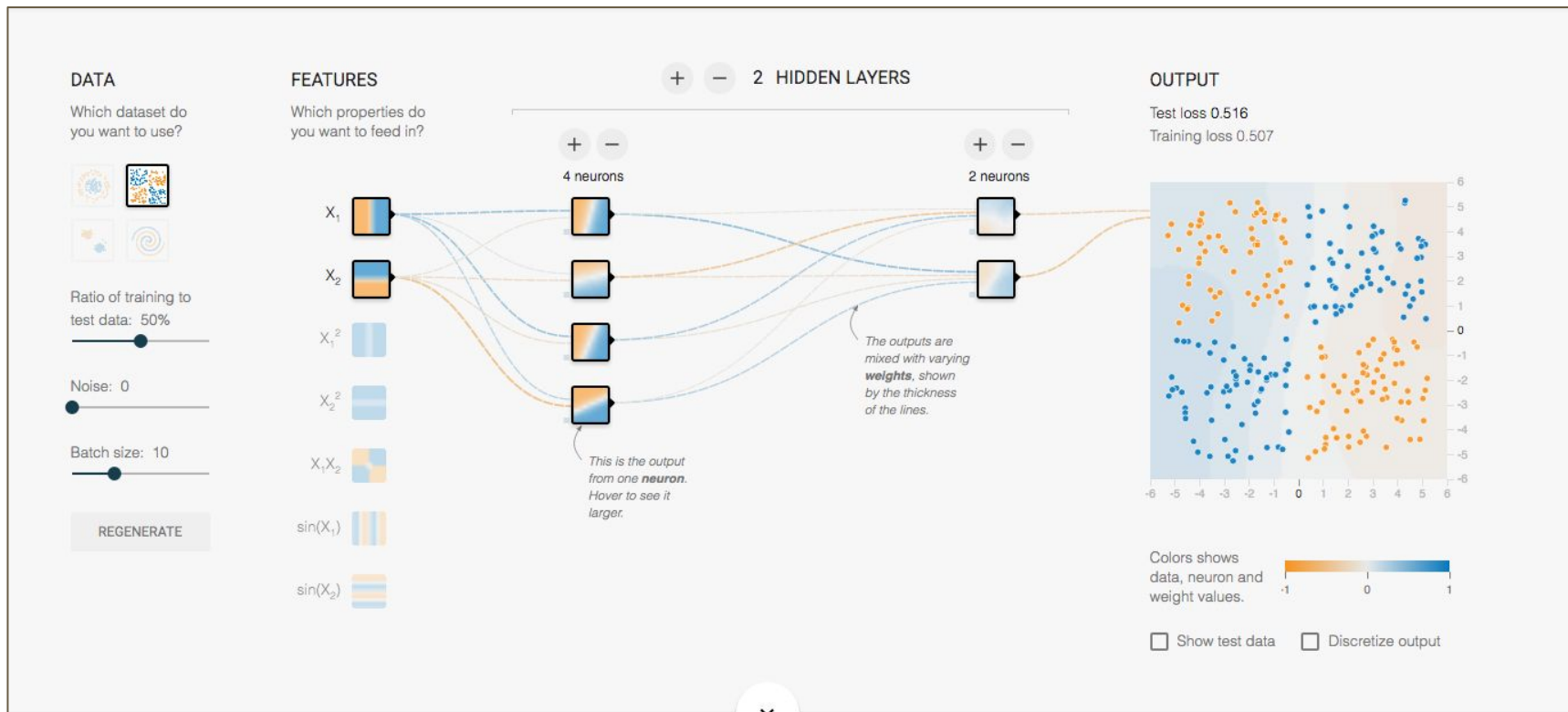

Intuitions for Neural Networks

— Trại Hè Toán và Khoa học MaSSP —

Môn Tin học

Hà Nội, Tháng 7/2016

Previously



Linear regression

- NN + No hidden layer = Linear Regression
- Why? Kết hợp tuyến tính của các trọng số inputs.



Epoch
000,328

Learning rate
0.03

Activation
Tanh

Regularization
None

Regularization rate
0

Problem type
Classification

DATA
Which dataset do you want to use?

Ratio of training to test data: 50%

Noise: 0

Batch size: 10

REGENERATE

FEATURES
Which properties do you want to feed in?

+ - 0 HIDDEN LAYERS

X_1

X_2

X_1^2

X_2^2

$X_1 X_2$

$\sin(X_1)$

$\sin(X_2)$

OUTPUT
Test loss 0.000
Training loss 0.000

Colors shows data, neuron and weight values.

☐ Show test data ☐ Discretize output

Um, What Is a Neural Network?

It's a technique for building a computer program that learns from data. It is based very loosely on how we think the human brain works. First, a

Logistics regression

- NN + One hidden layer = Logistics Regression
- Why? Feeding tổ hợp tuyến tính của inputs để bẻ thành phi tuyến tính
- Simple network + simple feature set (X, Y) can do a lot !



