Factored 3-Way Restricted Boltzmann Machines for Modeling Natural Images

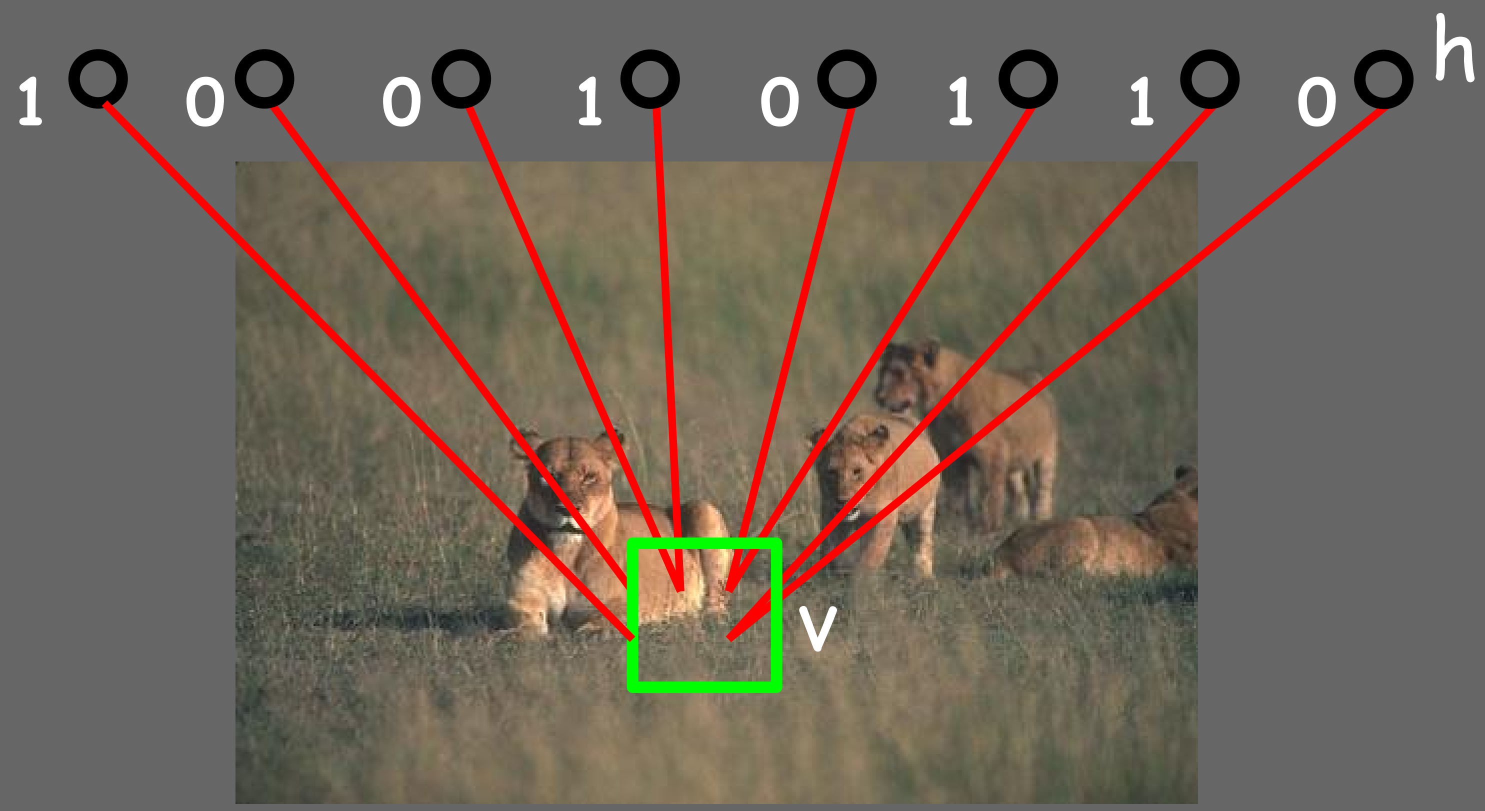
Marc'Aurelio Ranzato, Alex Krizhevsky and Geoff Hinton
Dept. of Computer Science – University of Toronto

- Want to model natural images by using a generative model

$$p(\text{image } v; W)$$

- Want to use the model to produce representations

$$p(\text{image } v, \text{hidden units } h; W)$$



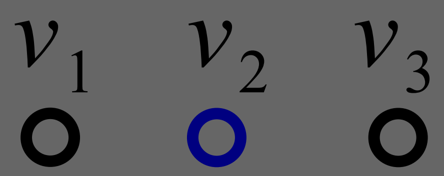
Goal: define $p(v, h)$

Q.: What is the key property of natural images?

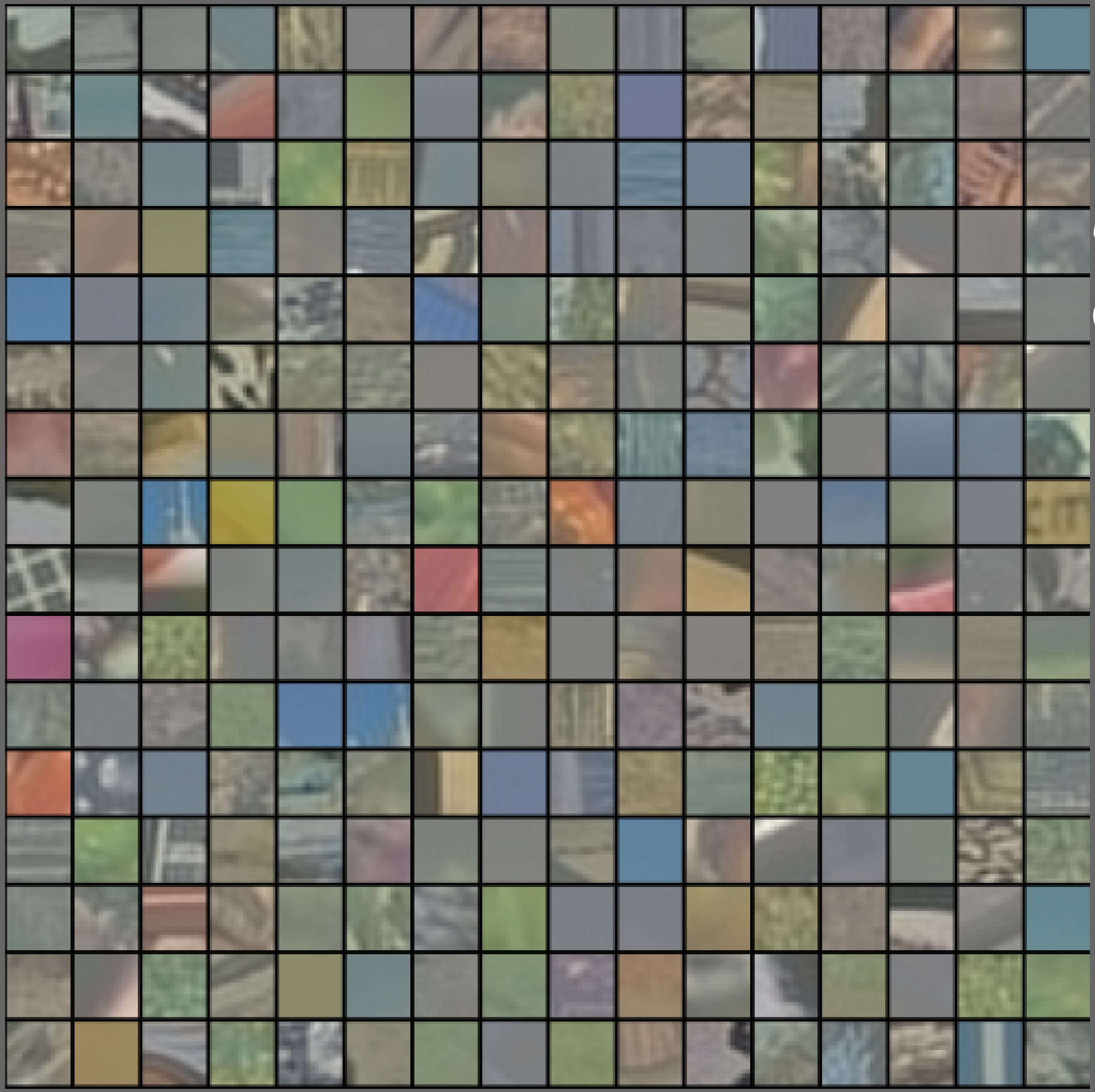
Goal: define $p(v,h)$

Q.: What is the key property of natural images?

