum_{z=l+1}^{\infty} \exp\left(-F(\mathbf{v}, l) + \sum_{i=l+1}^{z} \text{soft}_+(\mathbf{W}_{i\cdot}\mathbf{v} + b_i^{\text{h}}) - \beta_i\right)$$

INFINITE RESTRICTED BOLTZMANN MACHINE

$$\beta_i = \beta \mathrm{soft}_+(b_i^{\mathrm{h}})$$

$$\beta > 1$$

$$Z(\mathbf{v}) = \sum_{z=1}^{l} \exp(-F(\mathbf{v}, z)) + \sum_{z=l+1}^{\infty} \exp(-F(\mathbf{v}, z))$$

$$= \sum_{z=1}^{l} \exp(-F(\mathbf{v}, z)) + \sum_{z=l+1}^{\infty} \exp\left(-F(\mathbf{v}, l) + \sum_{i=l+1}^{z} \mathrm{soft}_+(\mathbf{W}_{i\cdot}\mathbf{v} + b_i^{\mathrm{h}}) - \beta_i\right)$$

$$= \sum_{z=1}^{l} \exp(-F(\mathbf{v}, z)) + \exp(-F(\mathbf{v}, l)) \sum_{z=l+1}^{\infty} \exp\left(\sum_{i=l+1}^{z} (1-\beta)\mathrm{soft}_+(0)\right)$$

INFINITE RESTRICTED BOLTZMANN MACHINE

$$\beta_i = \beta \mathrm{soft}_+(b_i^{\mathrm{h}})$$

$$\beta > 1$$

$$Z(\mathbf{v}) = \sum_{z=1}^{l} \exp(-F(\mathbf{v},z)) + \sum_{z=l+1}^{\infty} \exp(-F(\mathbf{v},z))$$

$$= \sum_{z=1}^{l} \exp(-F(\mathbf{v},z)) + \sum_{z=l+1}^{\infty} \exp\left(-F(\mathbf{v},l) + \sum_{i=l+1}^{z} \mathrm{soft}_+(\mathbf{W}_{i\cdot}\mathbf{v} + b_i^{\mathrm{h}}) - \beta_i\right)$$

$$= \sum_{z=1}^{l} \exp(-F(\mathbf{v},z)) + \exp(-F(\mathbf{v},l)) \sum_{z=l+1}^{\infty} \exp\left(\sum_{i=l+1}^{z} (1-\beta)\mathrm{soft}_+(0)\right)$$

$$= \sum_{z=1}^{l} \exp(-F(\mathbf{v},z)) + \exp(-F(\mathbf{v},l)) \sum_{z=1}^{\infty} \exp((1-\beta)\mathrm{soft}_+(0))^{z}$$

INFINITE RESTRICTED BOLTZMANN MACHINE

$$\beta_i = \beta \mathrm{soft}_+(b_i^{\mathrm{h}})$$

$$\beta > 1$$

$$Z(\mathbf{v}) = \sum_{z=1}^{l} \exp(-F(\mathbf{v}, z)) + \sum_{z=l+1}^{\infty} \exp(-F(\mathbf{v}, z))$$

$$= \sum_{z=1}^{l} \exp(-F(\mathbf{v}, z)) + \sum_{z=l+1}^{\infty} \exp\left(-F(\mathbf{v}, l) + \sum_{i=l+1}^{z} \mathrm{soft}_+(\mathbf{W}_{i\cdot}\mathbf{v} + b_i^{\mathrm{h}}) - \beta_i\right)$$

$$= \sum_{z=1}^{l} \exp(-F(\mathbf{v}, z)) + \exp(-F(\mathbf{v}, l)) \sum_{z=l+1}^{\infty} \exp\left(\sum_{i=l+1}^{z} (1-\beta)\mathrm{soft}_+(0)\right)$$

$$= \sum_{z=1}^{l} \exp(-F(\mathbf{v}, z)) + \exp(-F(\mathbf{v}, l)) \underbrace{\sum_{z=1}^{\infty} \exp((1-\beta)\mathrm{soft}_+(0))^z}_{\text{Geometric series}}$$

INFINITE RESTRICTED BOLTZMANN MACHINE

- Can perform Gibbs sampling
  - before sampling $\mathbf{h}$, first sample z given $\mathbf{v}$

$$P(z|\mathbf{v}) = \frac{\exp(-F(\mathbf{v}, z))}{Z(\mathbf{v})}$$

$$P(h_i = 1|\mathbf{v}, z) = \begin{cases} \sigma(\mathbf{W}_{i\cdot}\mathbf{v} + b_i^{\mathrm{h}}) & \text{if } i \leq z \\ 0 & \text{otherwise} \end{cases}$$

$$P(v_j = 1|\mathbf{h}, z) = \sigma\left(\sum_{i=1}^{z} W_{ij} h_i + b_j^{\mathrm{v}}\right)$$

INFINITE RESTRICTED BOLTZMANN MACHINE

- Can perform Gibbs sampling
  - before sampling $\mathbf{h}$, first sample z given $\mathbf{v}$

- If can perform Gibbs sampling, can perform contrastive divergence CD training
  - Gibbs sampling provides negative samples for update

