Feed-forward Neural Networks (Part 1)

Outline (part 1)

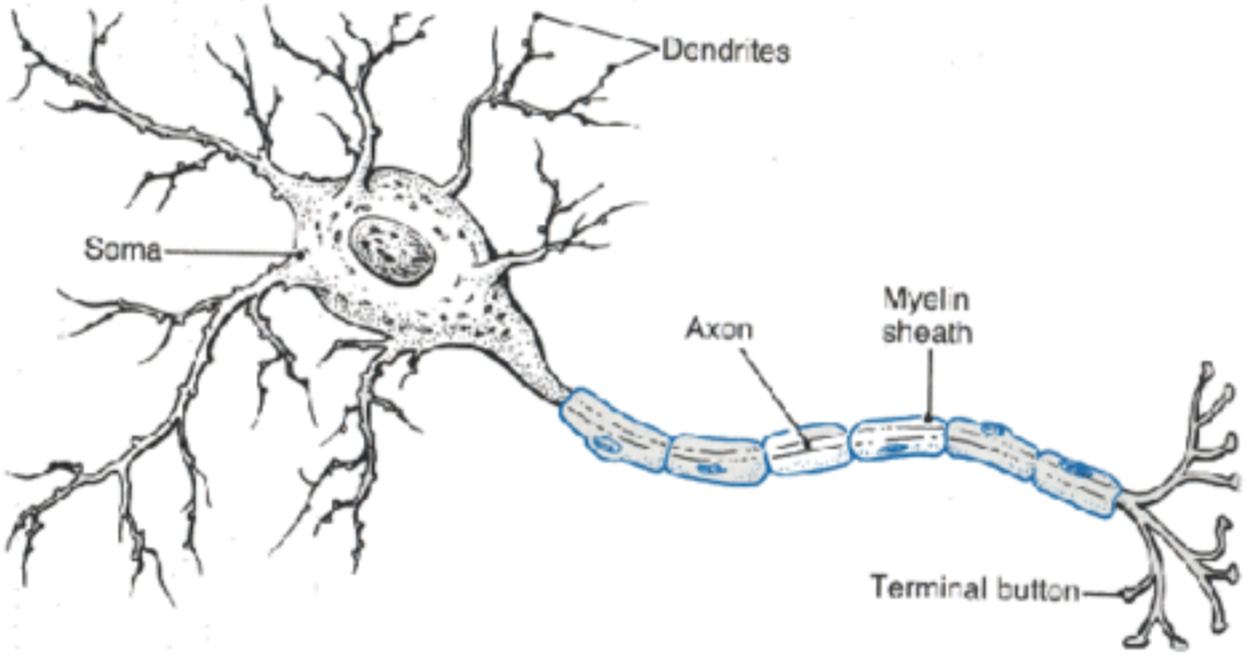
- Feed-forward neural networks
- The power of hidden layers
- Learning feed-forward networks
  - SGD and back-propagation

Motivation

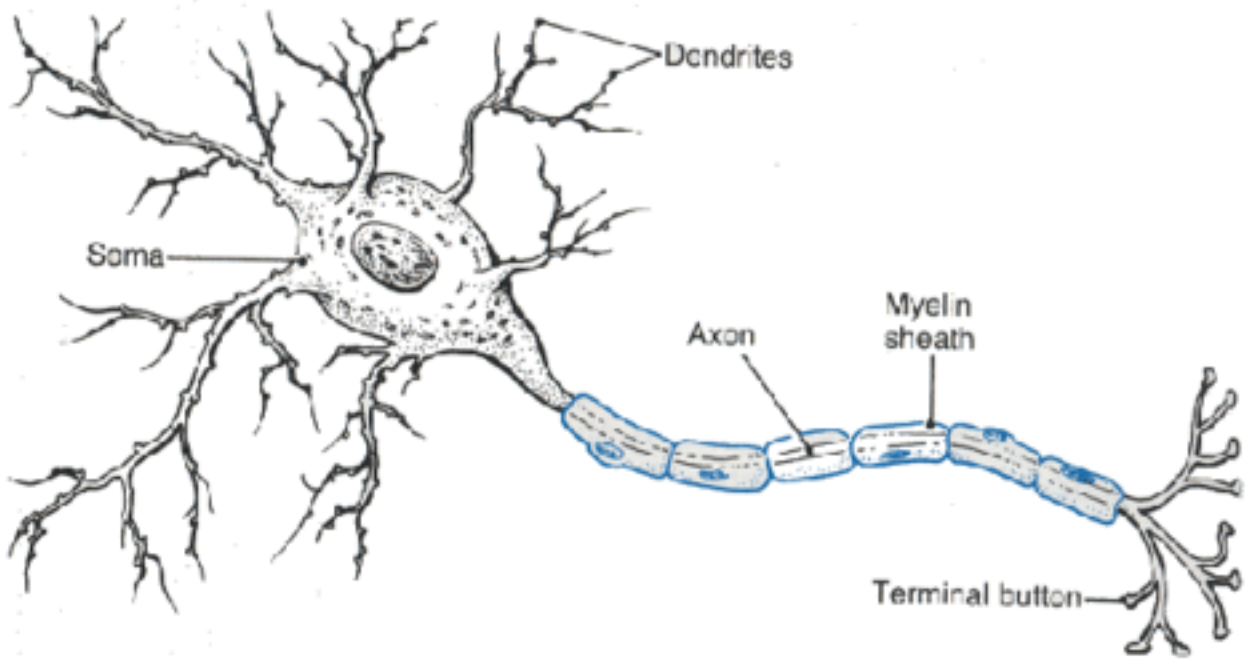
- So far our classifiers rely on pre-compiled features

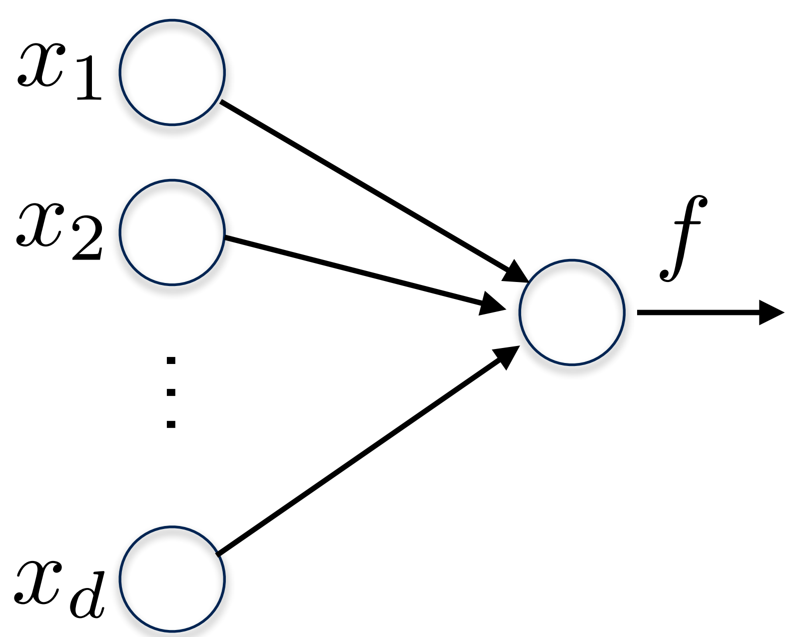
$$\hat{y} = \text{sign}\big(\, \theta \cdot \phi(x) \,\big)$$

Neural Networks



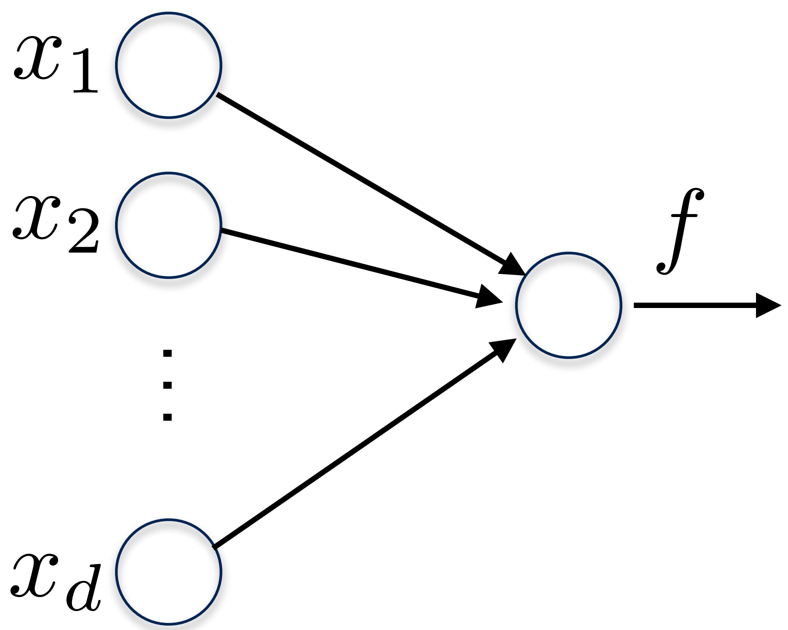
(Artificial) Neural Networks



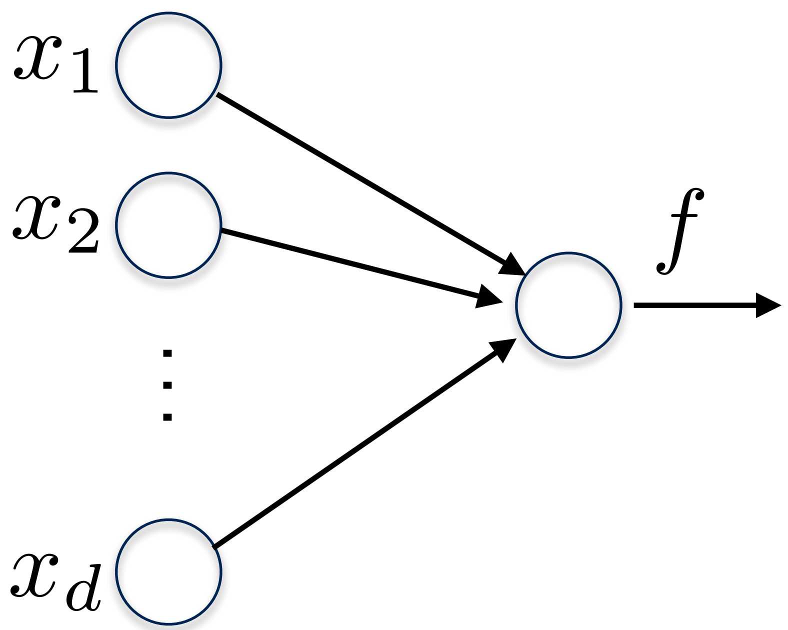


(e.g., a linear classifier)

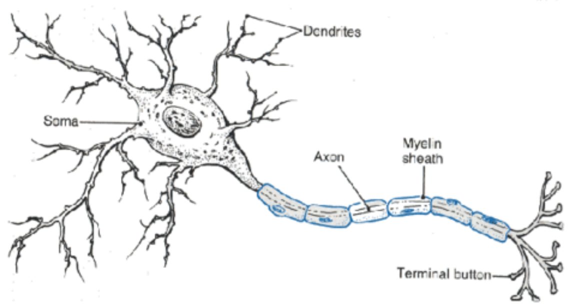
A unit in a neural network

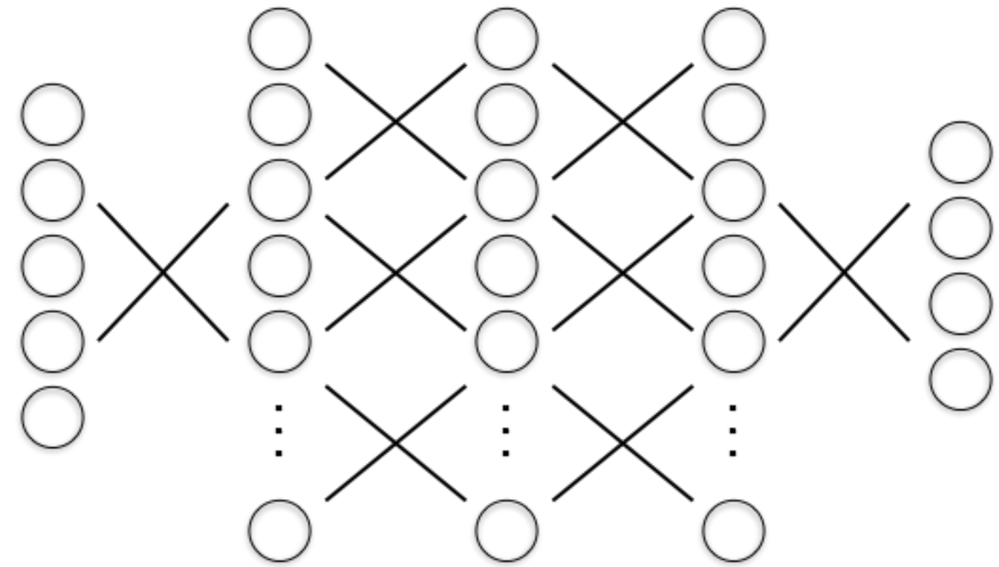


A unit in a neural network



Deep Neural Networks





- Deep neural networks
  - loosely motivated by biological neurons, networks
  - adjustable processing units (~ linear classifiers)
  - highly parallel, typically organized in layers
  - deep = many transformations (layers) before output

    e.g., edges -> simple parts-> parts -> objects -> scenes

Deep Learning

- Deep learning has overtaken a number of academic disciplines in just a few years
  - computer vision (e.g., image, scene analysis)
  - natural language processing (e.g., machine translation)
  - speech recognition
  - computational biology, etc.

Deep Learning

- Deep learning has overtaken a number of academic disciplines in just a few years
  - computer vision (e.g., image, scene analysis)
  - natural language processing (e.g., machine translation)
  - speech recognition
  - computational biology, etc.
- Key role in recent successes
  - self driving vehicles
  - speech interfaces
  - conversational agents
  - superhuman game playing

Deep Learning

- Deep learning has overtaken a number of academic disciplines in just a few years
  - computer vision (e.g., image, scene analysis)
  - natural language processing (e.g., machine translation)
  - speech recognition
  - computational biology, etc.
- Key role in recent successes
  - self driving vehicles
  - speech recognition
  - conversational agents
  - superhuman game playing
- Many more underway
  - personalized/automated medicine
  - chemistry, robotics, materials science, etc.

Deep learning … why now?