A.: smoothness

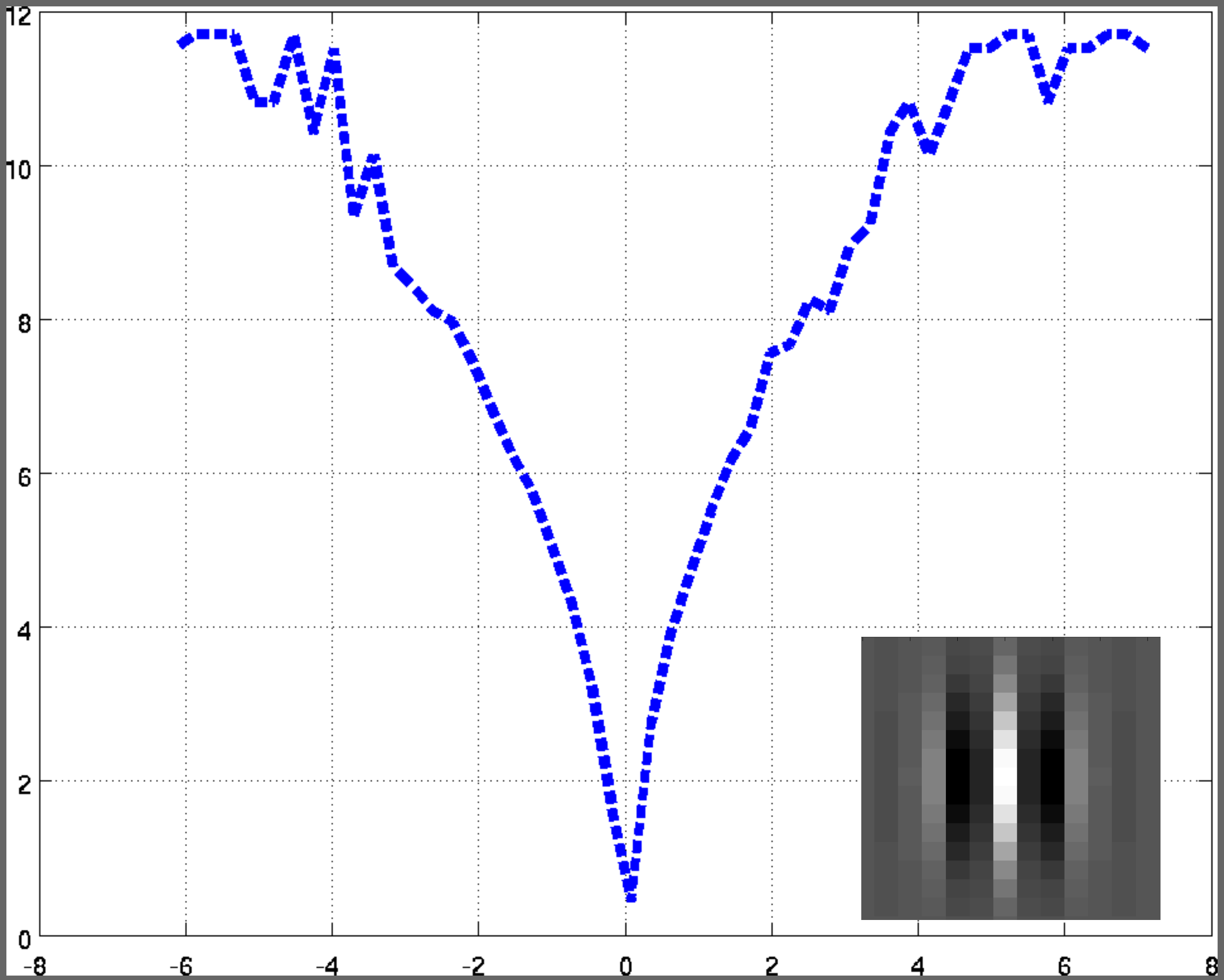


Most of the times, the value of one pixel can be well predicted from its neighbors



Most local neighborhoods are smooth

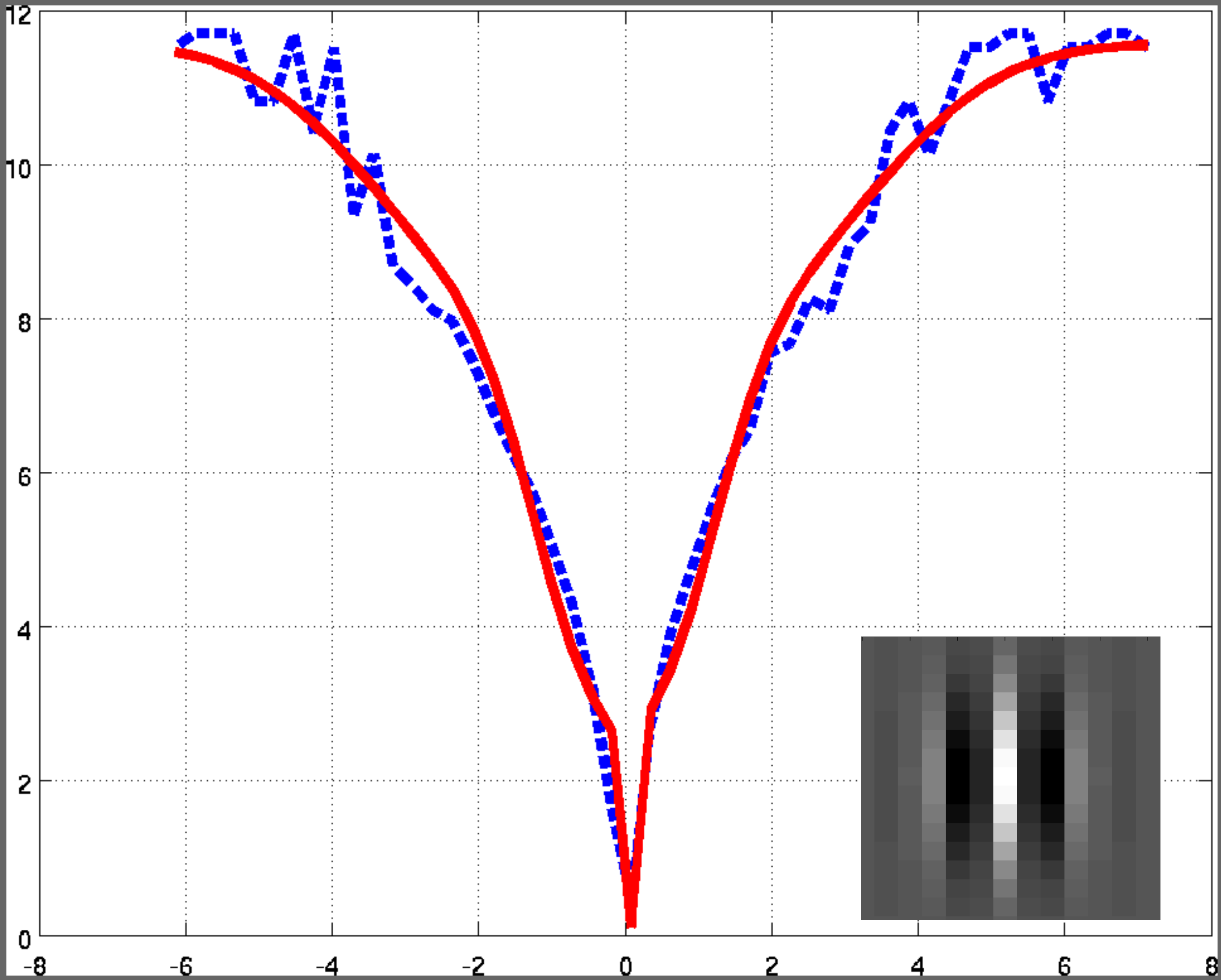
- log(empirical p.d.f. of filter response))



Finding an edge at this location and orientation is rare

- log(empirical p.d.f. of filter response))

- log(fit of model p.d.f. to filter response))

