Learning Features by Contrasting Natural Images with Noise

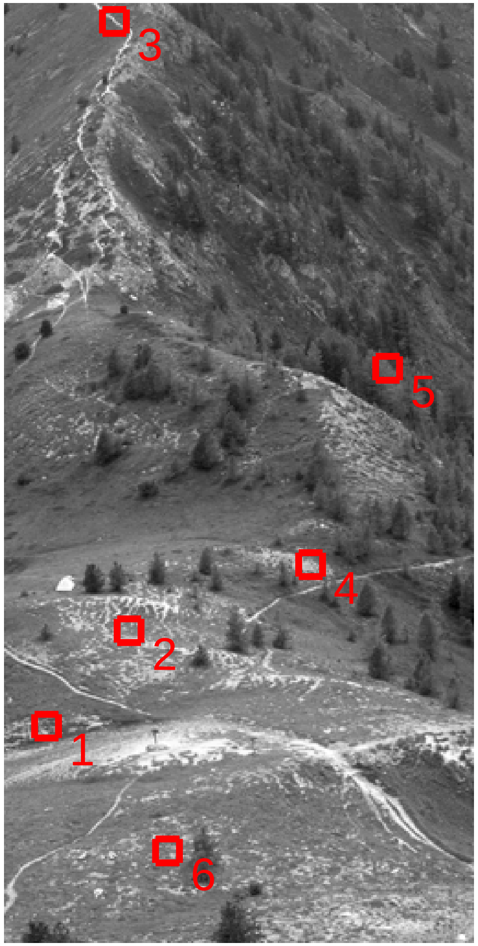
Michael Gutmann
University of Helsinki
michael.gutmann@helsinki.fi

Aapo Hyvärinen
University of Helsinki
aapo.hyvarinen@helsinki.fi

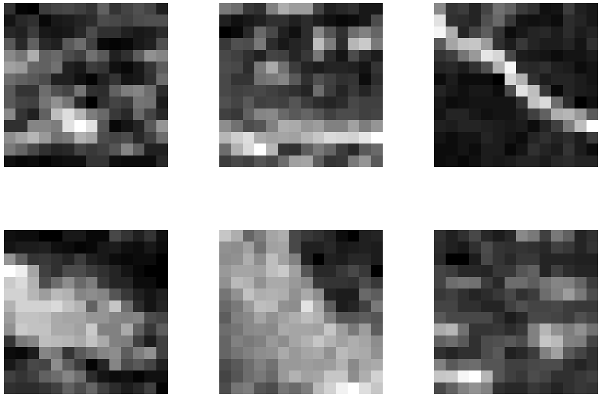
Introduction

Natural images?

A natural scene



Natural image patches



- Most prob. models are models of natural image *patches*.
- Difficult enough as the data is high dimensional.
- Can serve as building block for models of entire scenes.

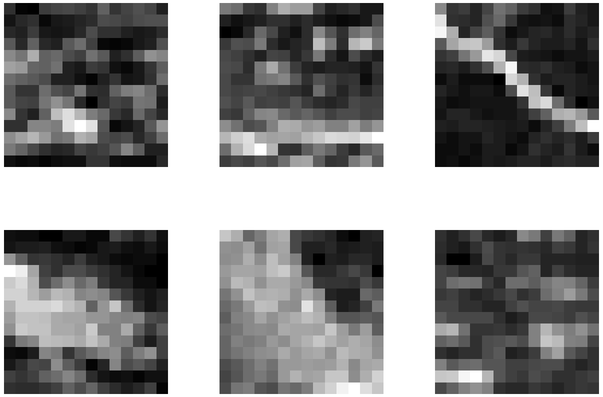
Why model natural images?