Epoch
000,276

Learning rate
0.03

Activation
Tanh

Regularization
None

Regularization rate
0

Problem type
Classification

DATA

Which dataset do you want to use?

Ratio of training to test data: 50%

Noise: 0

Batch size: 10

REGENERATE

FEATURES

Which properties do you want to feed in?

X_1

X_2

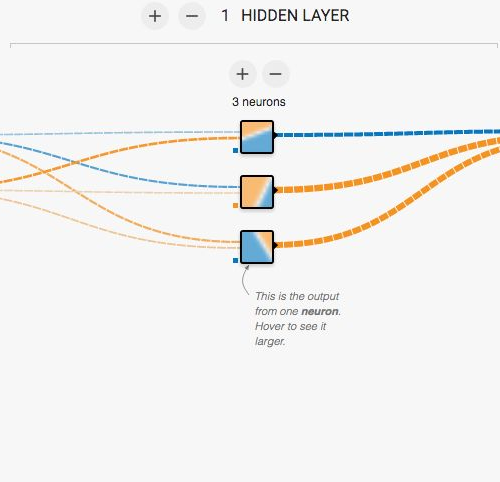
X_1^2

X_2^2

$X_1 X_2$

$\sin(X_1)$

$\sin(X_2)$



OUTPUT

Test loss 0.052

Training loss 0.074

Colors shows data, neuron and weight values.

☐ Show test data ☐ Discretize output

Um, What Is a Neural Network?

It's a technique for building a computer program that learns from data. It is based very loosely on how we think the human brain works. First, a



Epoch
000,280

Learning rate
0.03

Activation
Tanh

Regularization
None

Regularization rate
0

Problem type
Classification

DATA

Which dataset do you want to use?



Ratio of training to test data: 50%

Noise: 0

Batch size: 10

REGENERATE

FEATURES

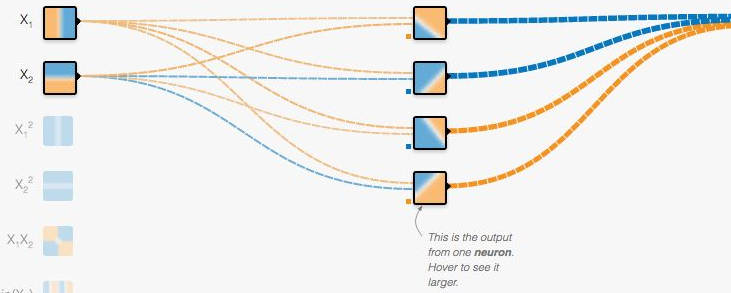
Which properties do you want to feed in?

- ☒ X_1
- ☒ X_2
- ☐ X_1^2
- ☐ X_2^2
- ☒ $X_1 X_2$
- ☐ $\sin(X_1)$
- ☐ $\sin(X_2)$

+ - 1 HIDDEN LAYER

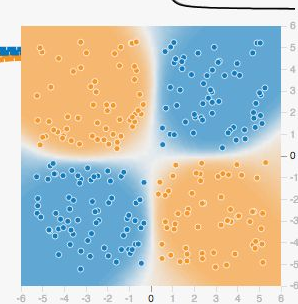
+ -

4 neurons



OUTPUT

Test loss 0.017
Training loss 0.014



Colors shows data, neuron and weight values.

☐ Show test data ☐ Discretize output



Um, What Is a Neural Network?

It's a technique for building a computer program that learns from data. It is based very loosely on how we think the human brain works. First, a

Sức mạnh của hidden layers

- Càng nhiều hidden layers thì NNs sẽ càng mạnh hơn để xấp xỉ được bất kỳ hàm nào

Điều chỉnh neural networks như nào

- Feature set
- Số hidden layers
- Số neurons trên mỗi layers
- Learning rates
- Activation
- Regularization & regularization rate

Điều chỉnh neural networks như nào

- Không đáp án chính xác nào, nhưng có thể hình thành trực giác về cách điều chỉnh các tham số của NNs
- Tối ưu hyper parameters là nghệ thuật của NNs.

Số hidden layers

- NNs càng nhiều hidden layers (càng sâu) thì càng mạnh.
- **Nhiều hidden layers (với ít neurons) >> Nhiều neurons mỗi layers với ít layers.**
- Successive layers lead to more increasing abstractions.

Caveat:

- Càng nhiều hidden layers thì càng tốn nhiều chi phí tính toán.
- Nhiều hidden layers có thể gây overfit.

Số neurons của mỗi layer

- Thêm neurons ở các layers cũng tăng sức mạnh cho networks.

Caveat:

- Poor generalization, dễ overfit hơn.
- Chi phí tính toán lớn

Playground

Classifying “Spiral Network”

4 x 4 x 2 with ReLU vs 6 x 6 with ReLU

4 x 4 x 3 with ReLU vs 6 x 6 x 2 with ReLU

Assignment:

- Hover từng neuron ở mỗi layer để xem output của mỗi neuron
- Chú ý xem ý nghĩa của các output trên là gì
- Để ý thấy có 1 ví dụ, tỉ lệ train/test là 70%. Điều này có nghĩa là gì ?.