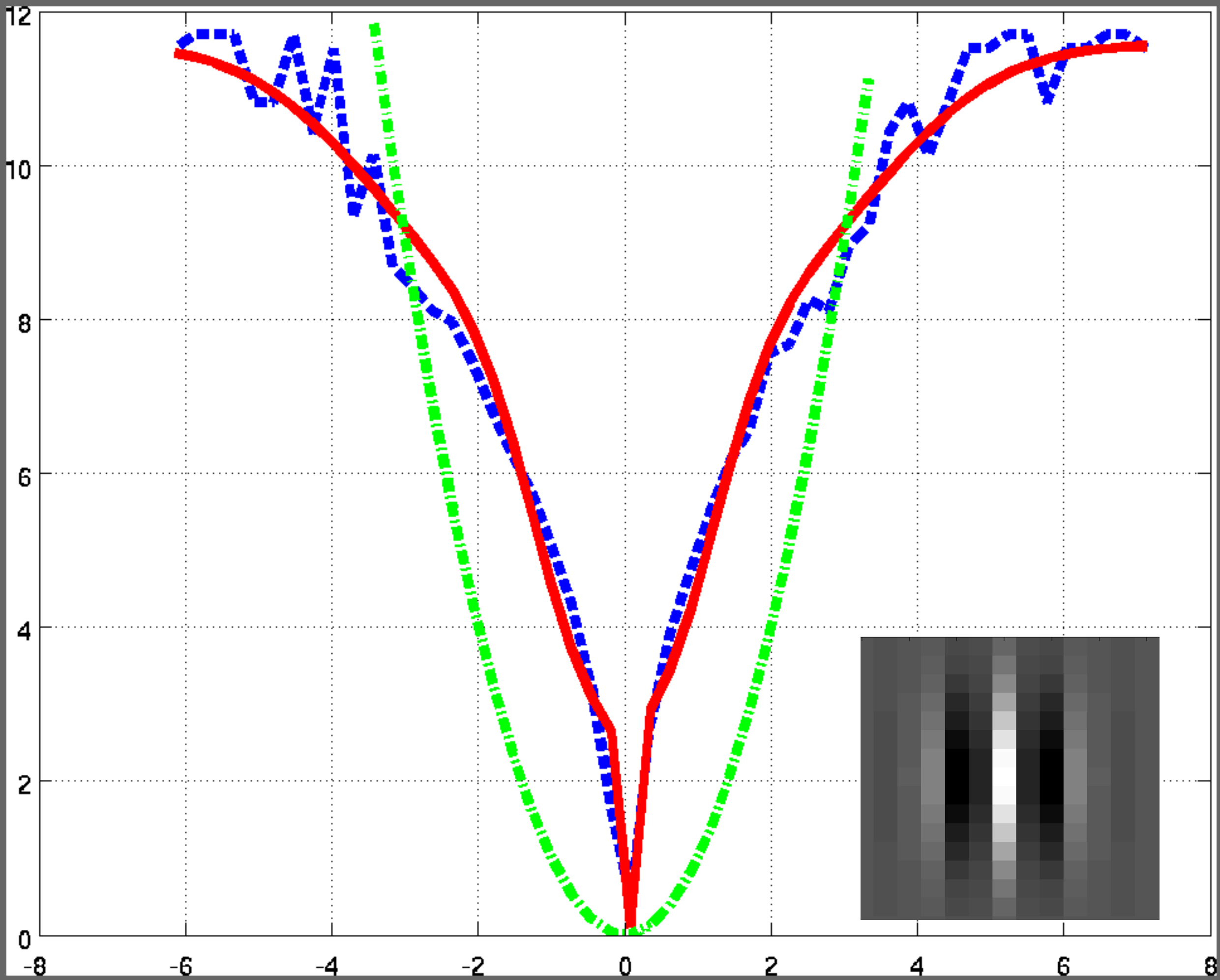


Finding an edge at this location and orientation is rare

- log(empirical p.d.f. of filter response))

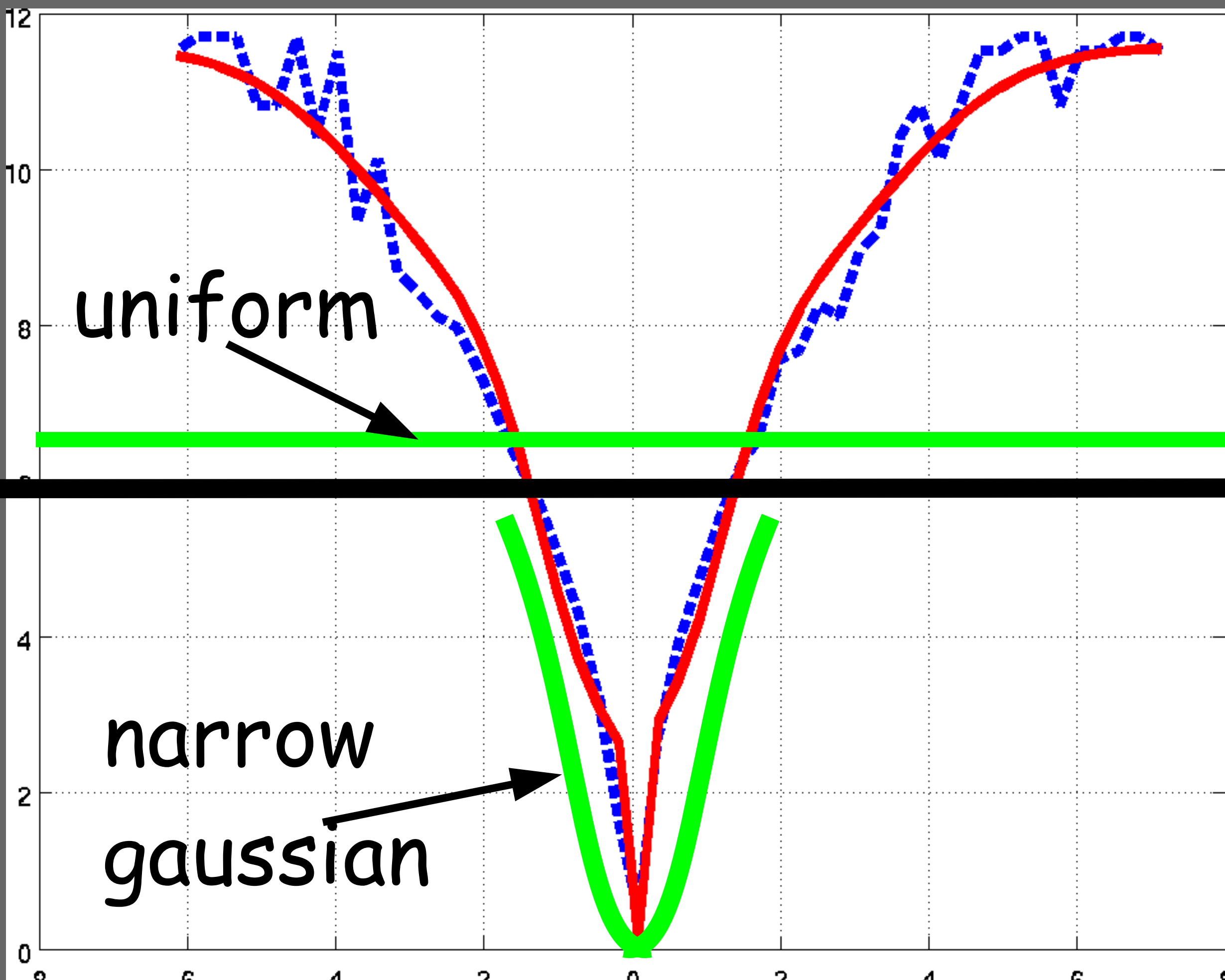
- log(fit of model p.d.f. to filter response))

- log(fit of Gaussian p.d.f. to filter response))