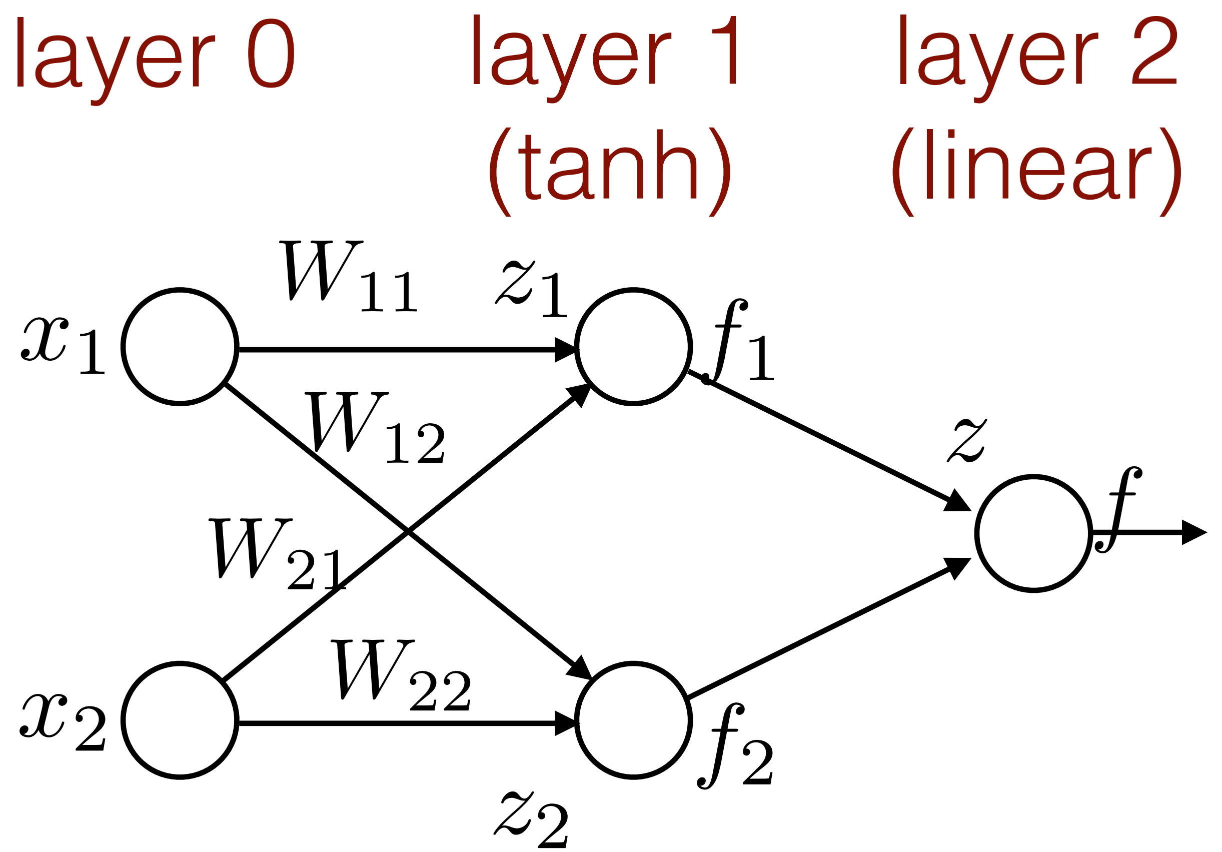
- **Reason #1:** lots of data
  - many significant problems can only be solved at scale

- **Reason #2:** computational resources (esp. GPUs)
  - platforms/systems that support running deep (machine) learning algorithms at scale

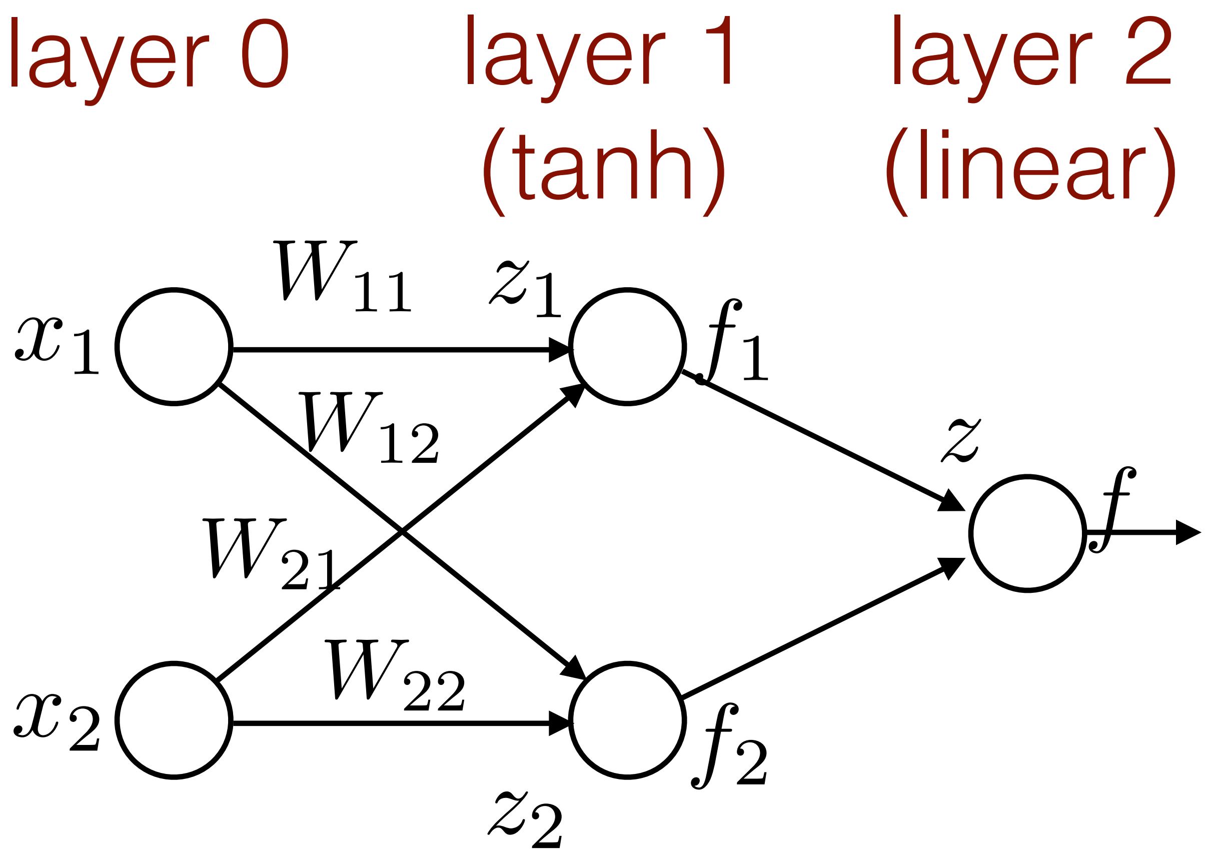
- **Reason #3:** large models are easier to train
  - large models can be successfully estimated with simple gradient based learning algorithms

- **Reason #4:** flexible neural “lego pieces”
  - common representations, diversity of architectural choices