INFINITE RESTRICTED BOLTZMANN MACHINE

- Can perform Gibbs sampling
  - before sampling $\mathbf{h}$, first sample z given $\mathbf{v}$
- If can perform Gibbs sampling, can perform contrastive divergence CD training
  - Gibbs sampling provides negative samples for update
- CD training is well defined
  - only selected hidden units get a non-zero gradient on their weights
  - any amount of regularization ensures that training does not diverge to infinitely many units with non-zero weights

Training of iRBM

EXPERIMENTS

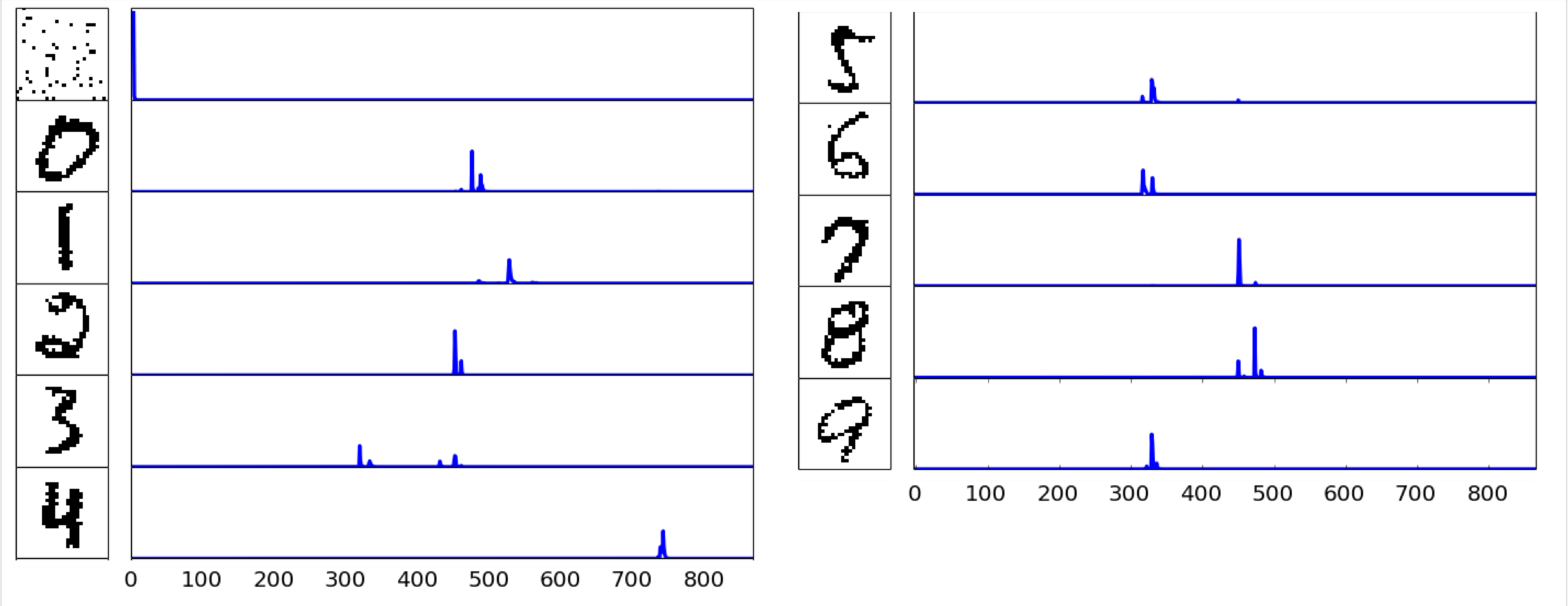
- Binarized MNIST and CalTech101 Silhouettes

	Binarized MNIST			CalTech101 Silhouettes		
Model	Size	$\ln(\hat{Z} \pm 3\sigma)$	Avg. NLL	Size	$\ln(\hat{Z} \pm 3\sigma)$	Avg. NLL
RBM	100	[600.88, 600.95]	98.17 ± 0.52	100	[2511.62, 2512.56]	177.37 ± 2.81
RBM	500	[613.24, 613.31]	86.50 ± 0.44	500	[2385.68, 2386.10]	119.05 ± 2.27
RBM	2000	[1098.94, 1099.17]	85.03 ± 0.42	2000	[3349.85, 3354.15]	118.29 ± 2.25
oRBM	500	[39.90, 40.19]	88.15 ± 0.46	500	[1782.88 1783.02]	114.99 ± 1.97
iRBM	1208	[40.03, 40.54]	85.65 ± 0.44	915	[1999.93, 2000.22]	121.47 ± 2.07

- Use adagrad for training
- Training robust to value of β
  - used 1.01 in all experiments

EXPERIMENTS

- Distribution $p(z\,|\,\mathbf{v})$ (Binarized MNIST)



FUTURE WORK

- Can be extended to other types of representations
  - feed-forward neural networks
  - RBM with softmax units, for quantization-based fast search
  - RBM with tree-based representations, for hierarchical topic modeling
  - word representations (infinite Skip-Gram, Nalisnick and Ravi, 2015)

http://github.com/MarcCote/iRBM
http://arxiv.org/abs/1502.02476

Côté & Larochelle (2016) **An Infinite Restricted Boltzmann Machine**. *Neural Computation*

MERCI !

http://github.com/MarcCote/iRBM
http://arxiv.org/abs/1502.02476

Côté & Larochelle (2016) **An Infinite Restricted Boltzmann Machine**. *Neural Computation*