Feature set

- Features tăng sức mạnh cho networks.
- Càng nhiều features, càng nhiều inputs.

Bất lợi:

- Có thể overfit data và dẫn tới poor generalization.
- Khó tìm tập feature nhỏ nhất để giải quyết bài toán.

Note: end-end neural networks không yêu cầu featureset

=> input => vector representation => Neural Networks => output

Playground

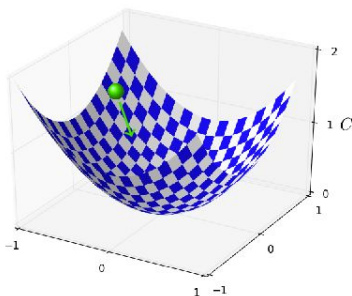
Tìm tập featureset nhỏ nhất để giải “Spiral” in a 6 x 6 x 6 ReLU

Learning rate

- learning rates cao làm cho networks train nhanh hơn.
- liệu learning rate càng cao càng tốt không?

Learning rate

- Mục đích chính là kiểm soát bước đi trong gradient descent.
- learning rate cao quá có thể dẫn tới oscillations (i.e. bước quá điểm cực trị).



Playground

Thử learning rates cho 6 x 6 x 2 with ReLU & report results.

Case:

- Learning rate = 0.00001
- Learning rate = 0.03
- Learning rate = 0.3
- Learning rate = 3
- Learning rate = 10

Activation function

- Không có một phương pháp cụ thể nào để chọn hàm activation
- Để ý tới neuron outputs để tìm ra activation phù hợp.
 - Chúng ta muốn trong signal hay softer signal?
- Math: NNs với ReLU (state of the arts) (với số layers đủ sâu) có thể xấp xỉ bất cứ hàm nào.

<https://arxiv.org/abs/1511.05678>

<https://papers.nips.cc/paper/5422-on-the-number-of-linear-regions-of-deep-neural-networks.pdf>

Optional

Regularization

- Add constraints to prevent overfitting

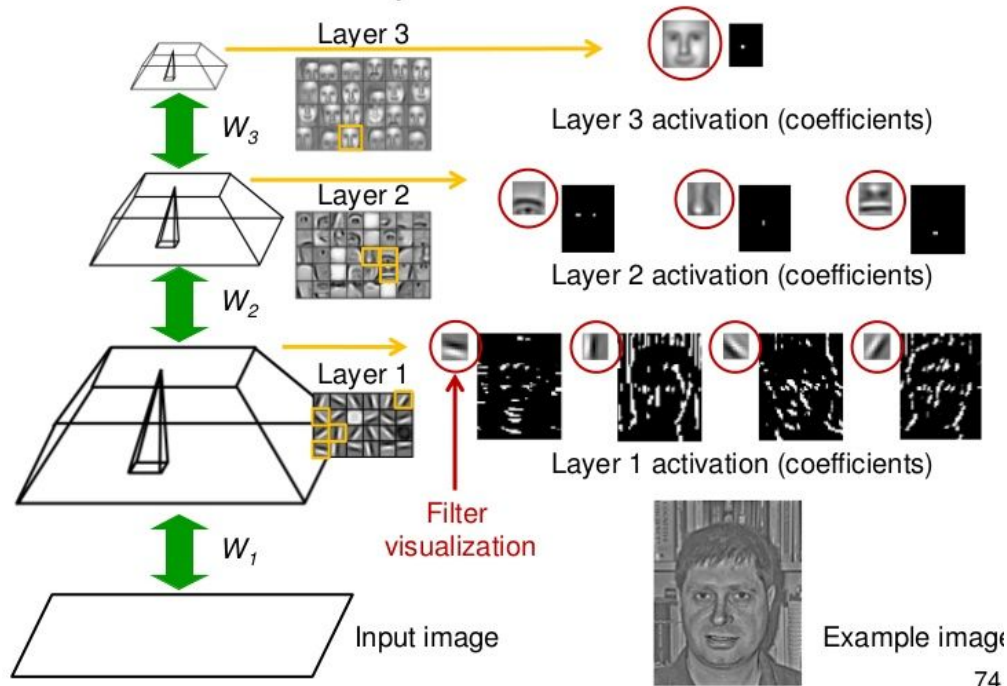
Why do we need end-end neural networks?

- Features need annotations / labels
- Annotations and labels are time consuming and expensive
- Basic label for outputs is needed for many problems because we have a lot of data

(ex: speech recognition is true end-end neural networks)

Layers of abstraction

Convolutional deep belief networks illustration



End-to-end NNs cannot solve all problems

- End-to-end NNs require a lot of data
- Ex:
 - Hand-tuned features NNs require 20k data points
 - End-to-end NNs require 200k data points
- Solution:
 - Hand-tuned features
 - Compositional networks
 - Each network can recognize / solve a sub problem
 - Each network can be end-to-end and need a small data points
 - $40k + 40k + 40k = 120k$ data points