One hidden layer model

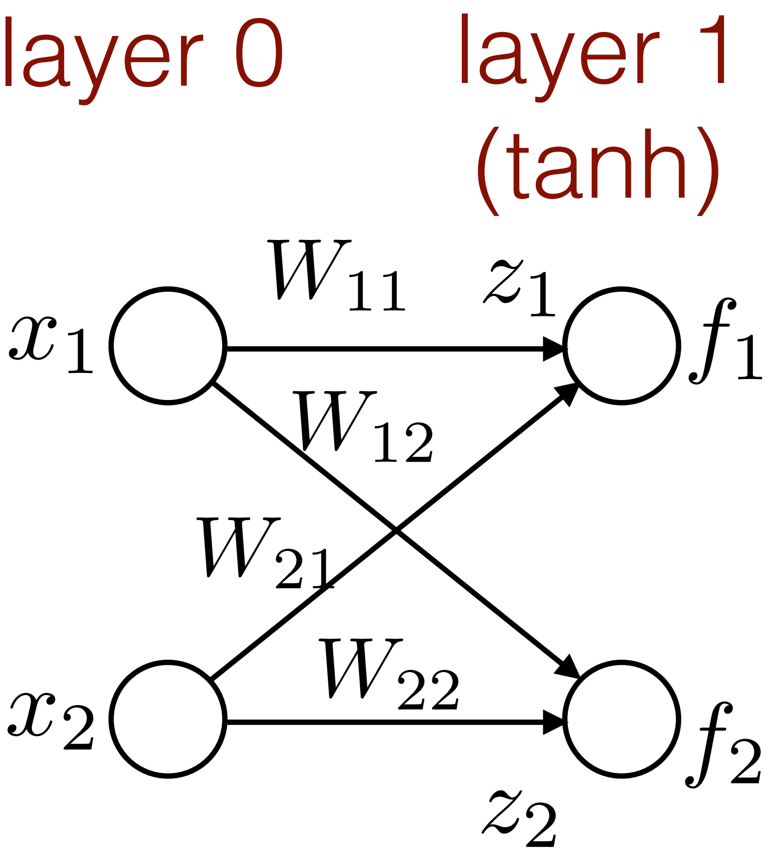


One hidden layer model

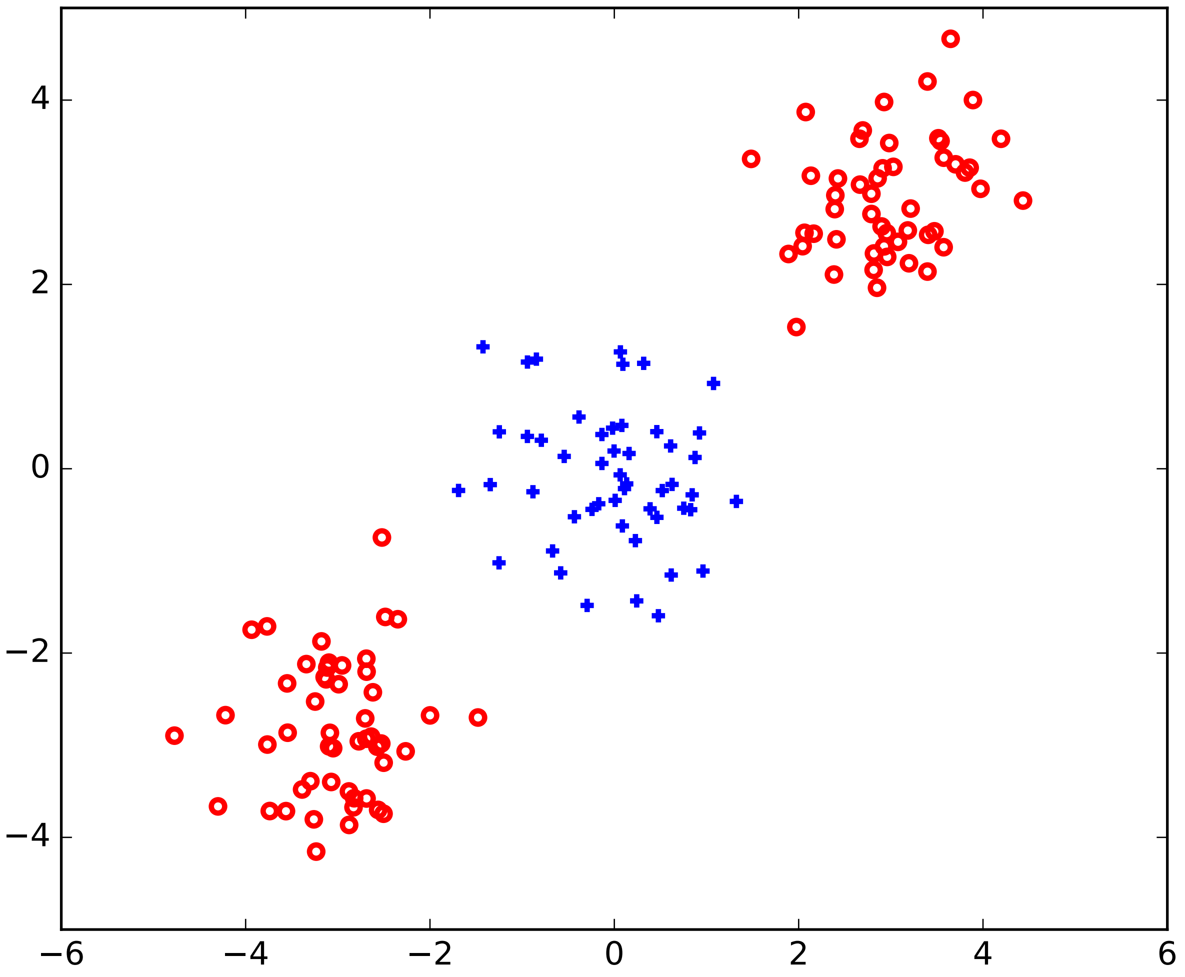


One hidden layer model

- Neural signal transformation

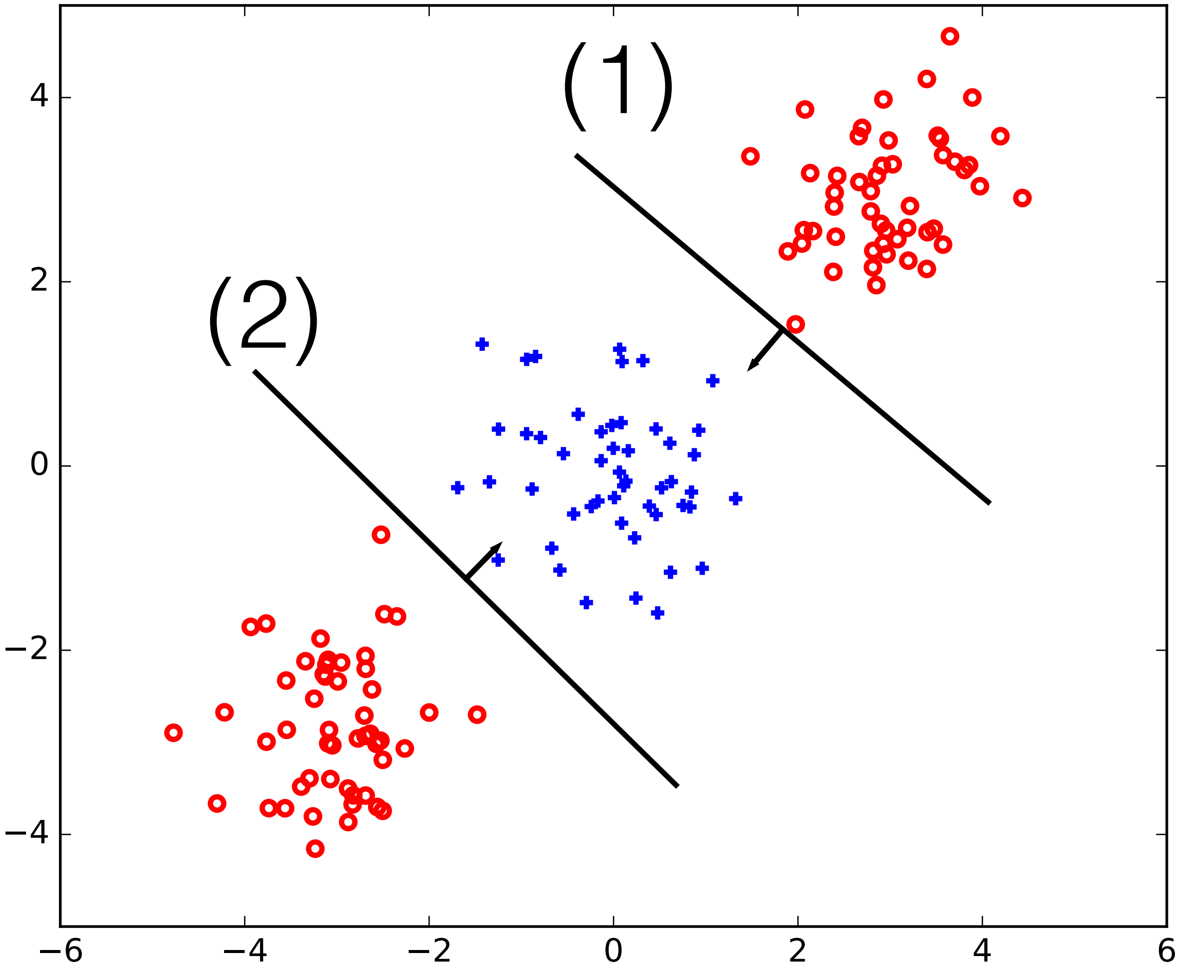


Example Problem



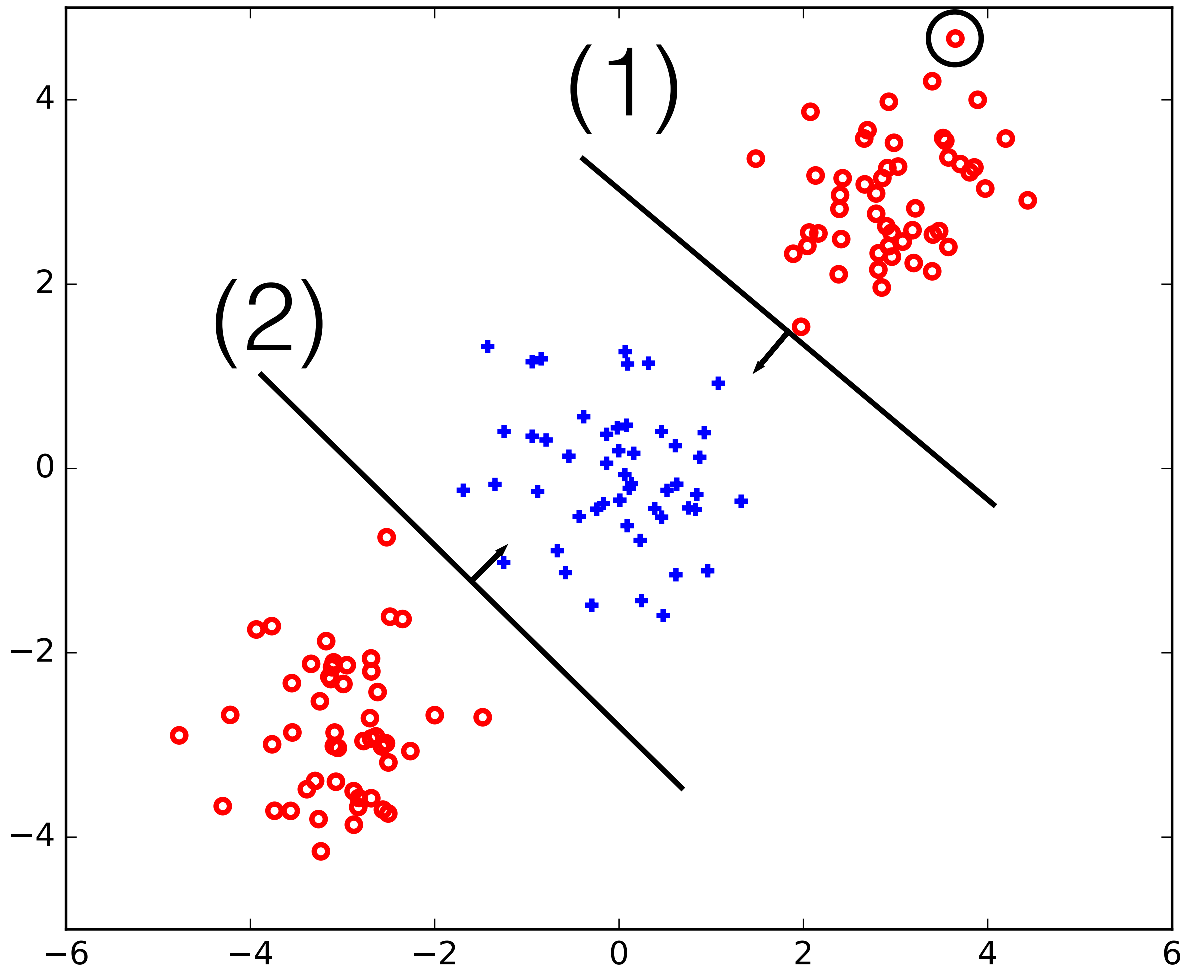
Hidden layer representation