Possible motivations for building statistical models of natural images:

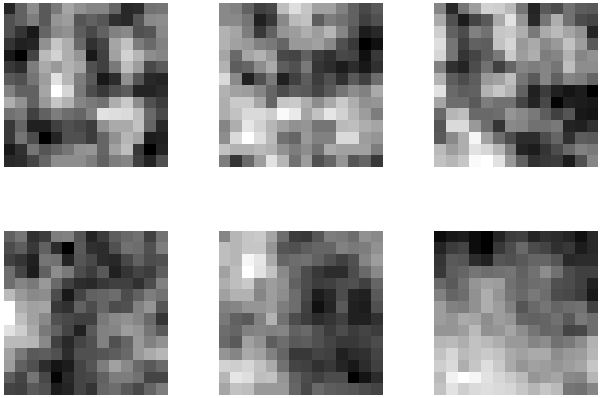
- You can use the model as prior in tasks which involve Bayesian inference (not the topic of this presentation).
- You can use the model to generate artificial natural images (not the topic of this presentation).
- There are some connections to visual neuroscience (not the topic of this presentation).
- *Finding interesting features of natural images:*
  - What kind of features appear in natural images?
  - What kind of structure is characteristic for natural images?
  - How do natural images differ from other, artificial, image data (noise)?

Natural images vs. noise

In what aspects do the two datasets differ from each other?

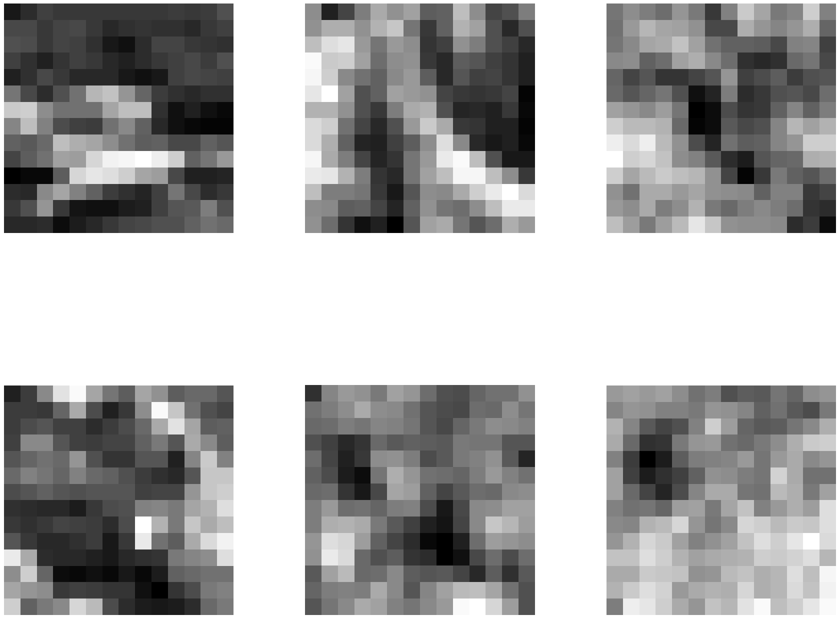


natural images



noise images

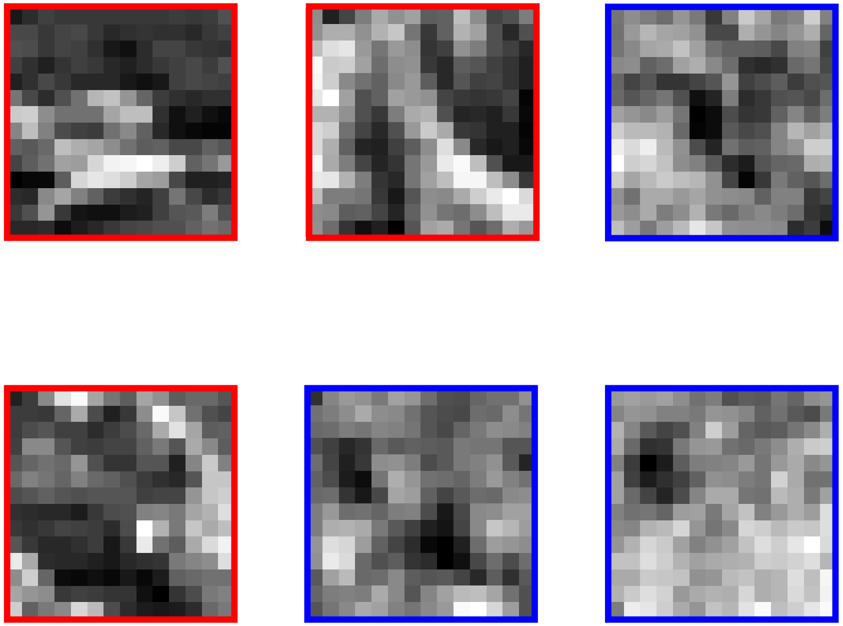
Learning features by classification



Key idea:

- Train a classifier to discriminate between natural images and some artificial noise.
- To succeed in the discrimination task, the classifier must “discover structure” in the data, i.e. *identify features of natural images.*

Learning features by classification



Key idea:

- Train a classifier to discriminate between natural images and some artificial noise.
- To succeed in the discrimination task, the classifier must "discover structure" in the data, i.e. *identify features of natural images.*

Contrastive feature learning

The elements of contrastive feature learning

1. Classifier: Assign $C = 1$ if input $\mathbf{x}$ is a natural image, and $C = 0$ if input is noise.
2. Estimation method: Fit the parameters in the classifier to the data (supervised learning!)
3. Noise: The reference data that is used for comparison with the natural images.

More on the classifier and its estimation

- Use a classification approach based on logistic regression

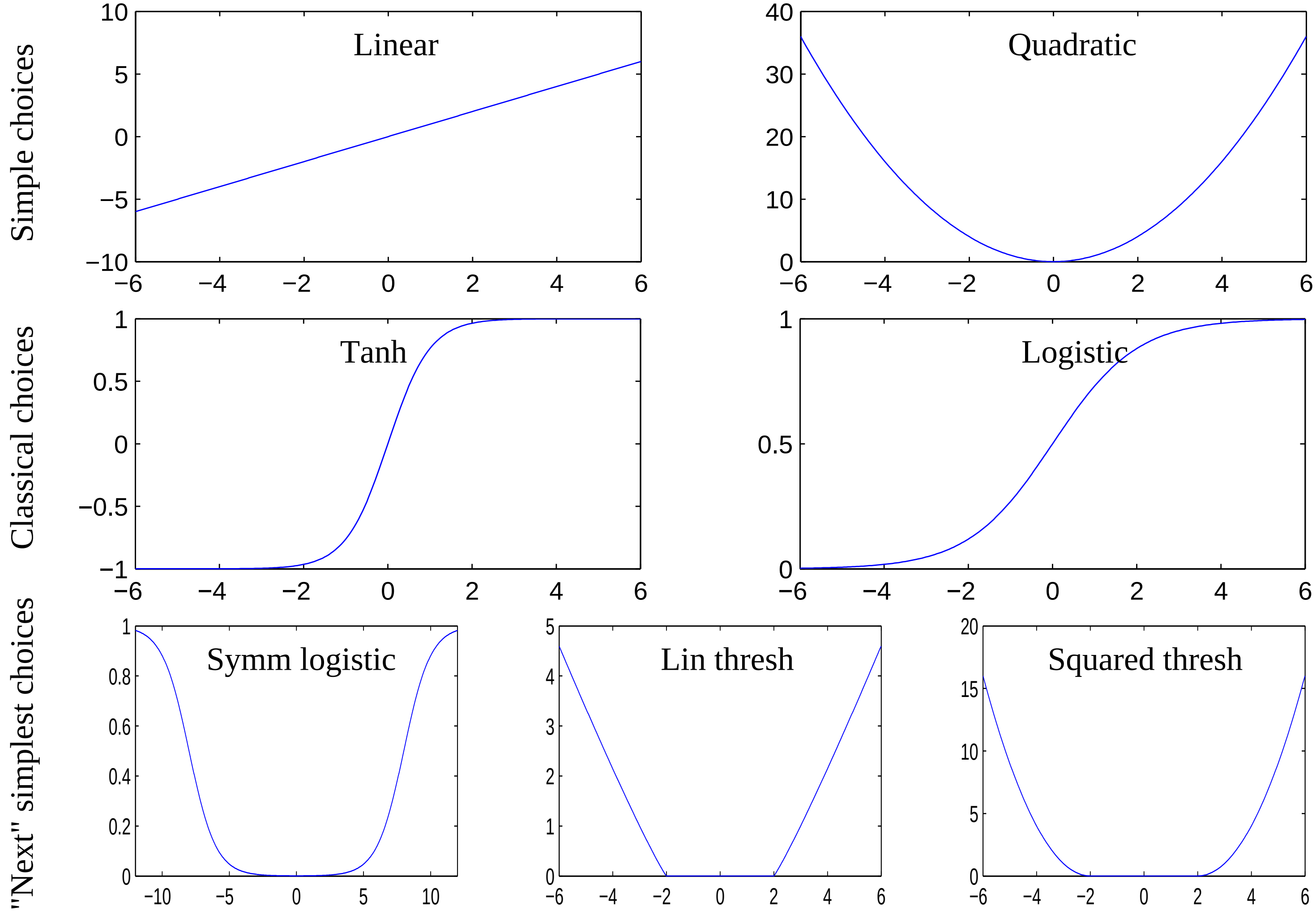
$$P(C=1|\mathbf{x}) = \frac{1}{1+\exp(-y(\mathbf{x}))} \qquad y(\mathbf{x}) = \sum_{m=1}^{M} g(\mathbf{w}_m^T\mathbf{x} + b_m) + \gamma$$