Finding an edge at this location and orientation is rare

KEY IDEA: use "switch" hidden variable



h = 0

structured images

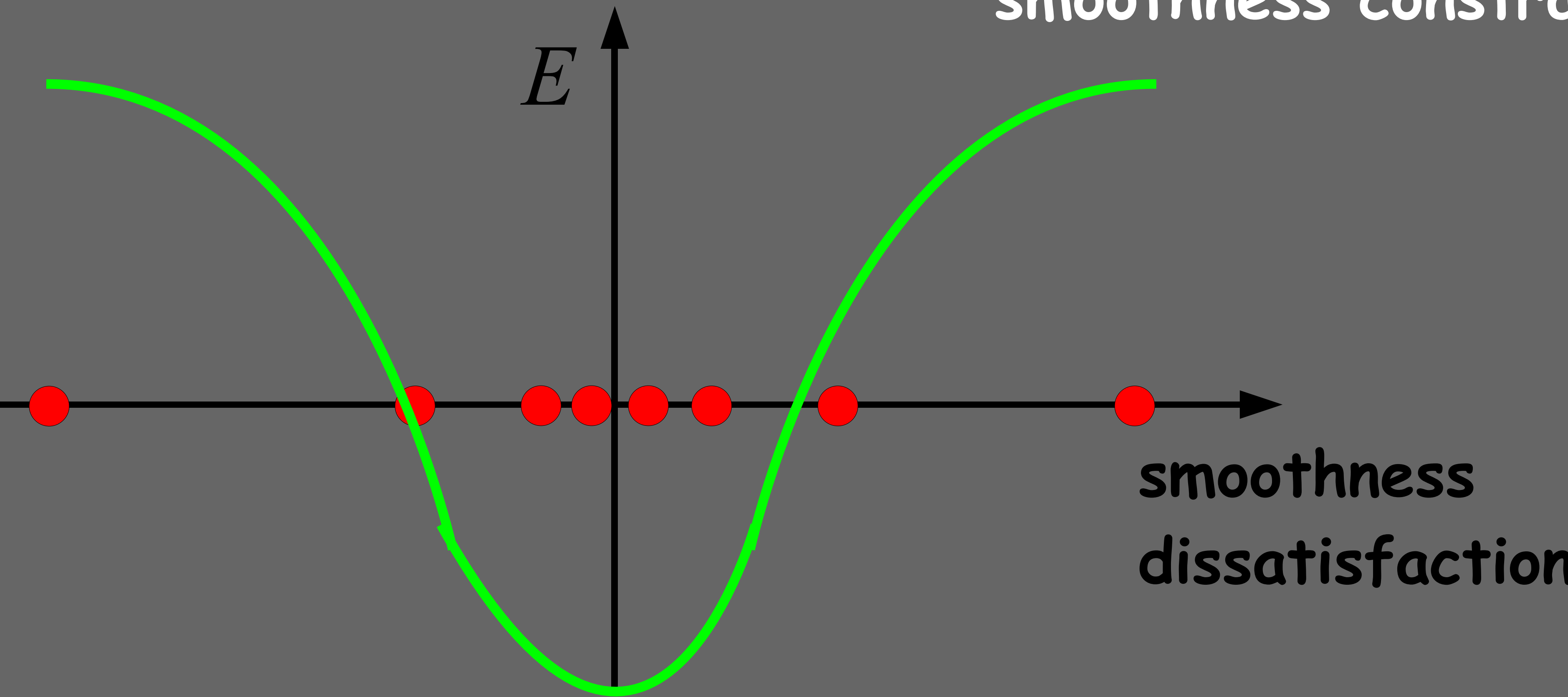
h = 1

smooth images

Goal: define $p(v,h)$

$$p(v,h) \propto \exp[-E(v,h)]$$

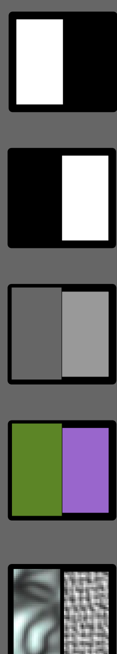
Energy measures dissatisfaction of smoothness constraints



- Smooth images are more likely
- (Rare) structured images are unlikely but possible

- What is a good measure of smoothness dissatisfaction?
- Edges are not smooth.
  E.g. what is a vertical edge?

- What is a good measure of smoothness dissatisfaction?
- Edges are not smooth.
  E.g. **what is a vertical edge?**



polarity

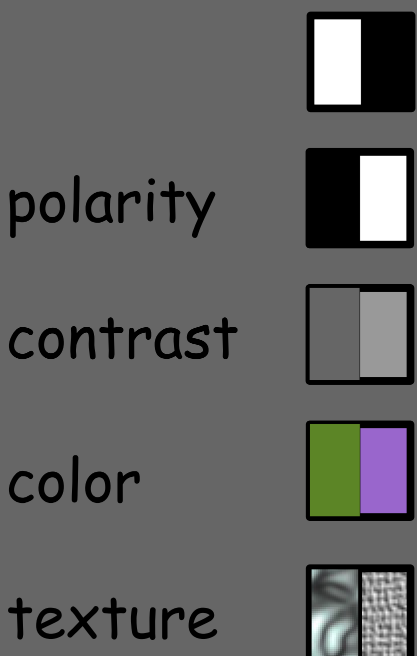
contrast

color

texture

...

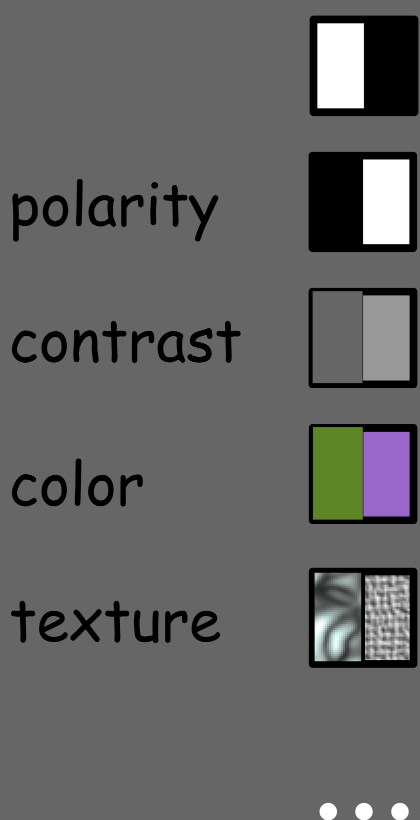
- What is a good measure of smoothness dissatisfaction?
- Edges are not smooth.
  E.g. what is a vertical edge?



...

Geoff: lack of horizontal "interpolation"

- What is a good measure of smoothness dissatisfaction?
- Edges are not smooth.
  E.g. what is a vertical edge?



Geoff: lack of horizontal "interpolation"

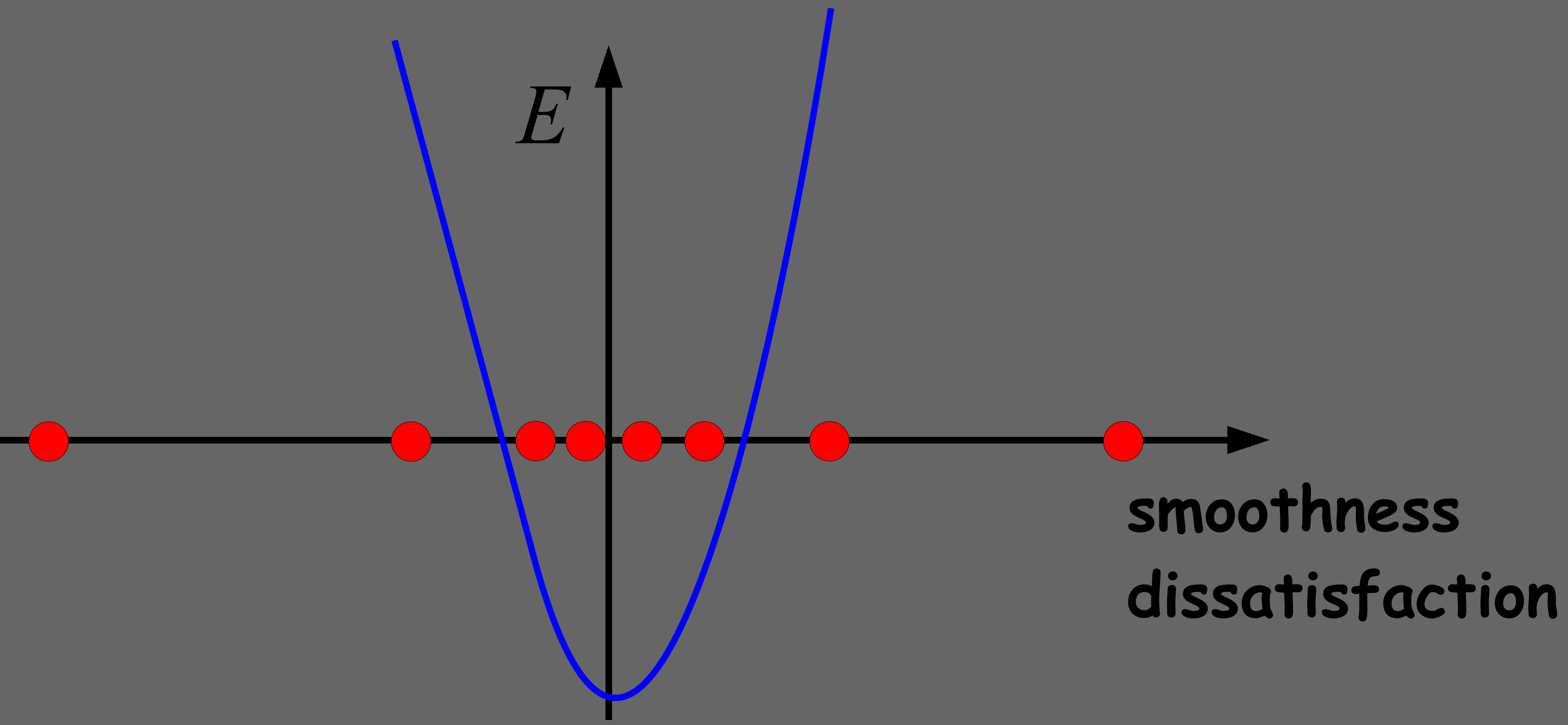
smoothness dissatisfaction
=
learned by set of linear filters
(patterns of "broken" interpolation)