Hidden layer units

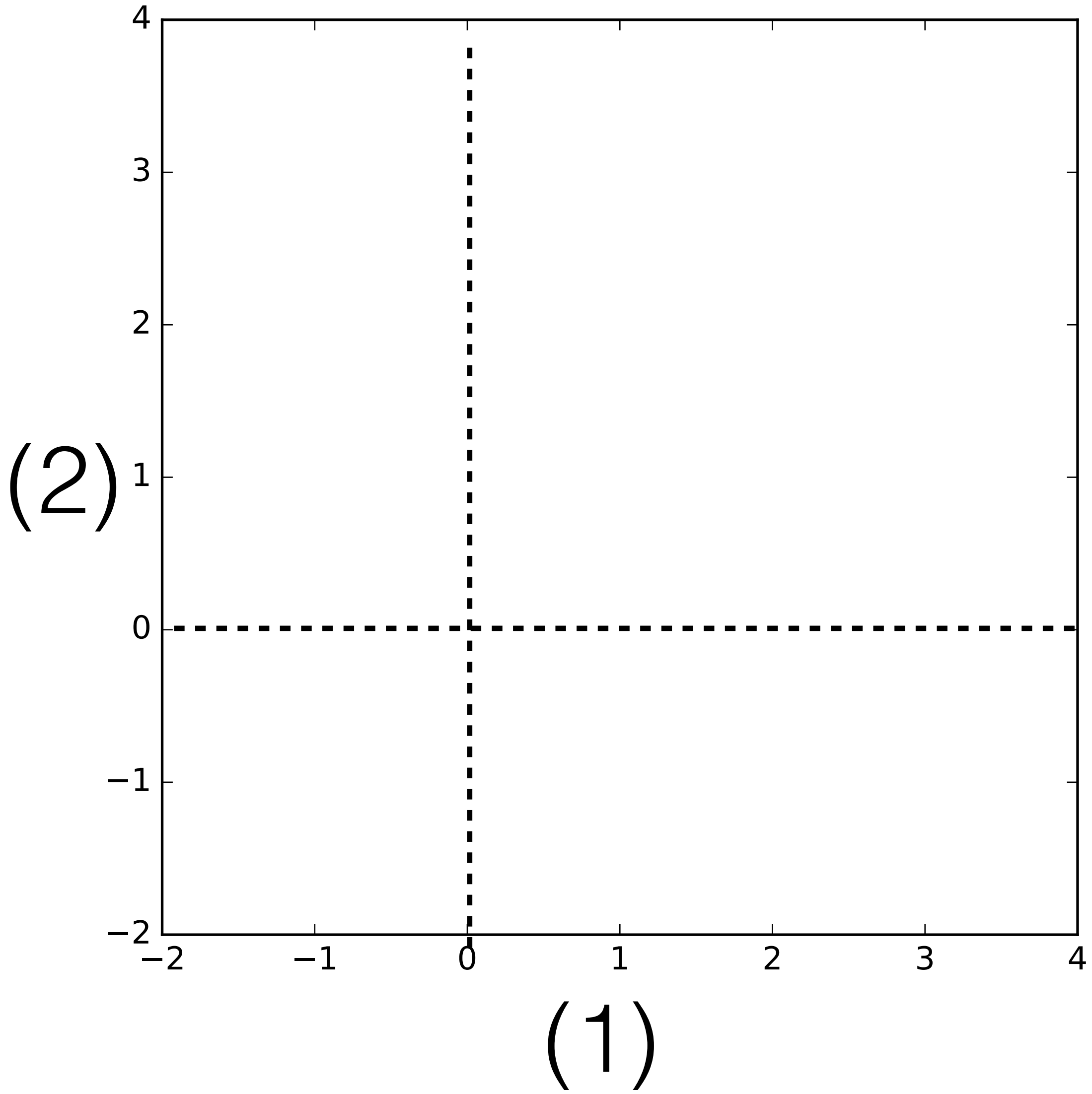


Hidden layer representation

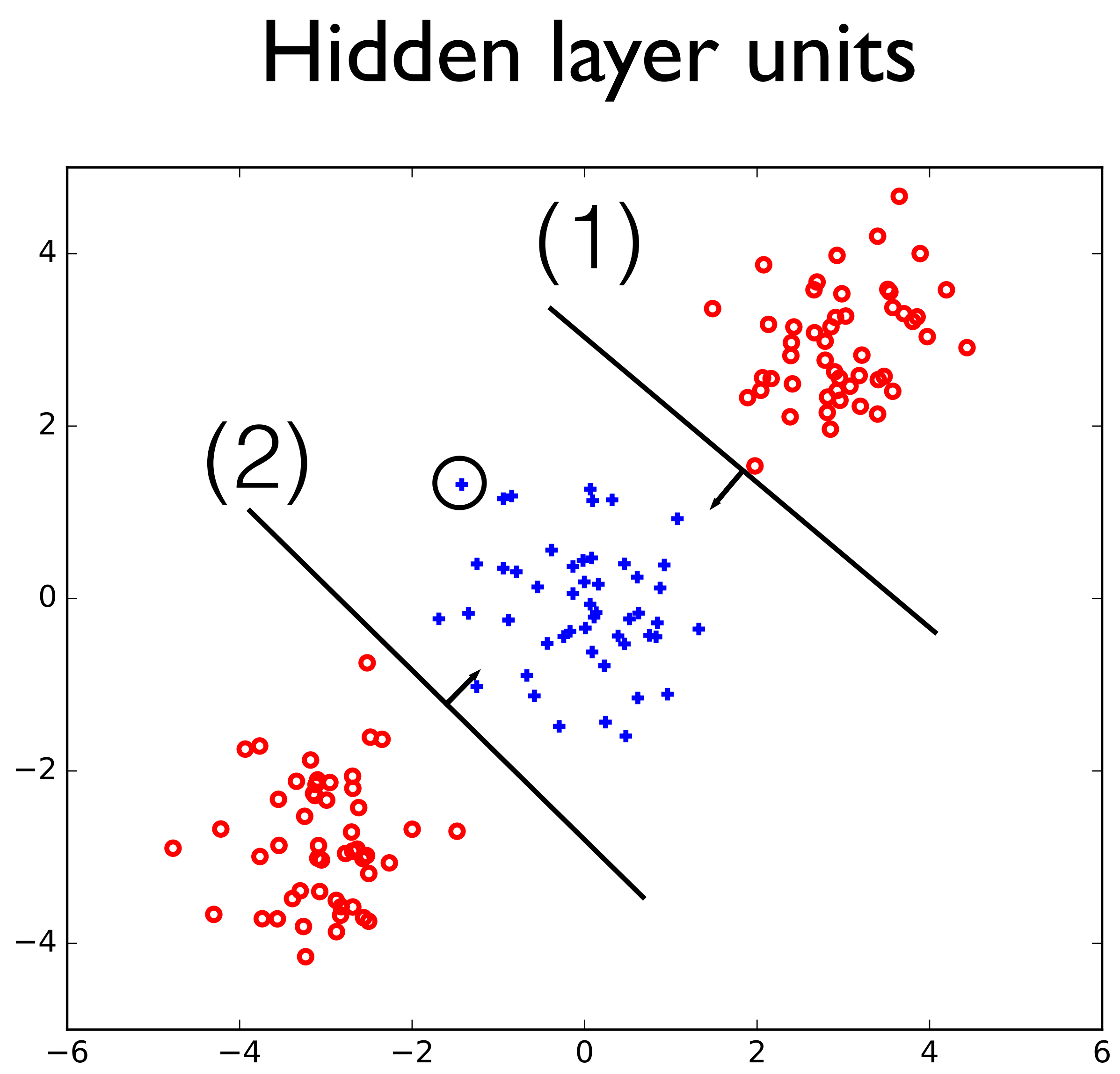
Hidden layer units

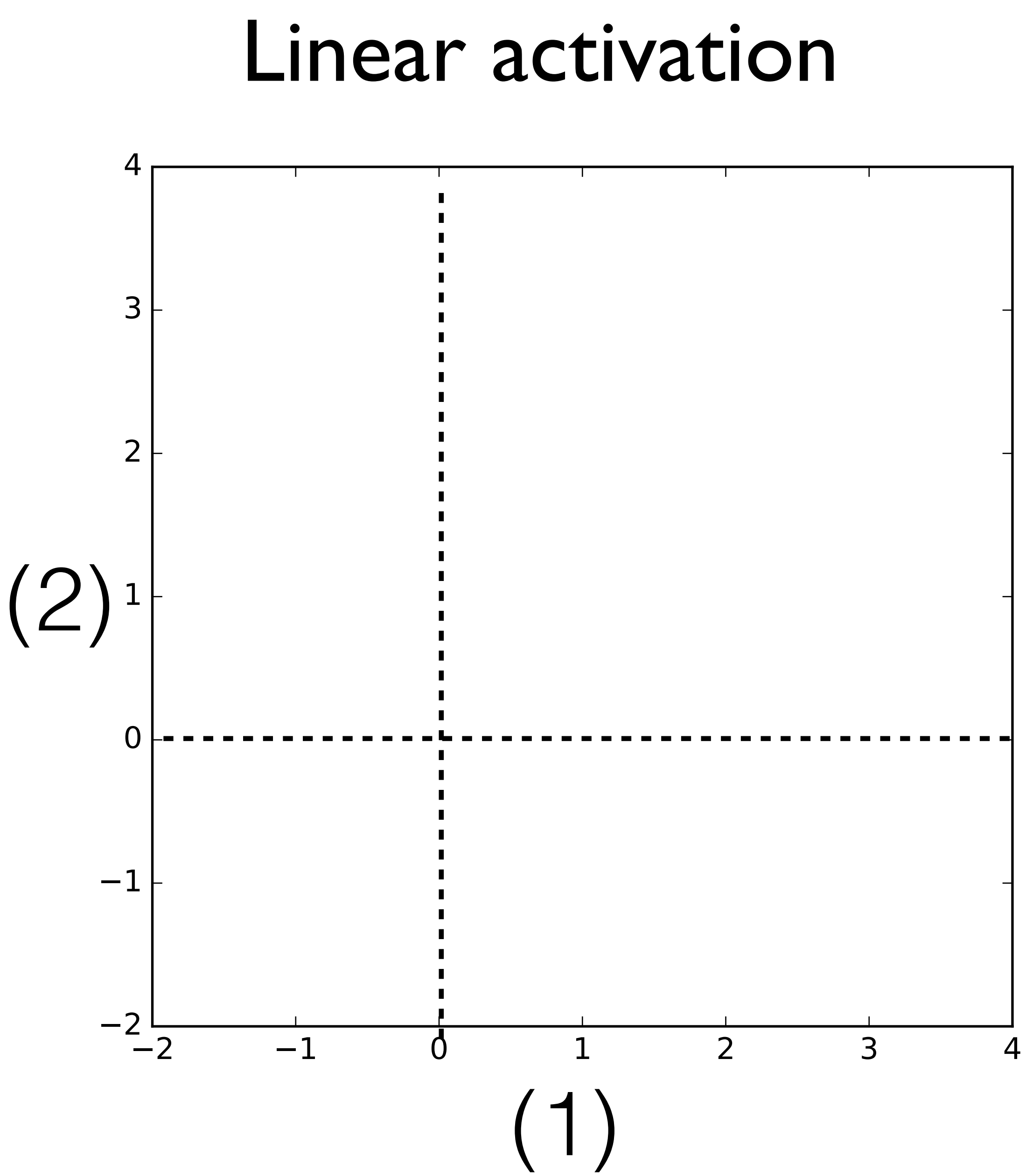


Linear activation

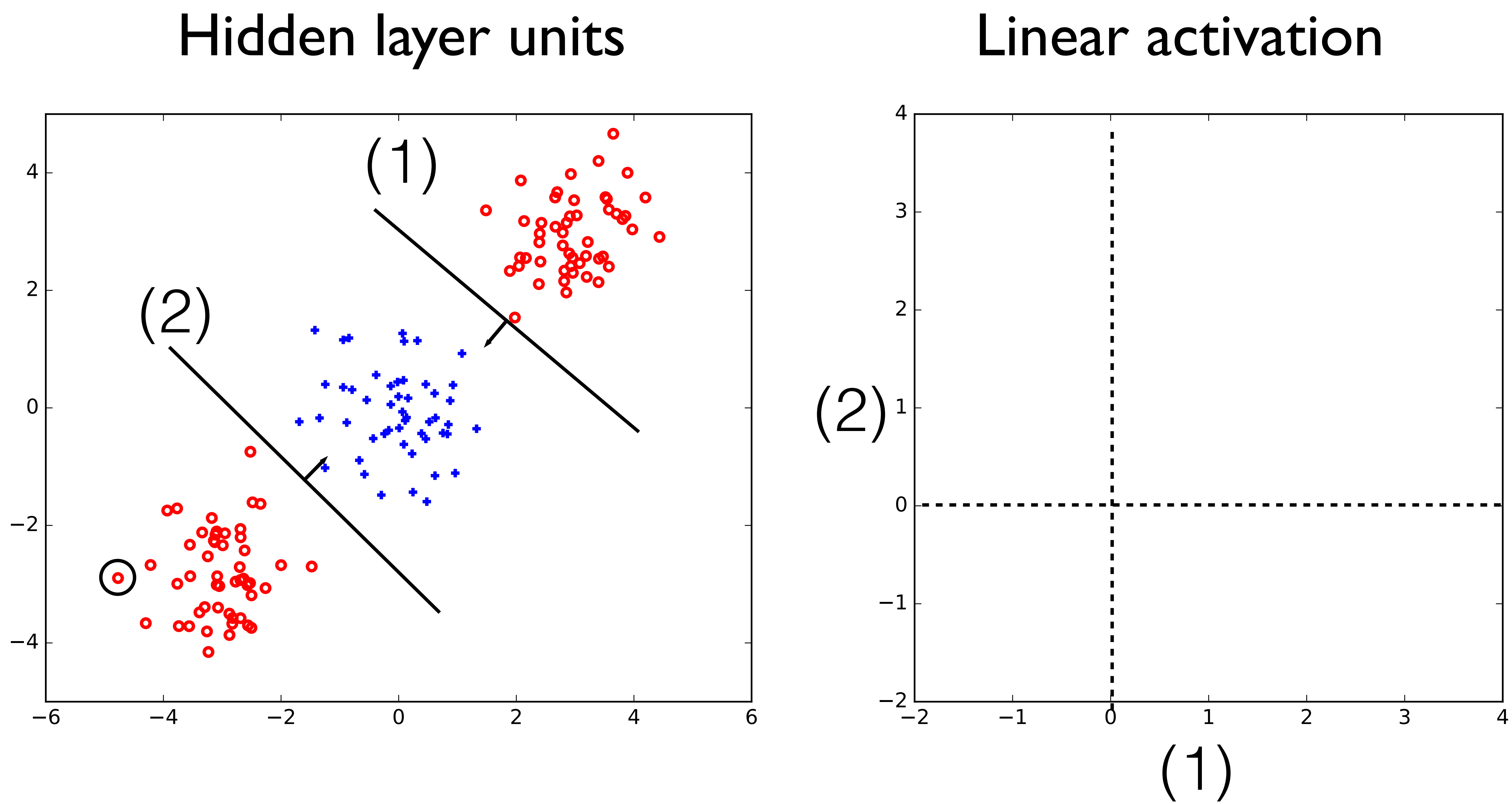


Hidden layer representation