- Parameters in the model are the features $\mathbf{w}_m$, the bias terms b_m, the offset γ, as well as possibly the function $g(u)$.
- The parameters can be estimated by maximum (conditional) likelihood. This is the same as minimization of the cross-entropy error J

$$J = \frac{1}{T}\sum_{t=1}^{T} -C_t \log\left[P(C_t=1|\mathbf{x}_t)\right] - (1-C_t)\log\left[1-P(C_t=1|\mathbf{x}_t)\right]$$

- Reference data: Use noise with the same covariance structure as natural images.

Choice of the nonlinearity in the discriminant $y(\mathbf{x})$



Parameterized nonlinearity:

$$g(u) = \alpha_1[\max(0, u - \beta_1)]^{\eta_1} + \alpha_2[\max(0, -(u - \beta_2))]^{\eta_2}$$

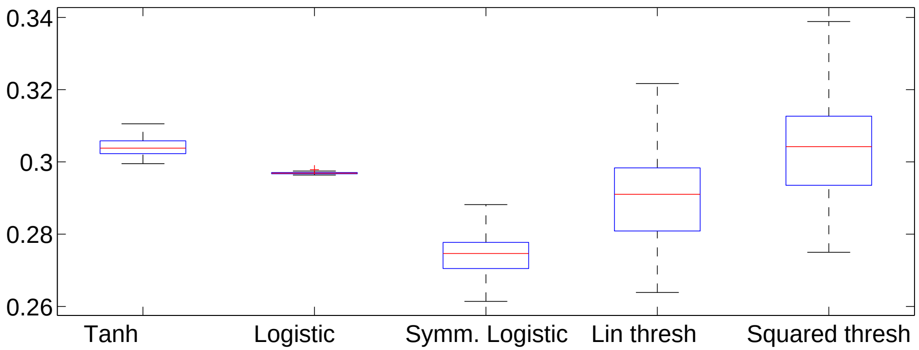
Simulations

The questions addressed

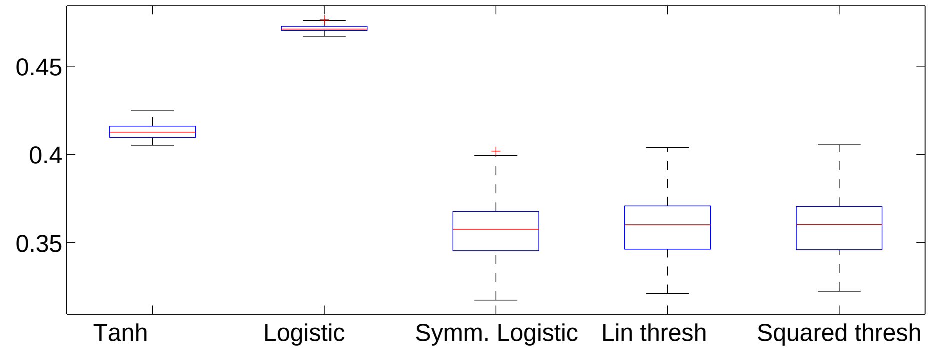
1. Which nonlinearity $g(u)$ gives the best performance?
2. How do the features $\mathbf{w}_m$ look like?
3. Which principle does the classifier use to solve the discrimination task?