$$E = \sum_i \text{smoothness\_dissatisfaction}_i^{۲} - b_i$$

$$\text{smoothness\_dissatisfaction}_i = W_i v$$

$$E = \sum_i \text{smoothness\_dissatisfaction}_i^2 - b_i$$

$$\text{smoothness\_dissatisfaction}_i = W_i v$$



$$E = \sum_i h_i \text{smoothness\_dissatisfaction}_i^2 - h_i b_i$$

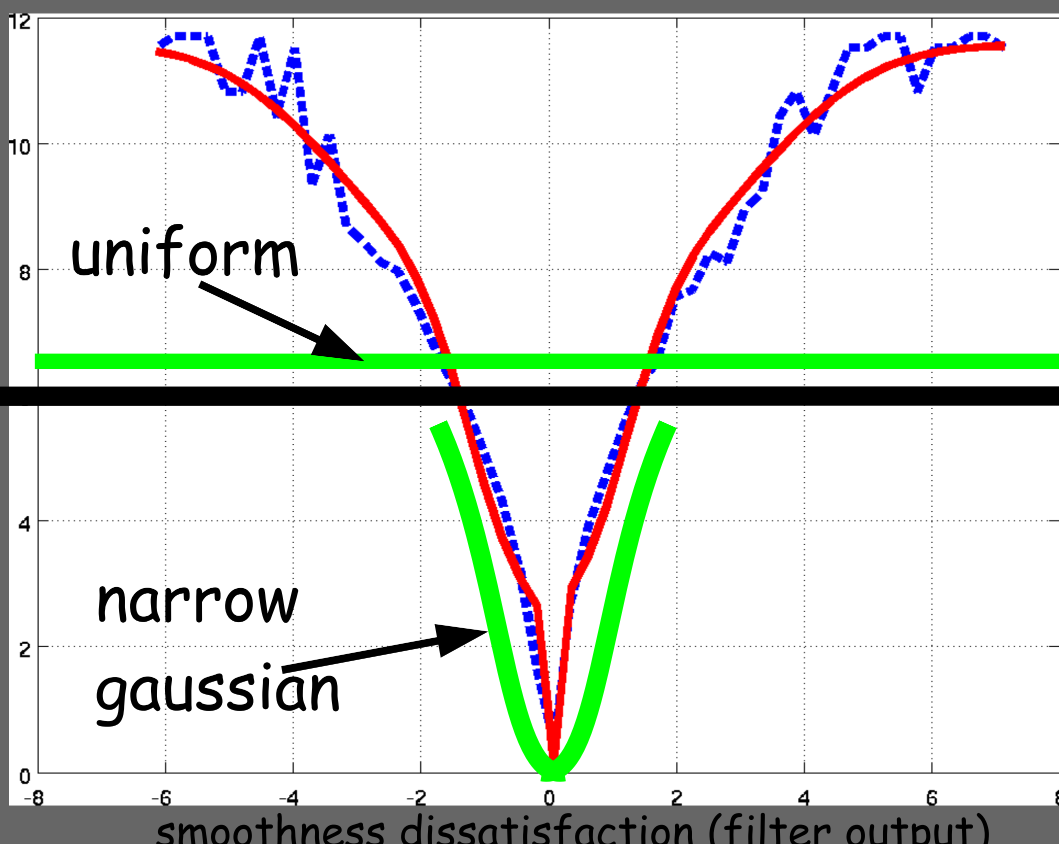
$$\text{smoothness\_dissatisfaction}_i = W_i v$$

$h_i \in \{0,1\}$ introduce hiddens to allow violations

$$E = \sum_i h_i \,\textbf{smoothness\_dissatisfaction}_i^2 - h_i b_i$$

$$\textbf{smoothness\_dissatisfaction}_i = W_i v$$

$h_i \in \{0,1\}$ **introduce hiddens to allow violations**



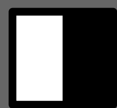
h = 0

structured images

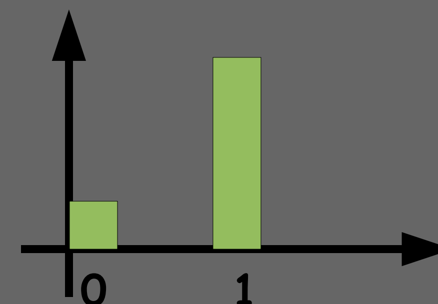
h = 1

smooth images

$$E = h_1 (W_1 v)^2 + h_2 (W_2 v)^2 - b_1 h_1 - b_2 h_2$$

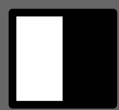
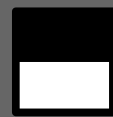
$W_1 =$  $W_2 =$  $b_1 = 10$ $b_2 = 10$

$$p(h_i = 1 | v) = \sigma(-(W_i v)^2 + b_i)$$

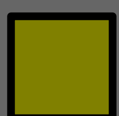


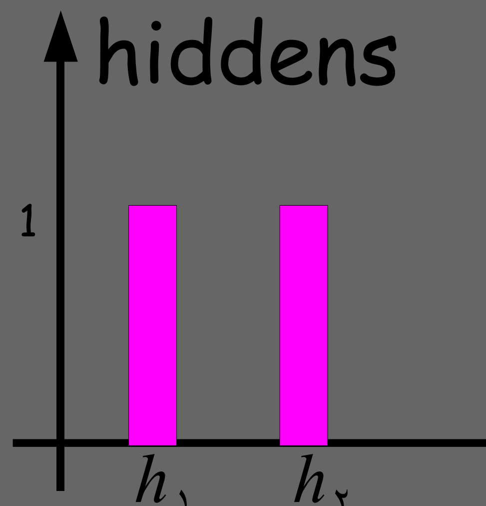
$$p(v|h) = N\left(0, \frac{1}{2} (h_1 W_1' W_1 + h_2 W_2' W_2)^{-1}\right)$$

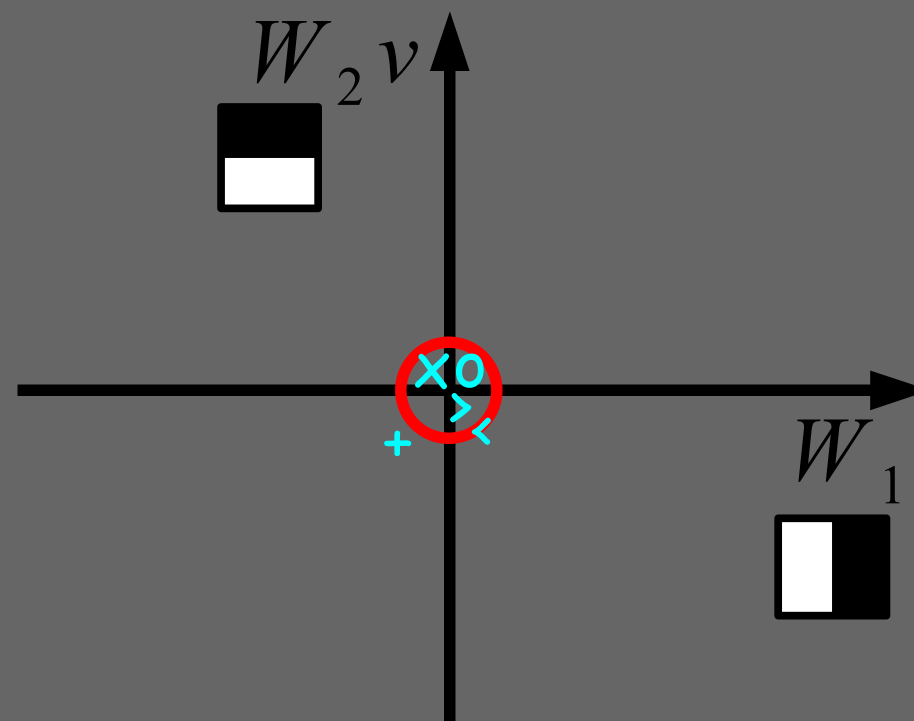
$$E = h_1 (W_1 v)^2 + h_2 (W_2 v)^2 - b_1 h_1 - b_2 h_2$$