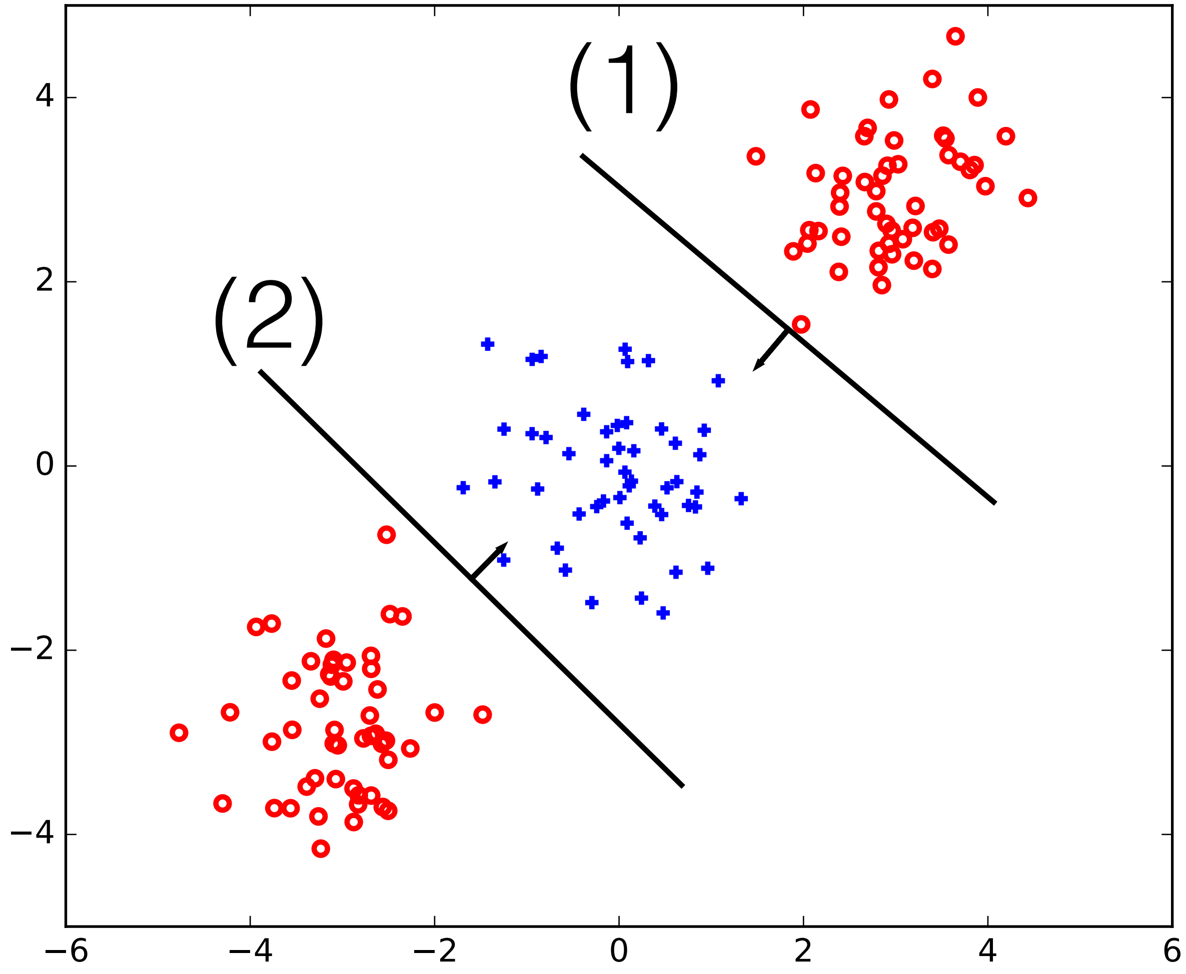


Hidden layer representation

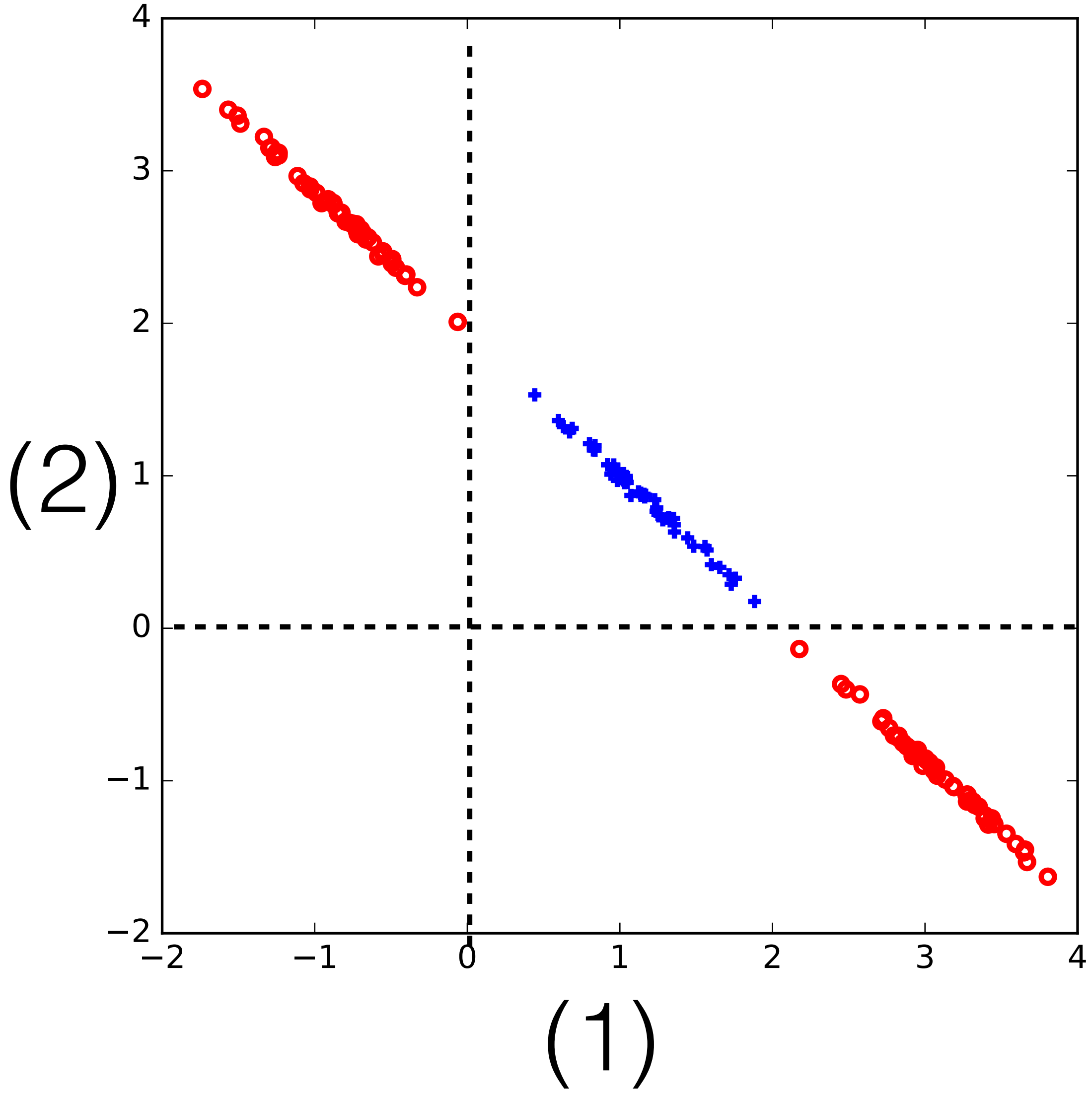


Hidden layer representation

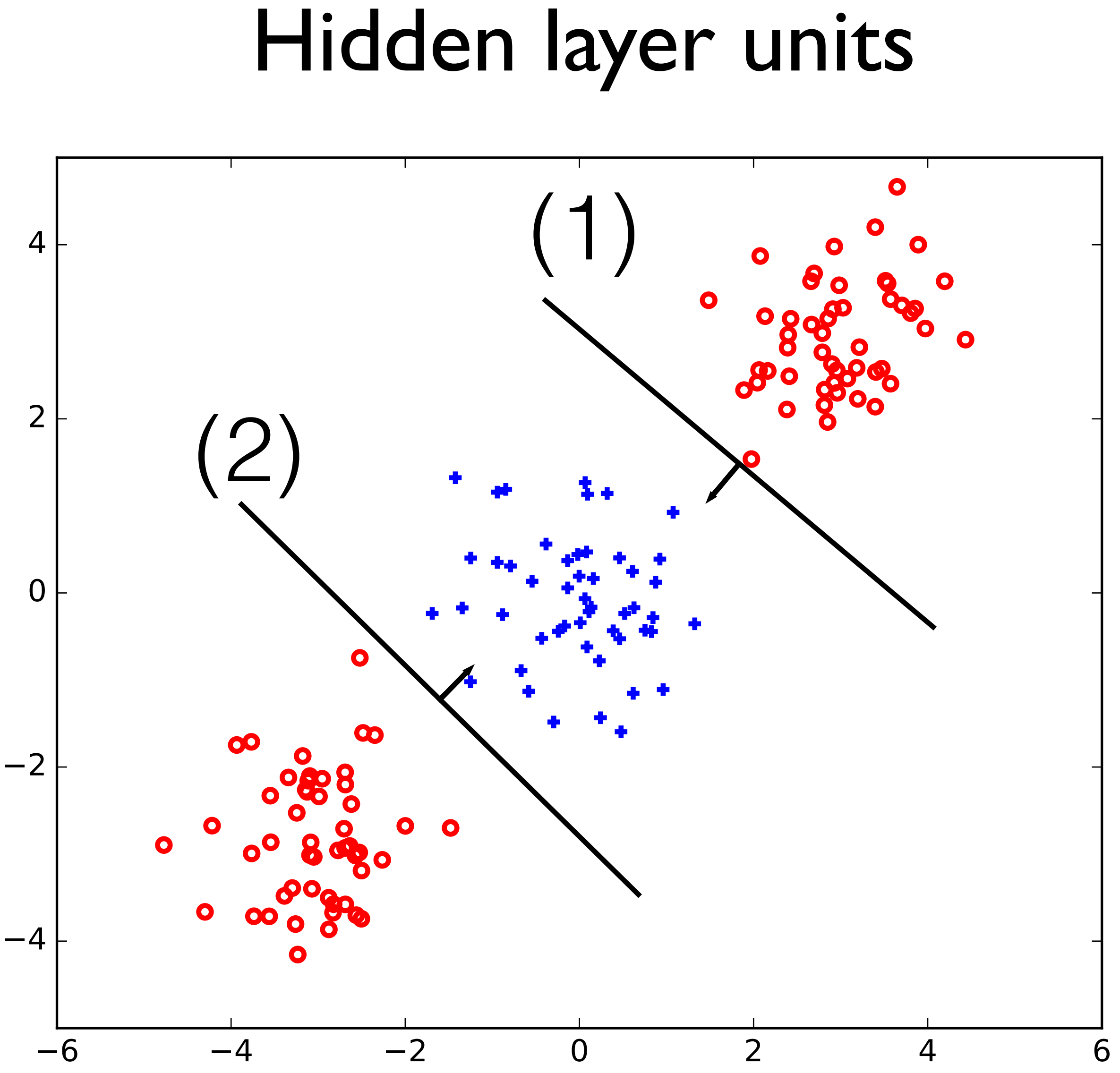
Hidden layer units

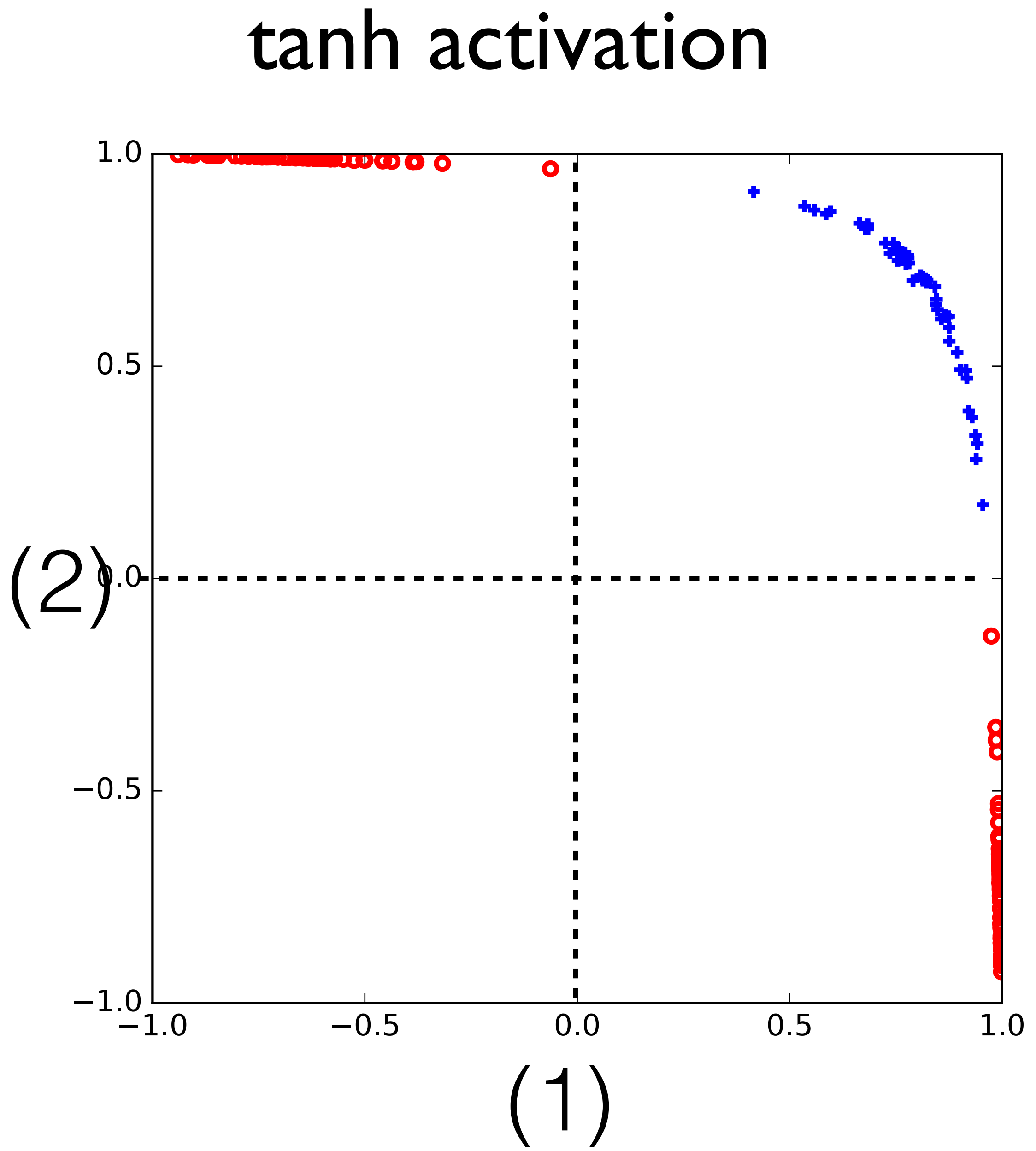


Linear activation



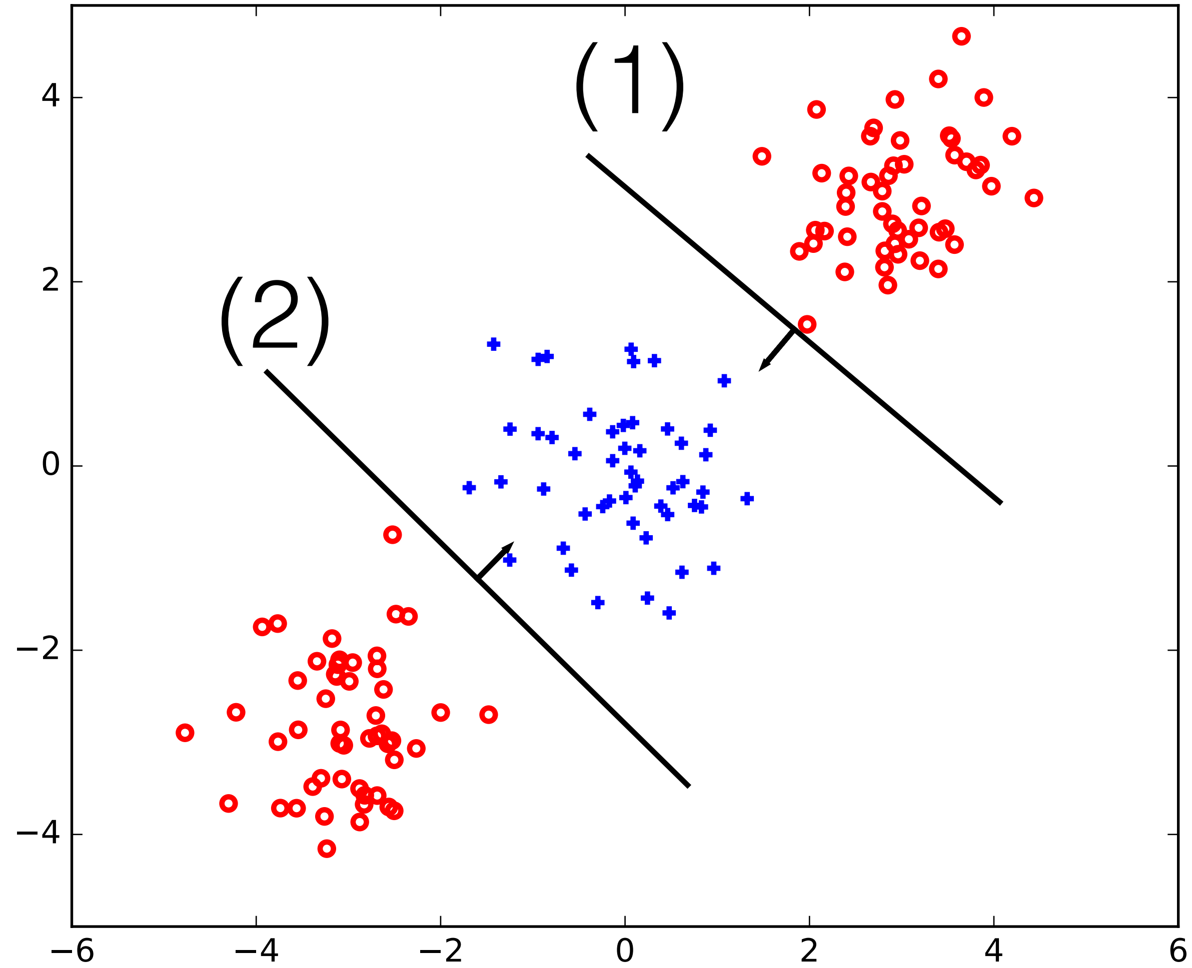
Hidden layer representation