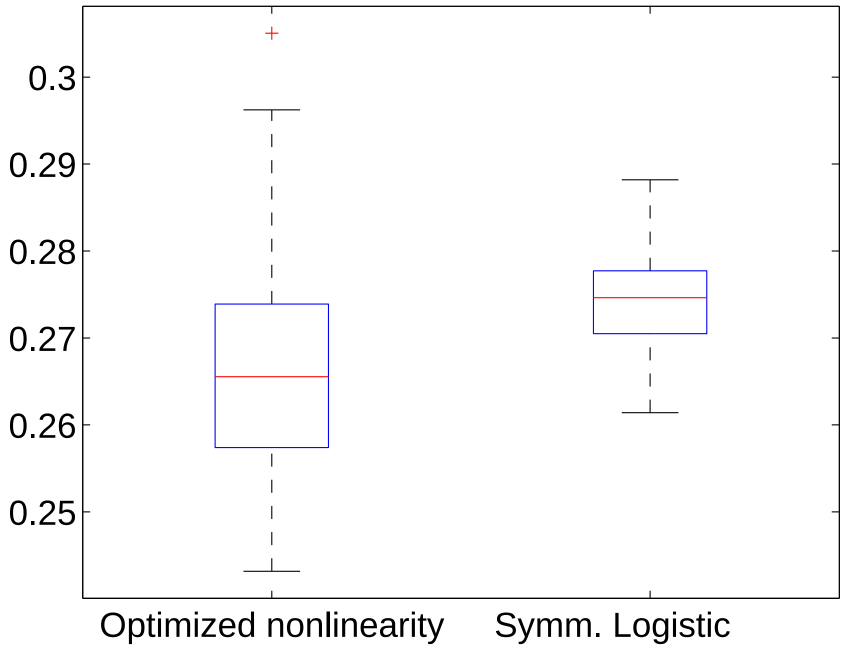
Classification performance

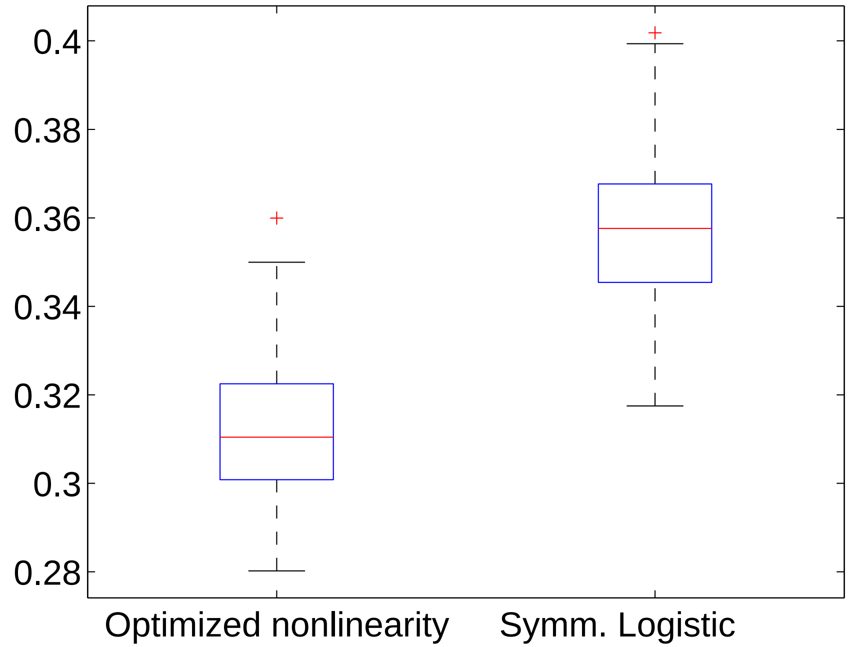
Cross entropy error



False classification rate

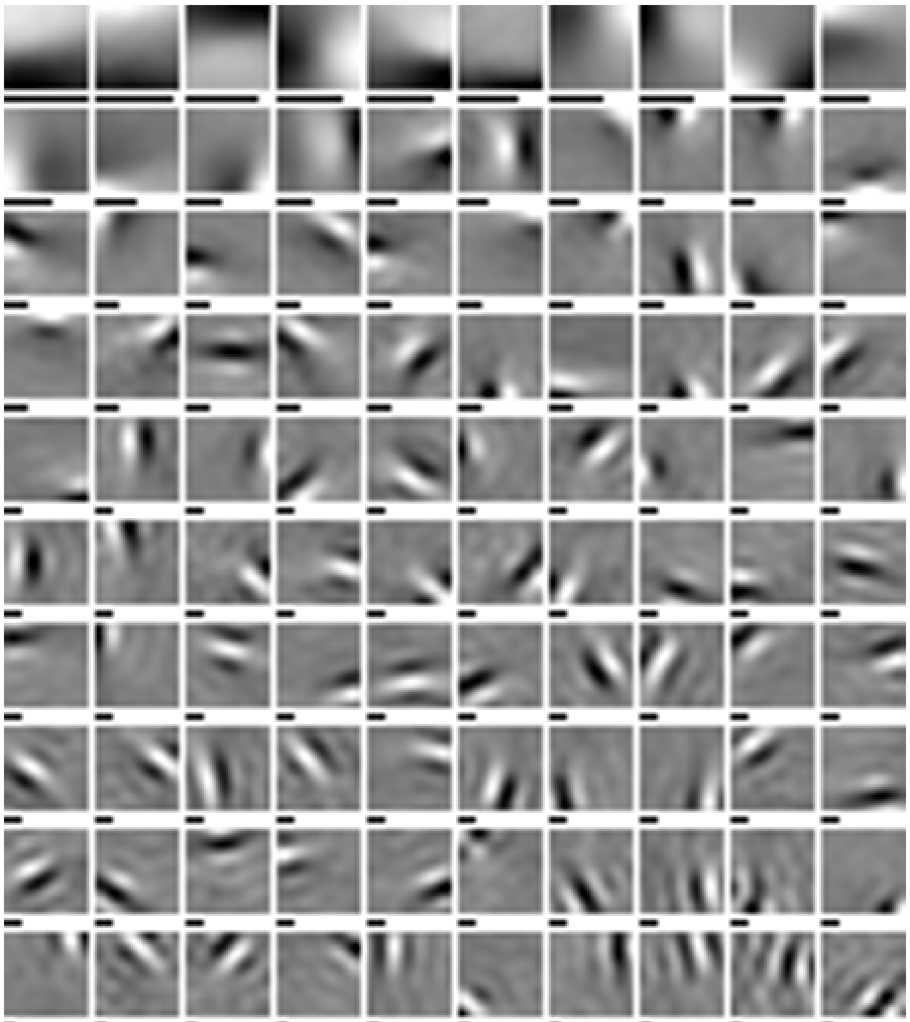




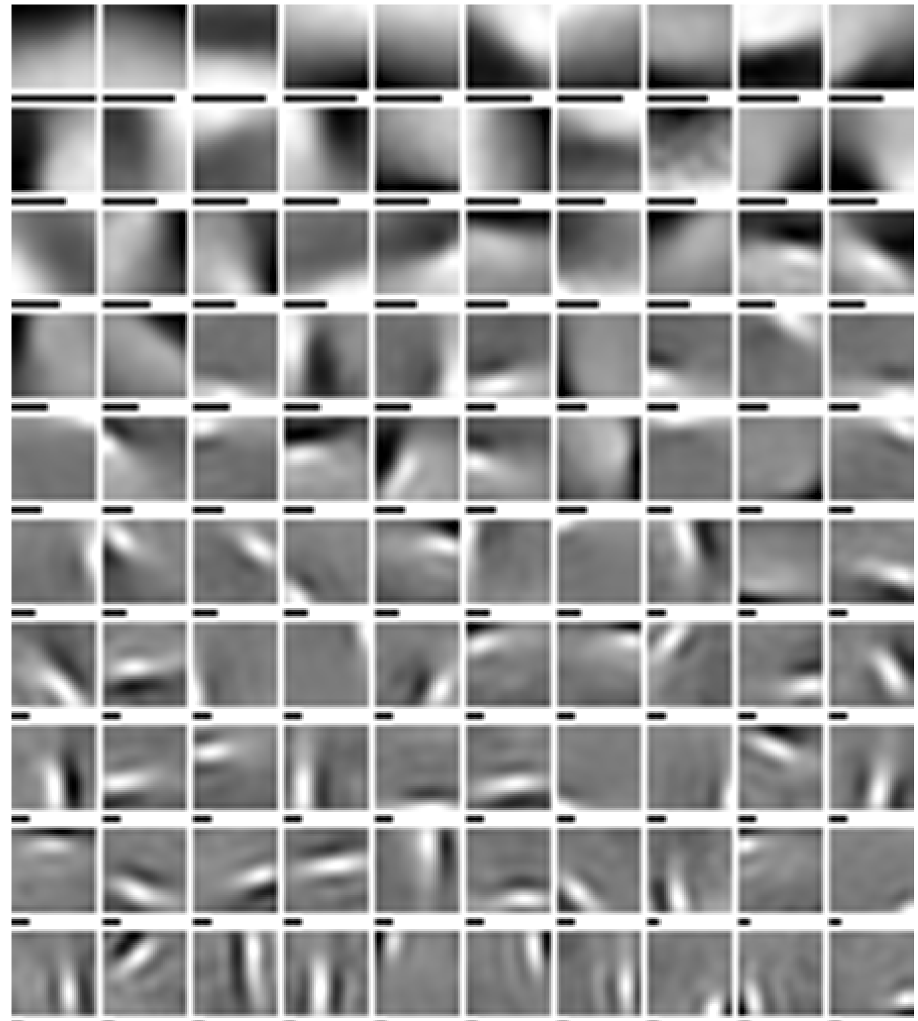


The learned features

Symm. logistic



Optimized nonlinearity

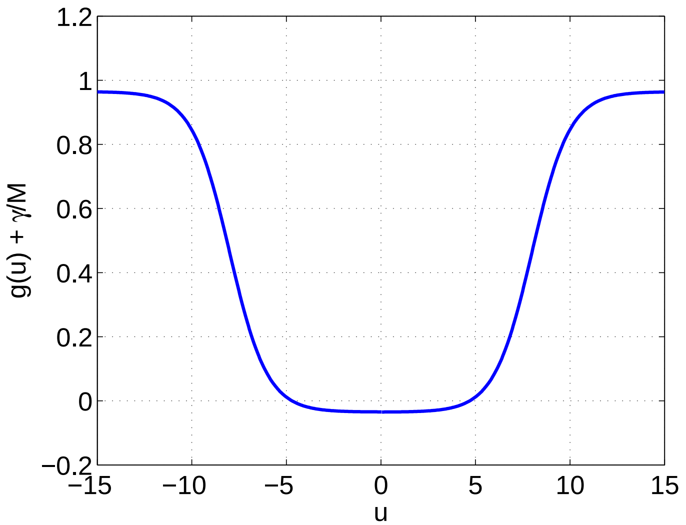


Classification principle

The discriminant $y(\mathbf{x}) = \sum_{m=1}^{M} \left(g(\mathbf{w}_m^T \mathbf{x}) + \gamma/M \right)$ rules:

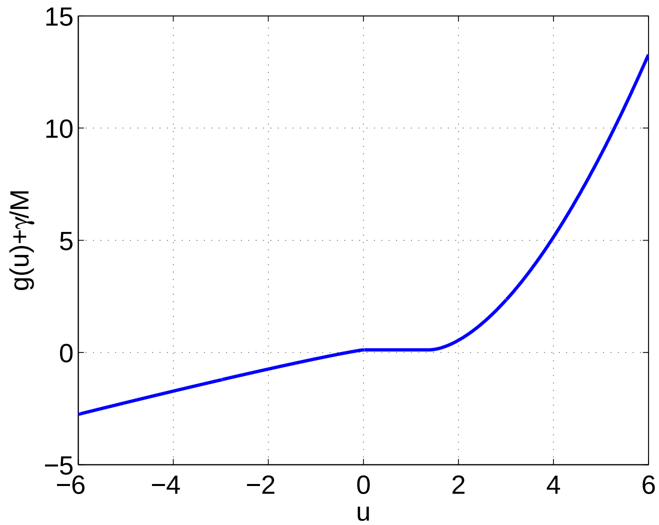
$y(\mathbf{x}) > 0 \Rightarrow \mathbf{x}$ is a natural image. $y(\mathbf{x}) < 0 \Rightarrow \mathbf{x}$ is noise.

Symm. logistic



Thresholding of each feature output

Optimized nonlinearity



Sign consistency across feature outputs

Summary

- The minimum to retain:
  1. The talk was about learning features in data (here: natural images).
  2. Features are learned by training a classifier to distinguish between the data and some artificial noise.
  3. We used nonlinear logistic regression to do the classification.
- Some more details:
  1. Classification by thresholding outputs of gabor-like feature detectors.
  2. Optimizing the nonlinearity gives an asymmetric solution. Classification performance improves.
  3. An alternative classification principle: the outputs of some gabor-like feature detectors need to have the same sign for natural images.