$W_1 =$  $W_2 =$  $b_1 = 10$ $b_2 = 10$

image

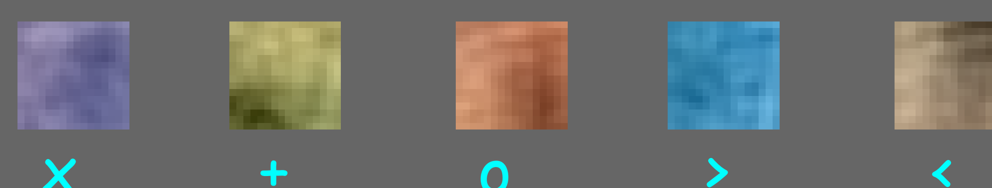




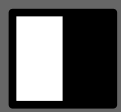
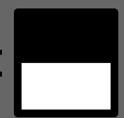


$$E(v, h) \approx -b_1 - b_2 = -20$$

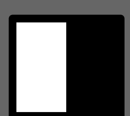
low energy

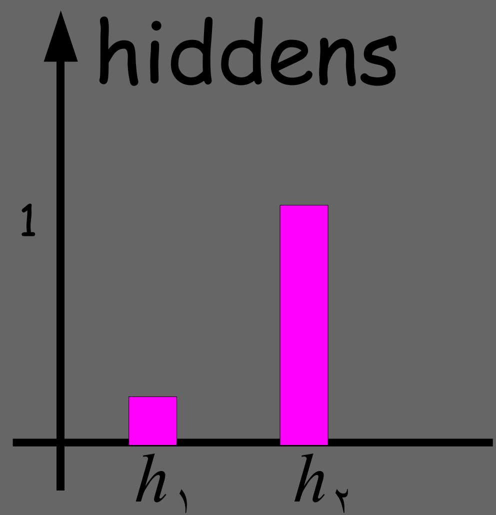


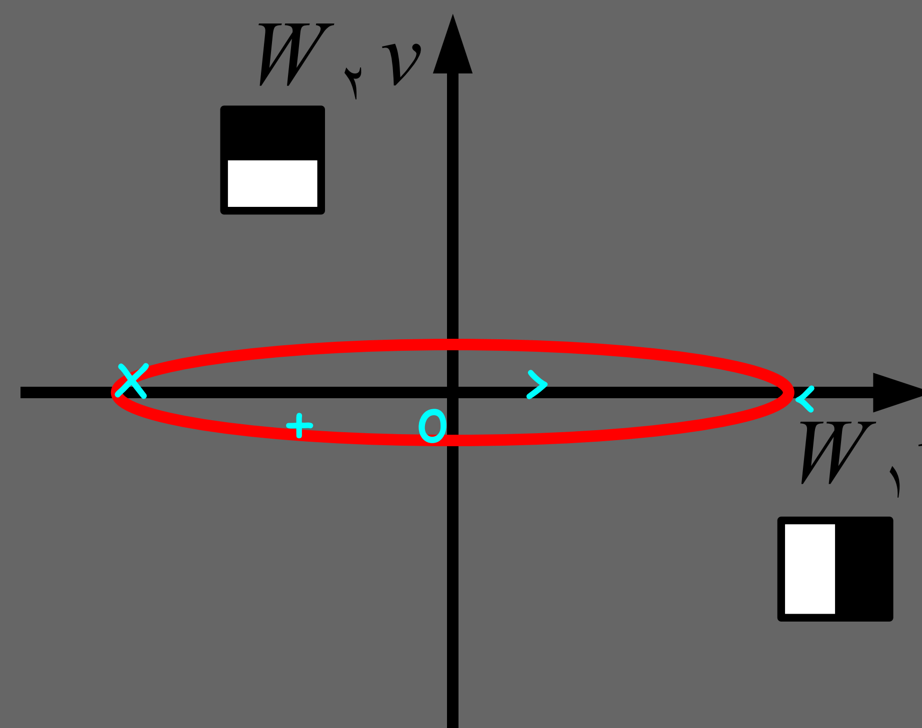
$$E = h_1 (W_1 v)^2 + h_2 (W_2 v)^2 - b_1 h_1 - b_2 h_2$$

$W_1 =$  $W_2 =$  $b_1 = 10$ $b_2 = 10$

image



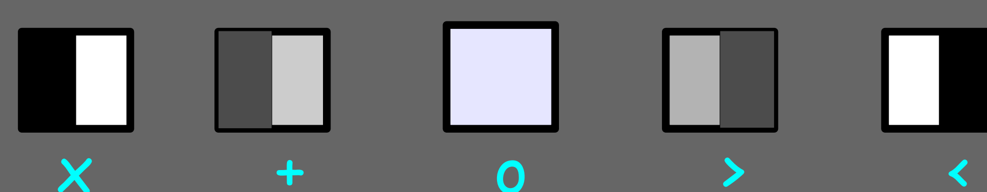




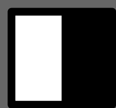
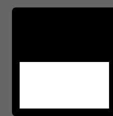
$$E(v, h) \approx -b_2 = -10$$

higher energy

(h_1 gave discount!)