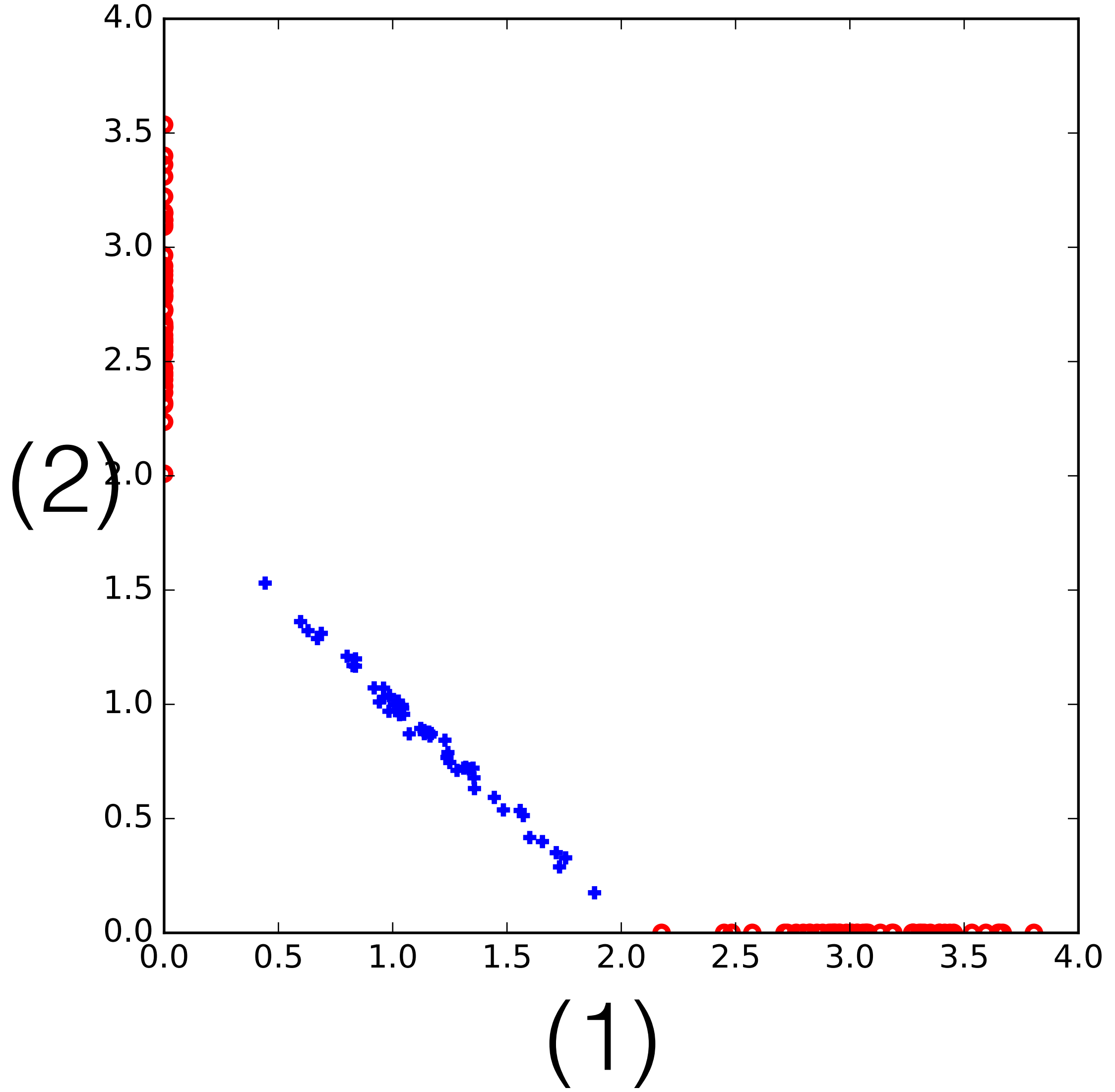


Hidden layer representation

Hidden layer units

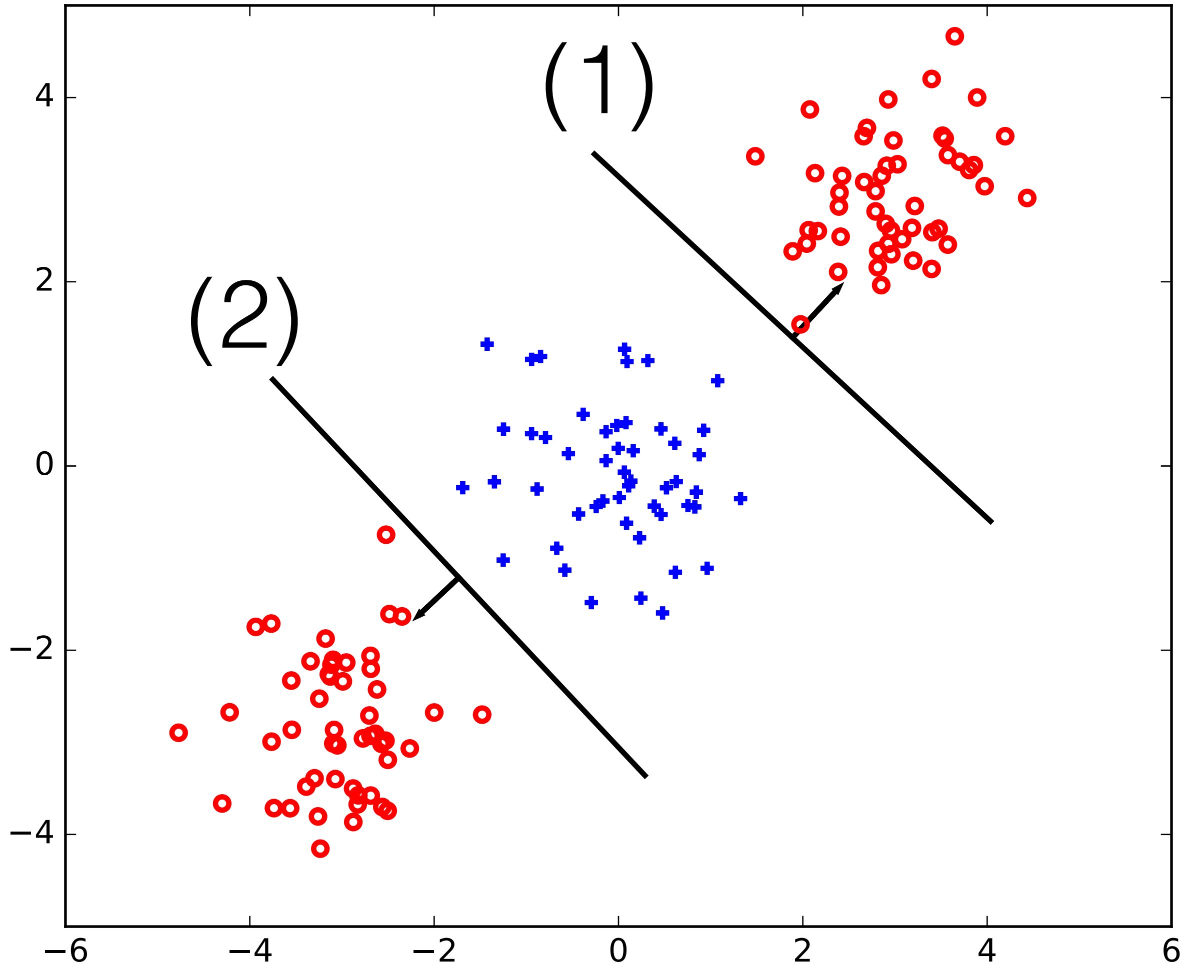


ReLU activation



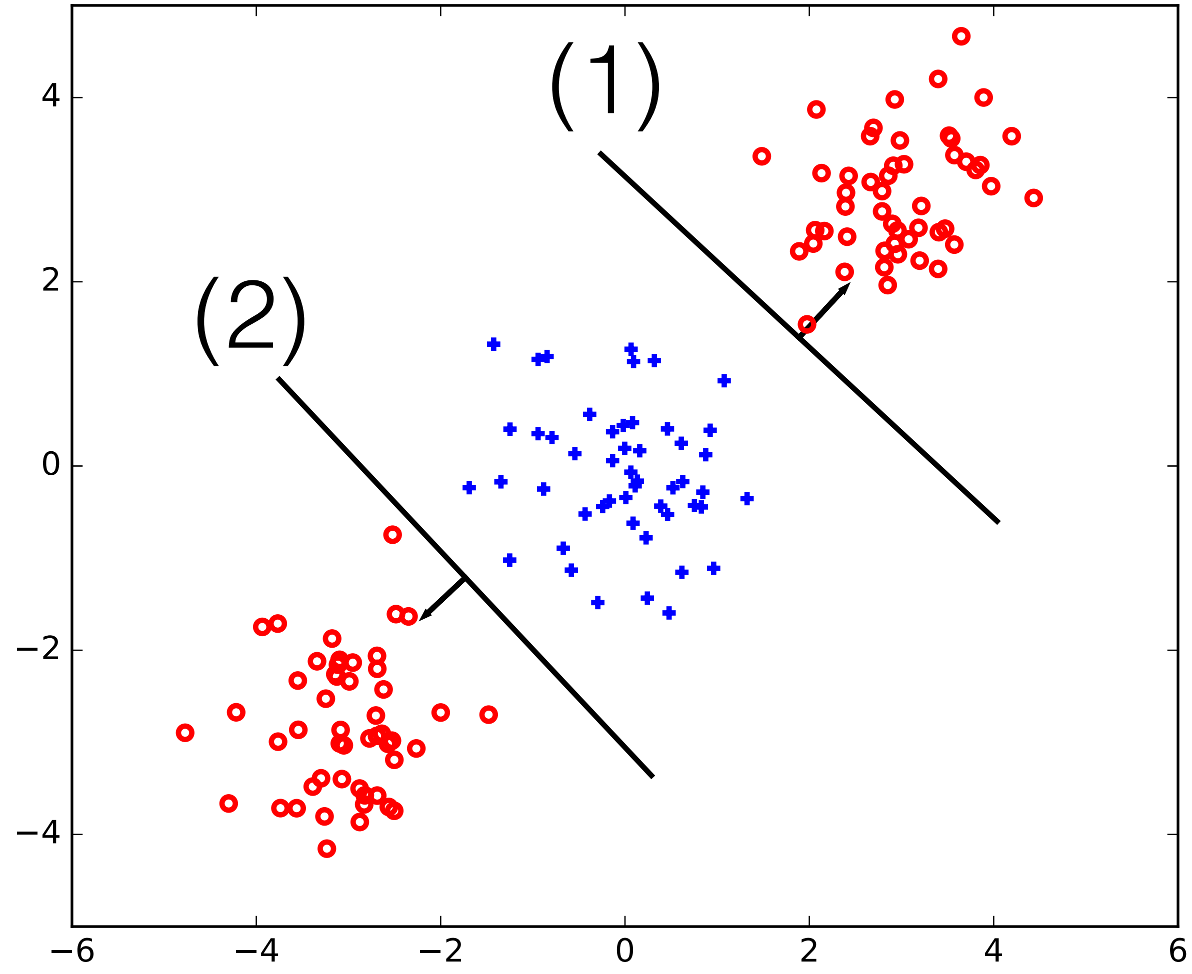
Does orientation matter?

Hidden layer units

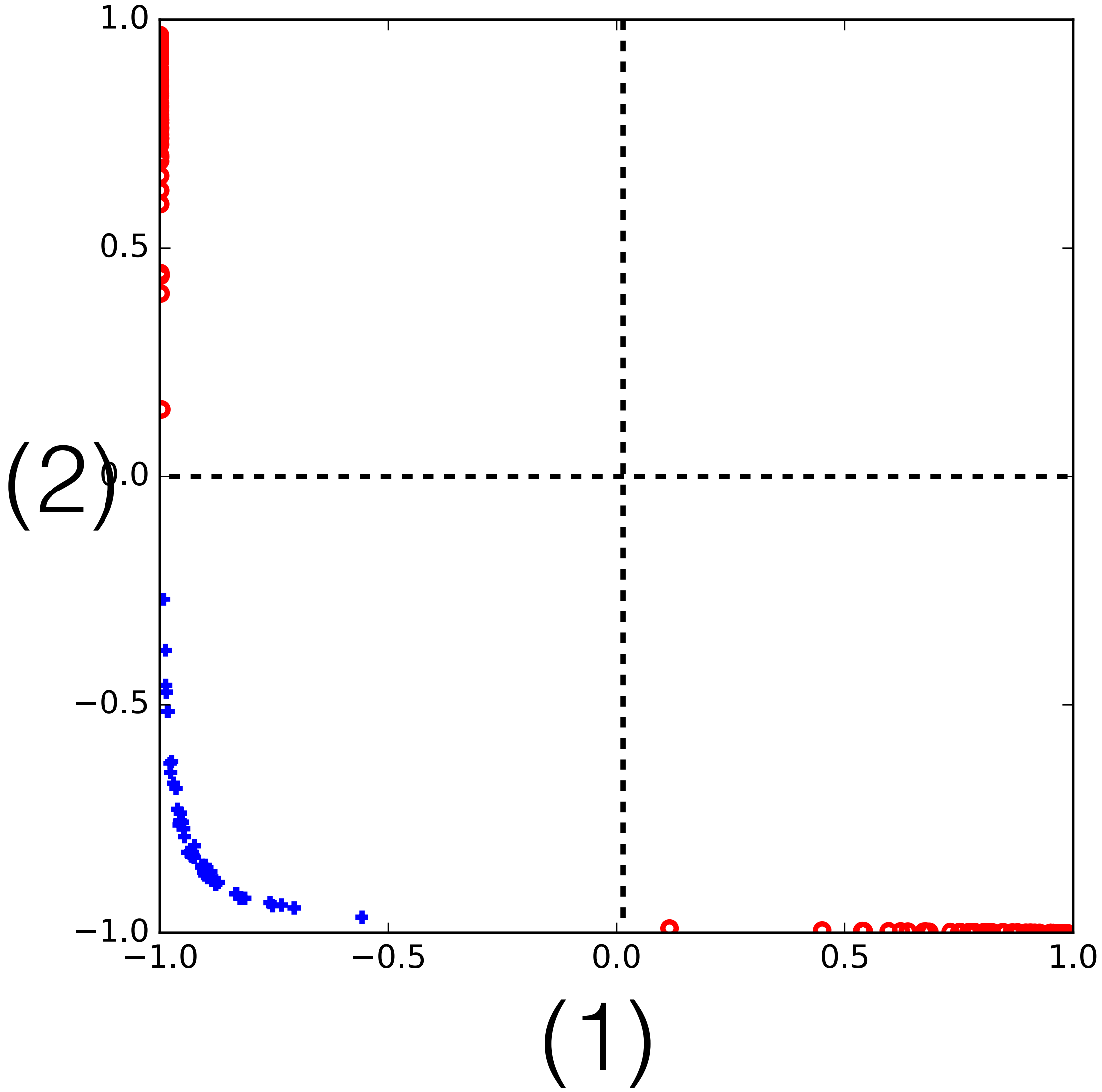


Does orientation matter?

Hidden layer units

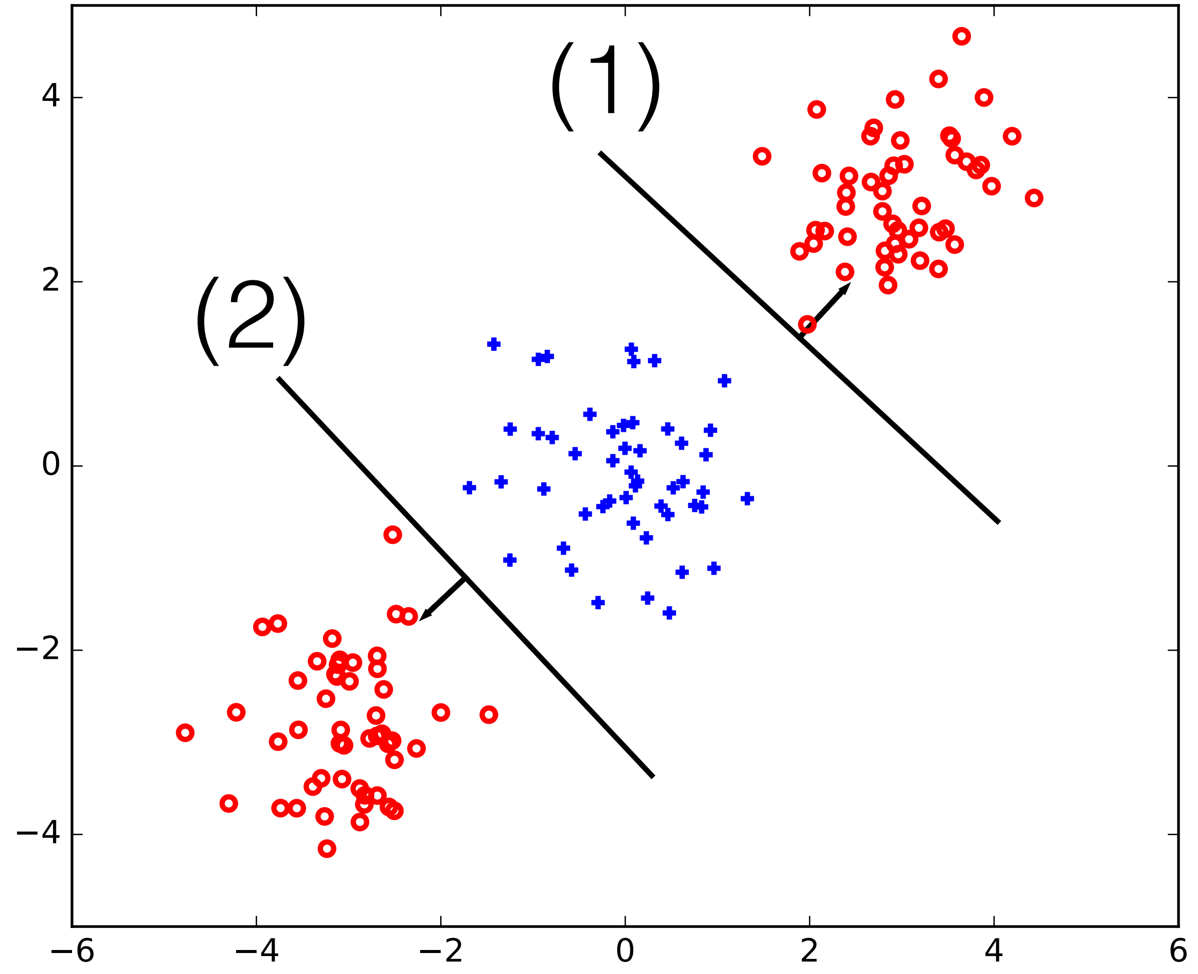


tanh activation

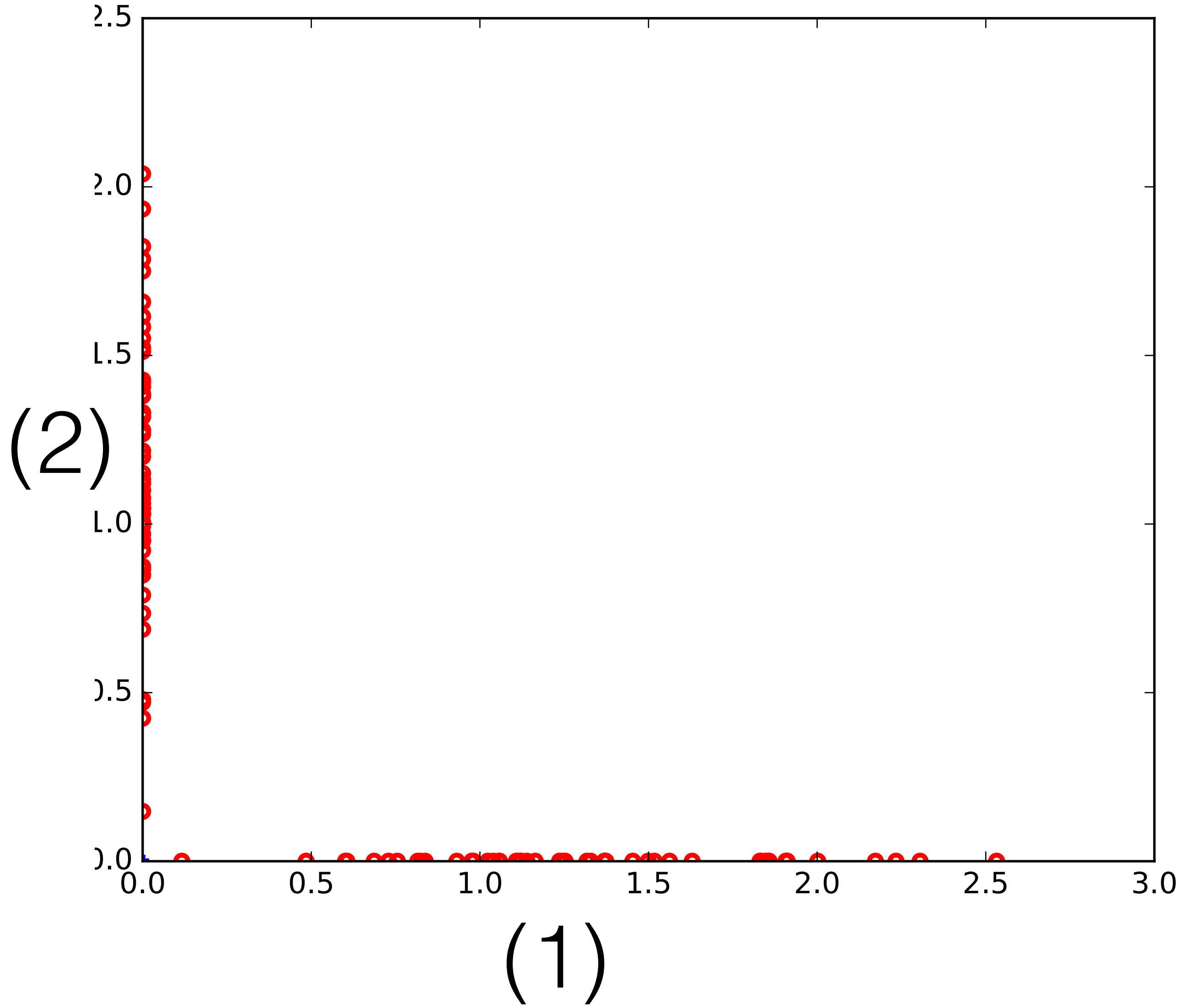


Does orientation matter?