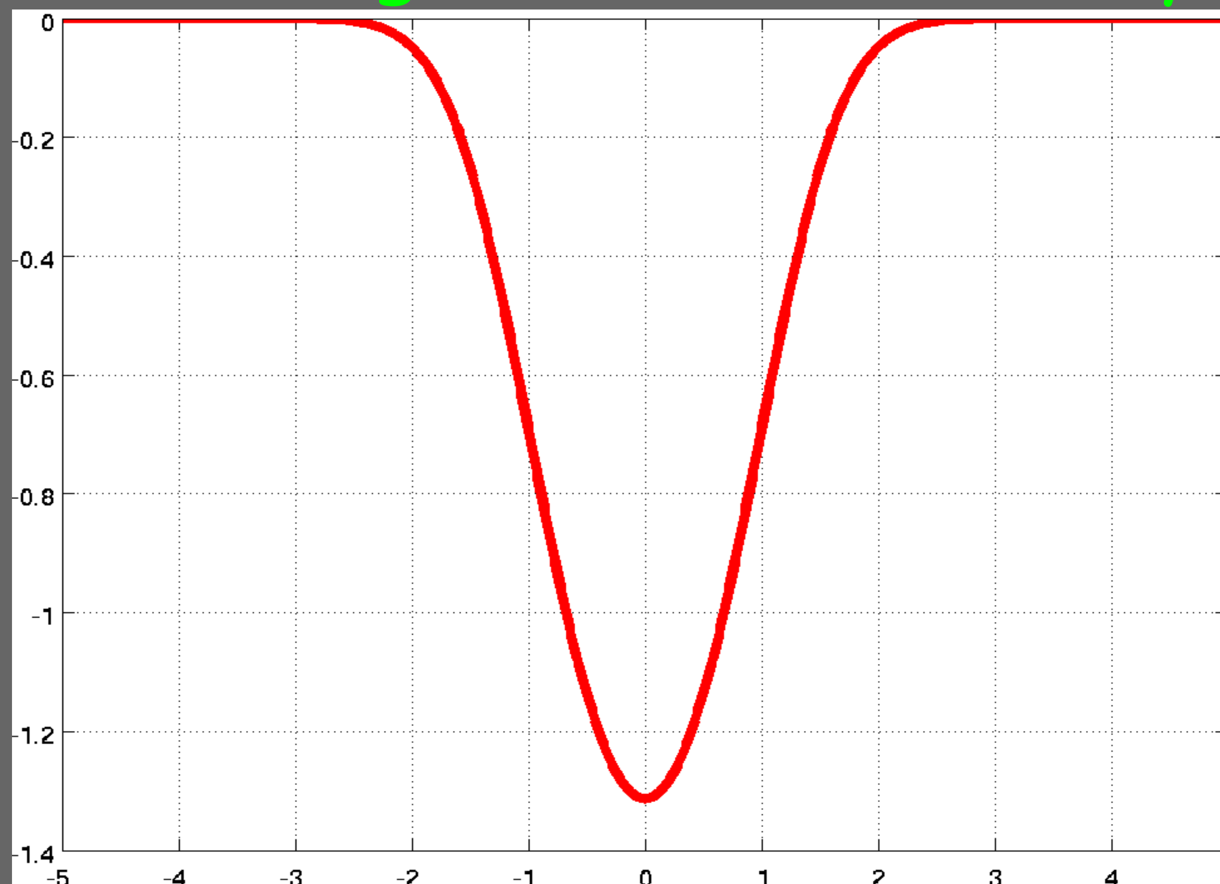


$$E = h_1 (W_1 v)^2 + h_2 (W_2 v)^2 - b_1 h_1 - b_2 h_2$$

$W_1 =$  $W_2 =$  $b_1 = 10$ $b_2 = 10$

Binary hiddens can be marginalized out exactly

- log(p(v)) + k



filter output

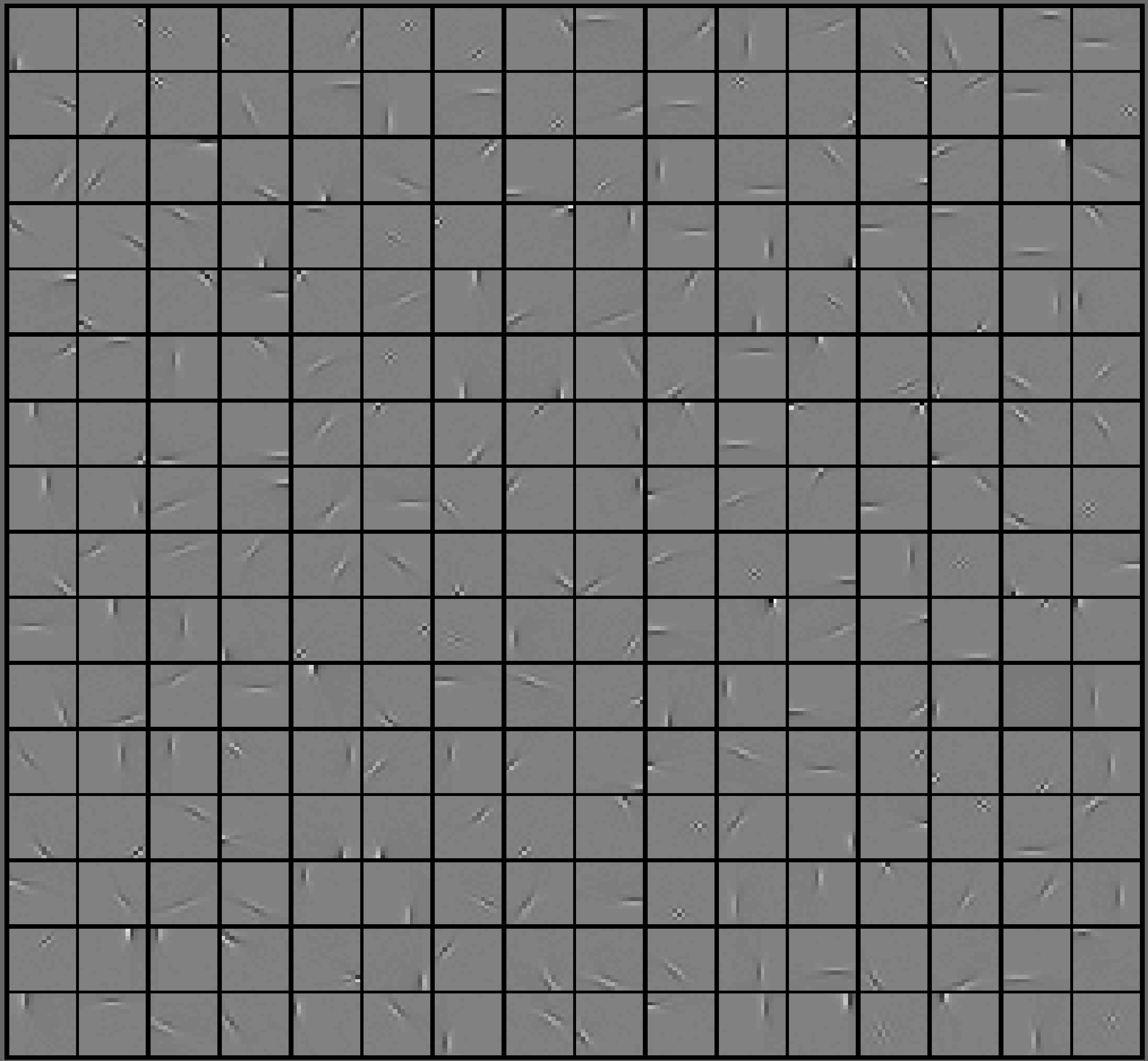
We allow (rare) smoothness violations

$$E = \frac{1}{2} \sum_k \sum_f h_k P_{fk} (W_f v)^2 - b_k h_k$$

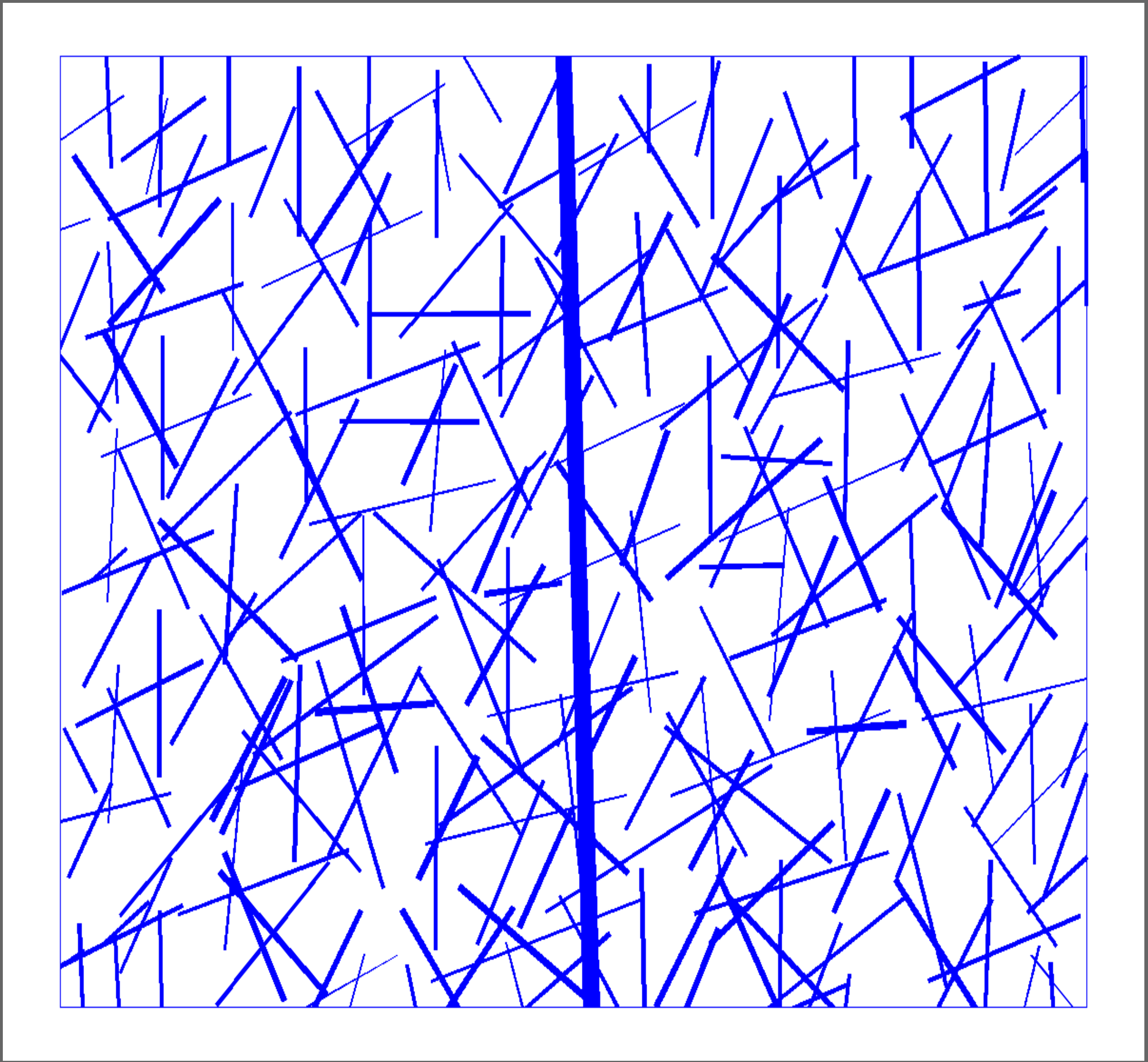
Learning: $\{W, P, b\}$

- approximate maximum likelihood
  - Contrastive Divergence
  - Hybrid Monte Carlo

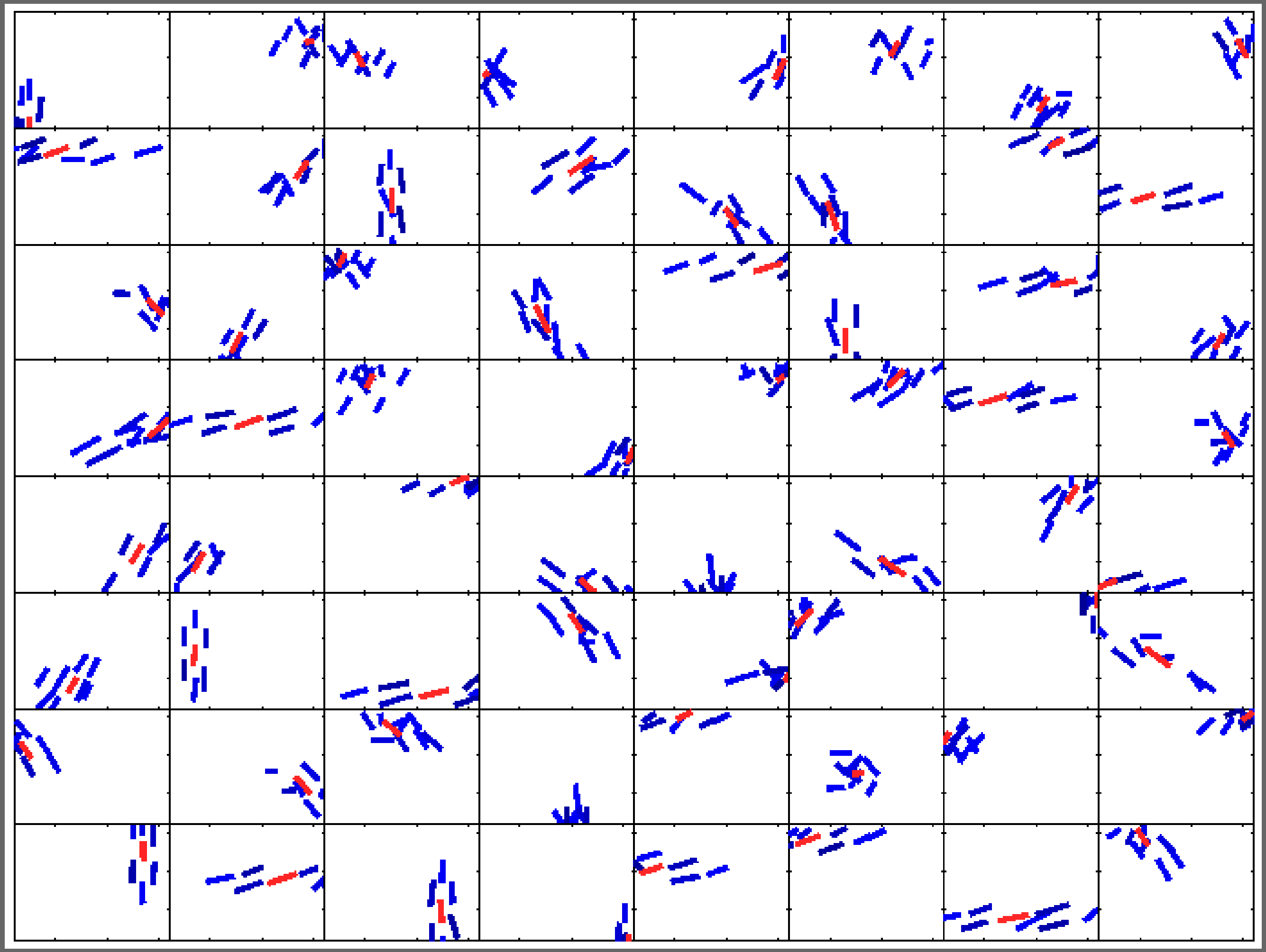
Learned filters (directions of smoothness dissatisfaction) on natural images



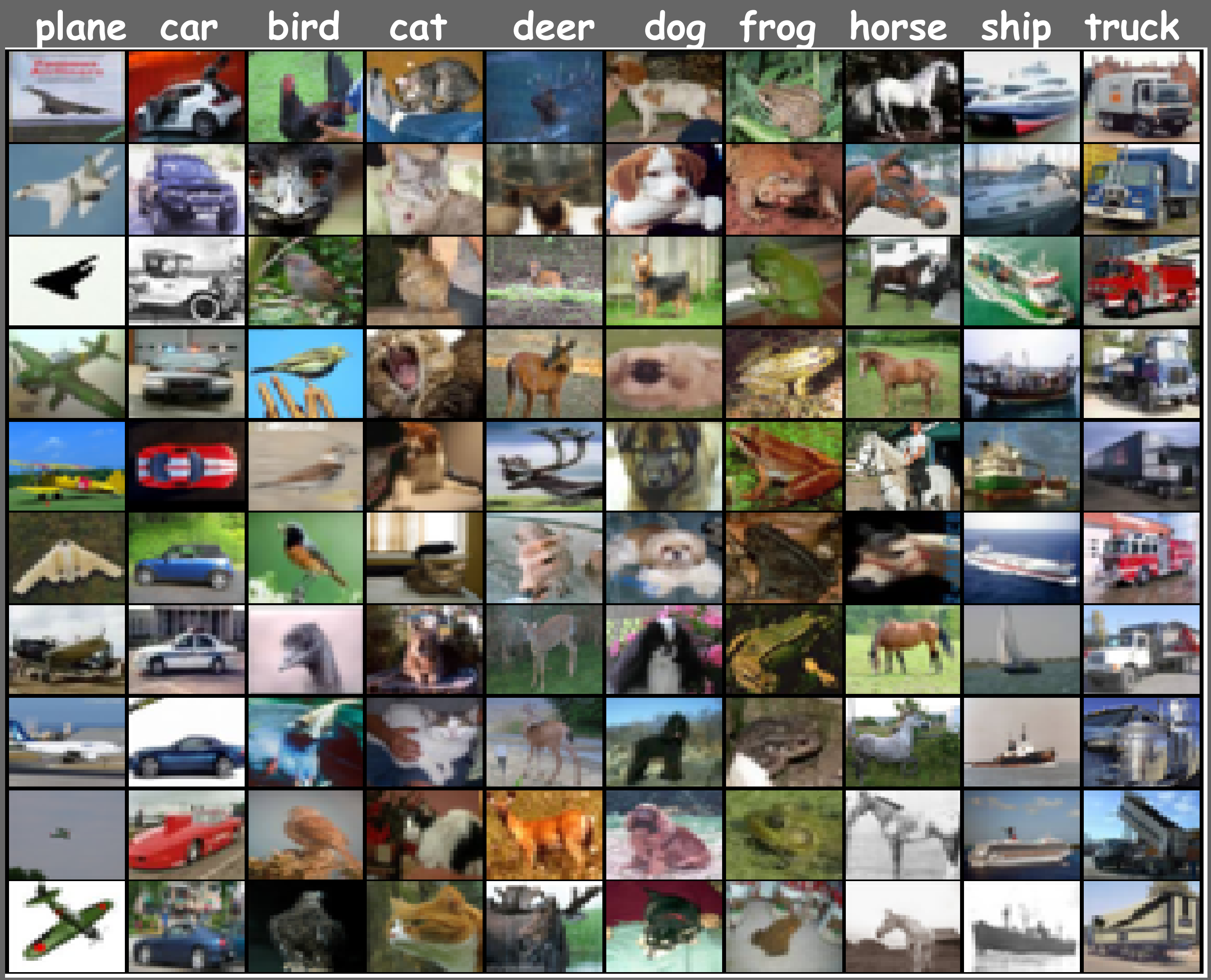
Learned filters: Gabor fit



Learned filters: grouping



RECOGNITION: CIFAR 10 dataset



RECOGNITION: CIFAR 10 dataset

	Accuracy
3way RBM - DBN input -> 9800 - 4096	64.7%
GRBM - DBN* input - 10000 - 10000	56.6%
3way RBM - DBN input -> 9800 - 4096 - 384	58.7%
GIST input - 384	54.7%

* from Krizhevsky 2009

Conclusions

- Model of natural images producing binary features
  - Invariance or robustness to distortions
  - Good for recognition
  - Probabilistic model of simple-complex cell model
  - Easy to integrate with DBN's

Conclusions

- Model of natural images producing binary features
  - Invariance or robustness to distortions
  - Good for recognition
  - Probabilistic model of simple-complex cell model
  - Easy to integrate with DBN's
- Code and more recent developments available at:

  http://www.cs.toronto.edu/~ranzato

THANK YOU!