Hidden layer units

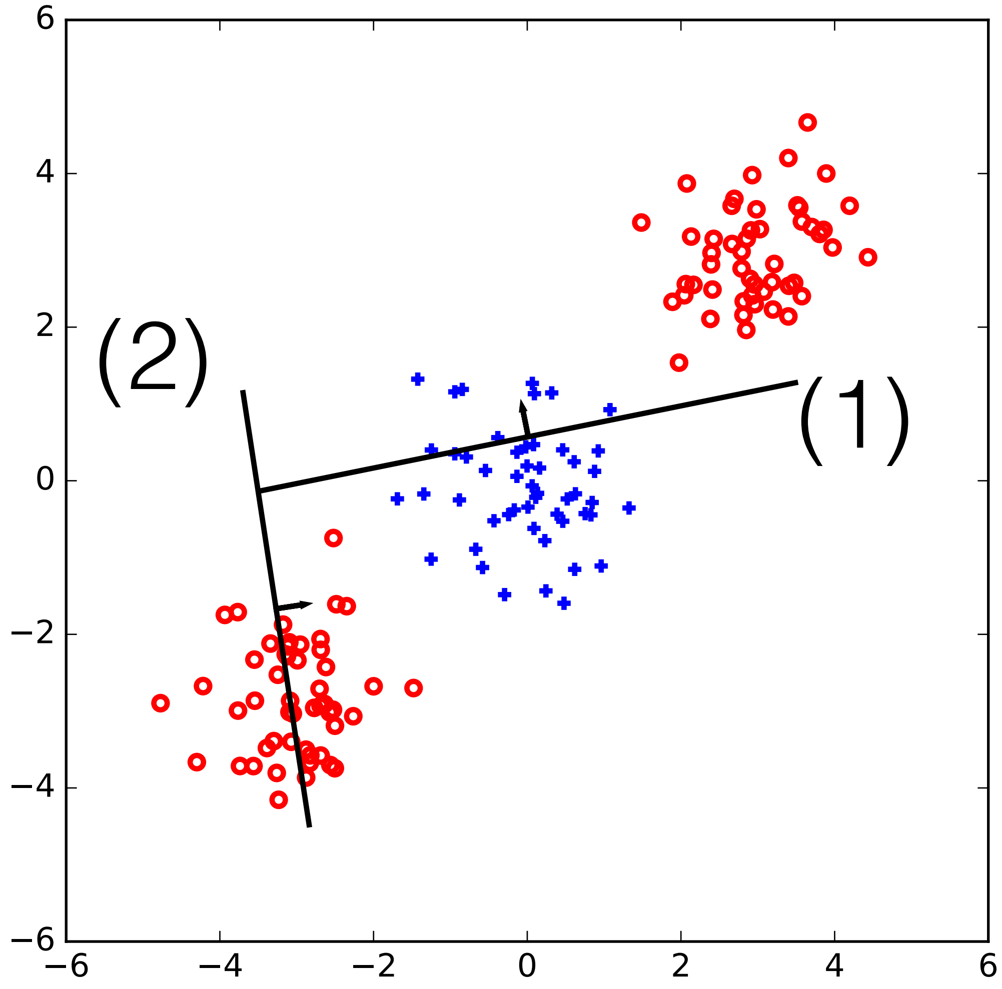


ReLU activation



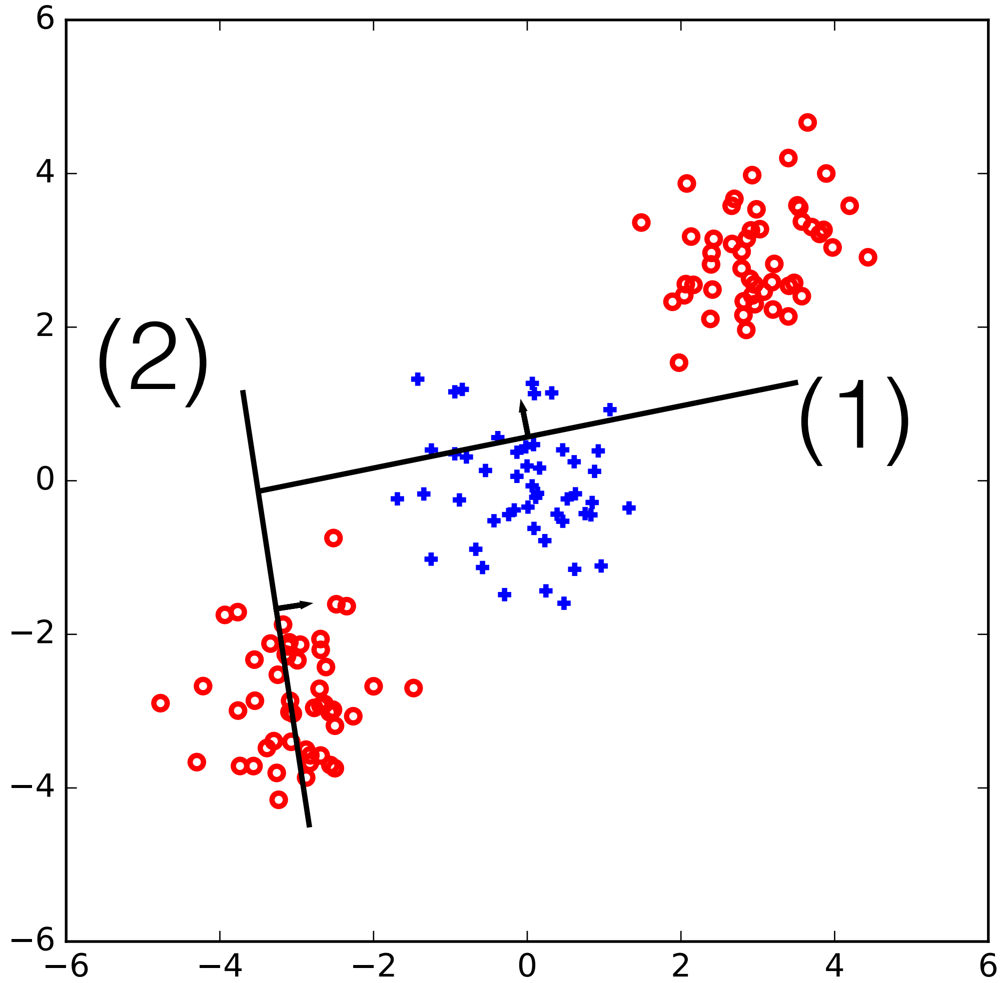
Random hidden units

Hidden layer units

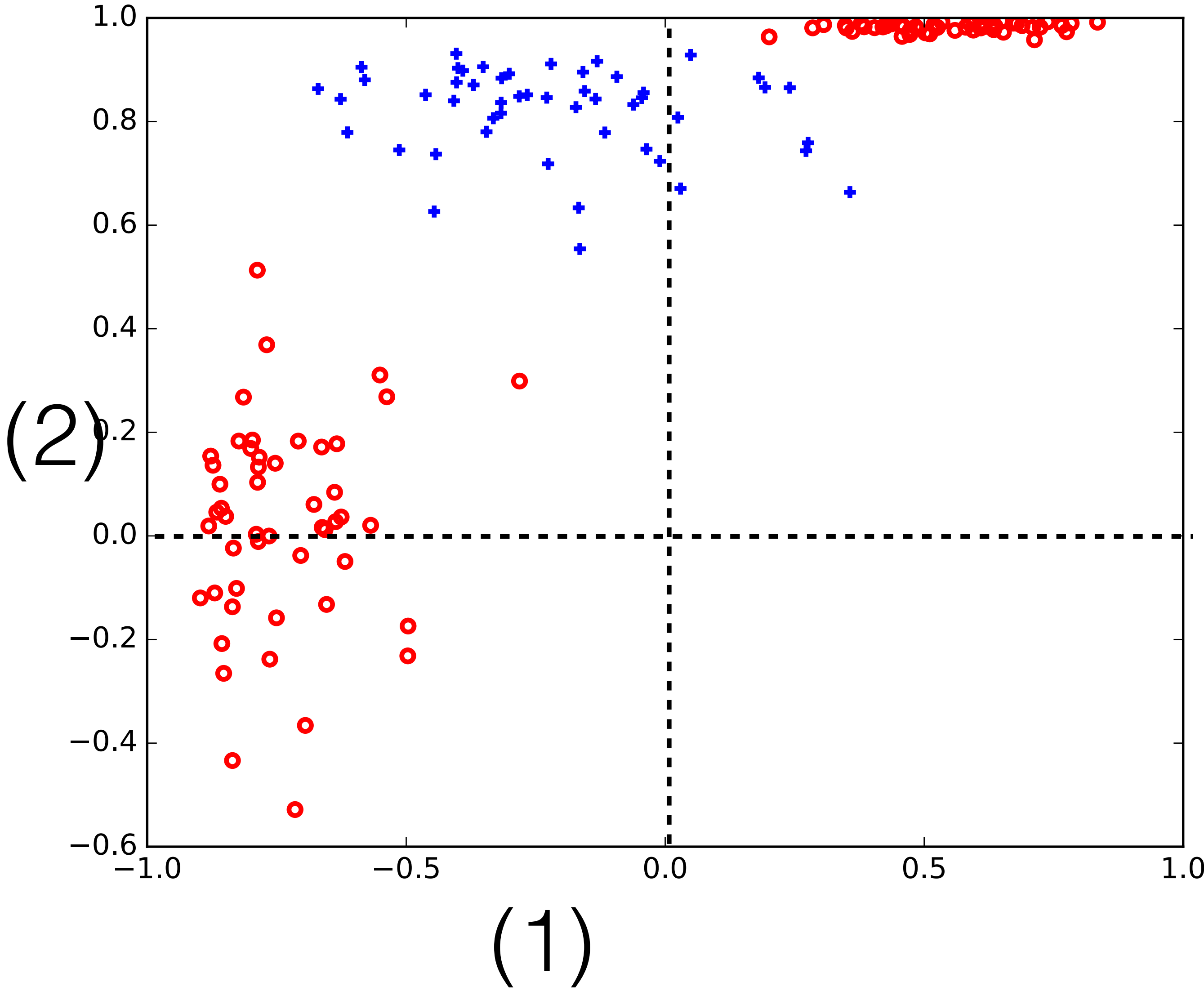


Random hidden units

Hidden layer units

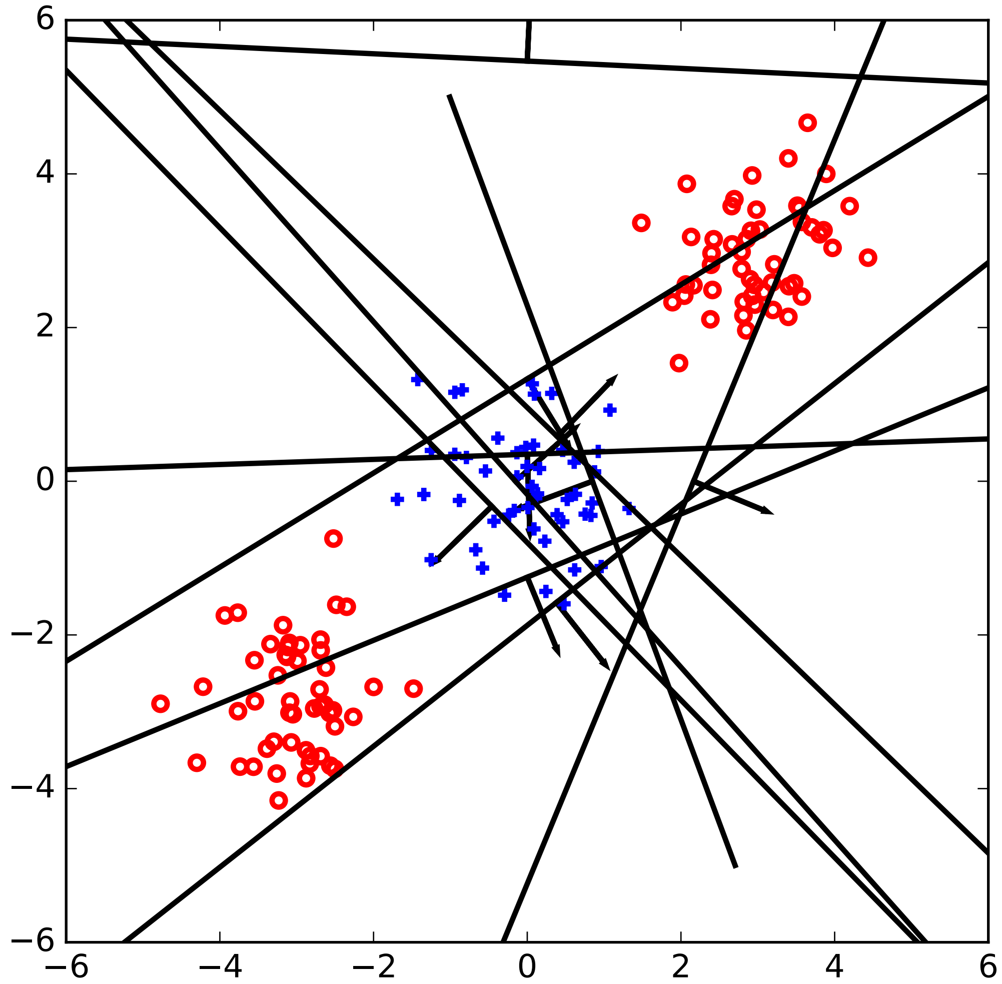


tanh activation



Random hidden units

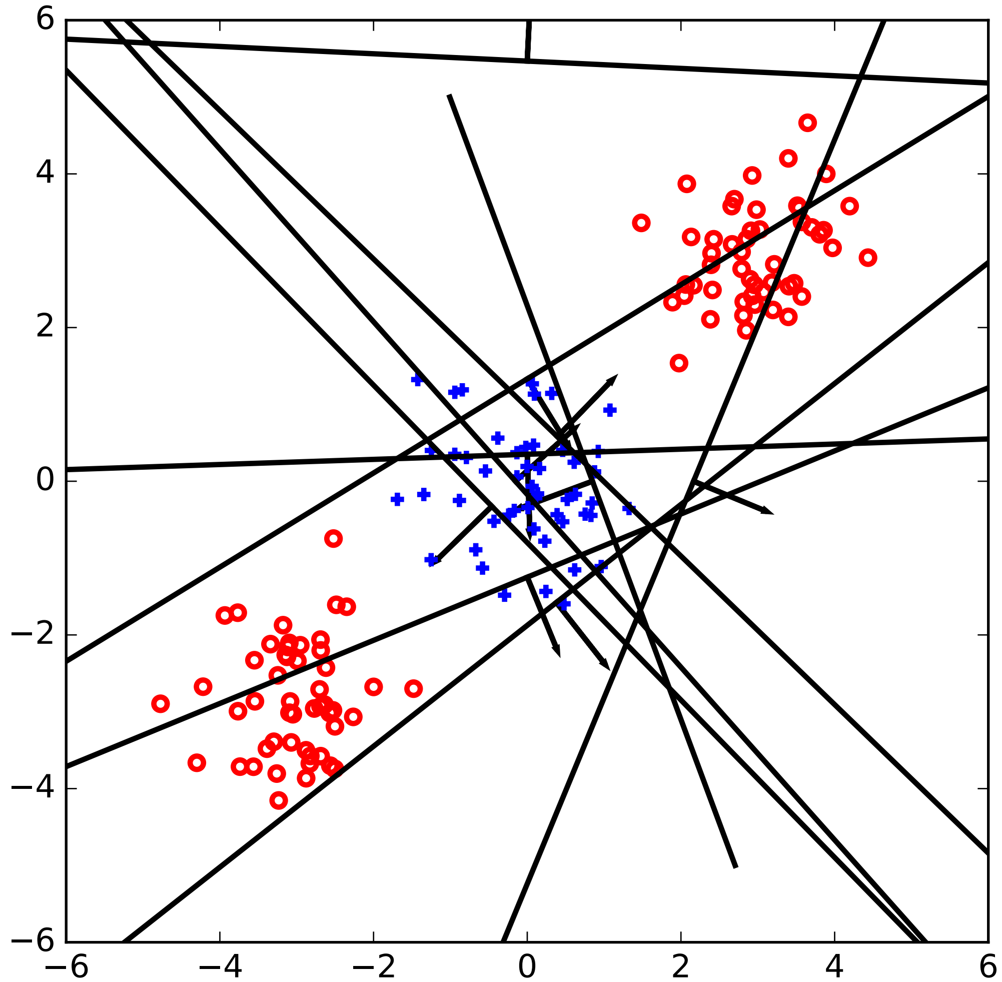
Hidden layer units



(10 randomly chosen units)

Random hidden units

Hidden layer units

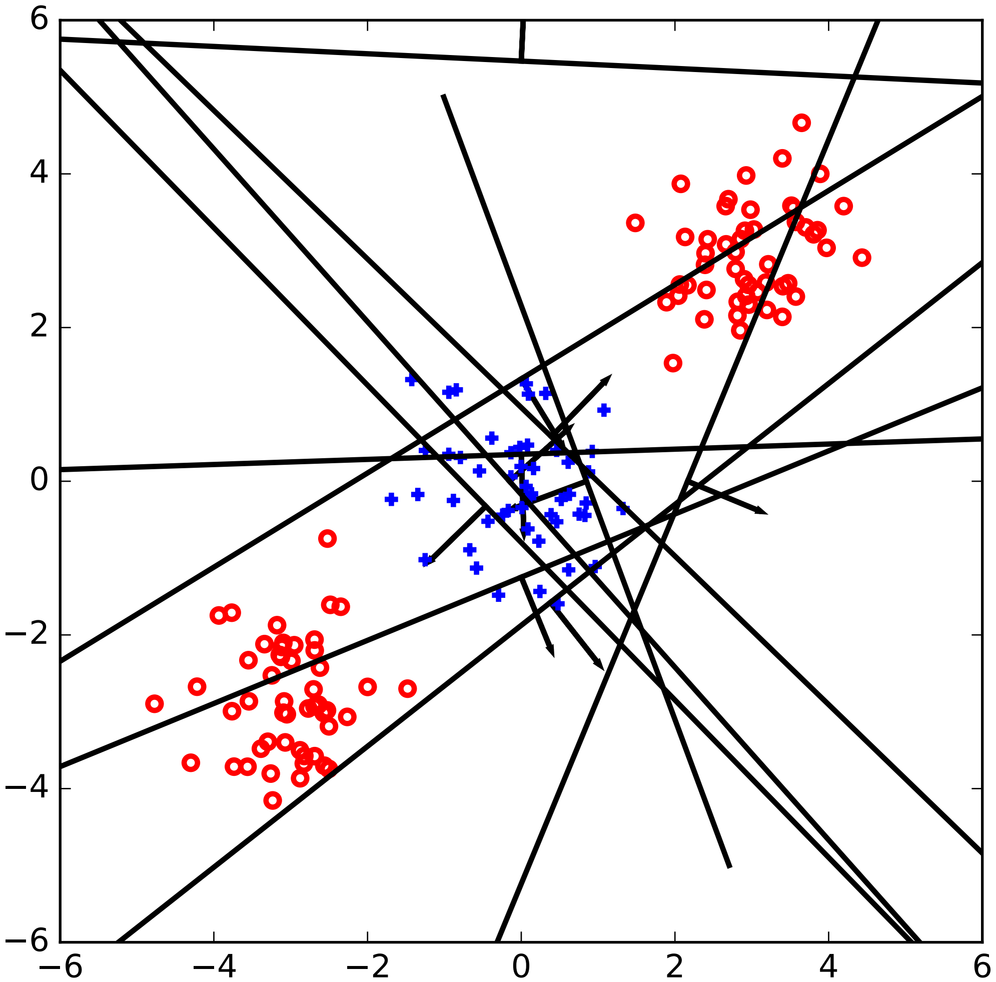


(10 randomly chosen units)

Are the points linearly separable in the resulting 10 dimensional space?

Random hidden units

Hidden layer units



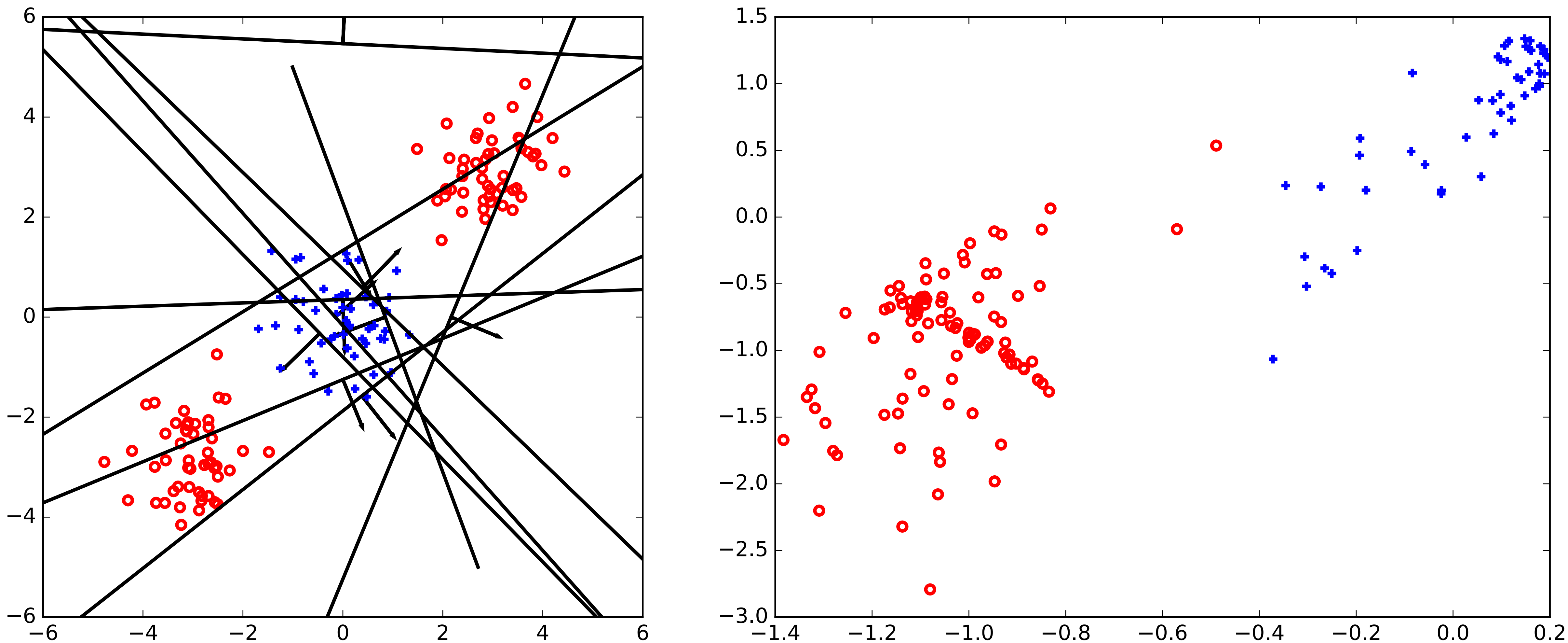
(10 randomly chosen units)

Are the points linearly separable in the resulting 10 dimensional space?

YES!

Random hidden units

Hidden layer units



(10 randomly chosen units)

what are the coordinates??

Summary

- Units in neural networks are linear classifiers, just with different output non-linearity
- The units in feed-forward neural networks are arranged in layers (input, hidden,…, output)
- By learning the parameters associated with the hidden layer units, we learn how to represent examples (as hidden layer activations)
- The representations in neural networks are learned directly to facilitate the end-to-end task
- A simple classifier (output unit) suffices to solve complex classification tasks if it operates on the hidden layer representations