3
Features

8000 Samples

X

Shuffle data

8000 Samples with shuffled order

In case there is an order in samples. Shuffling avoids creating batches with distinct set of data.

Create batches

10 batches each of size 800

Operation in one neuron

Value of variable X1 for one sample

x_1

w_1

pairwise connections

Value of variable X2 for one sample

x_2

$w2$

Artifical neuron in a hidden layer

$\sigma(w_1 x_1 + w_2 x_2 + b)$

σ can be any function, even linear. It is genrally differentiable.

$f(u) = \max(0, u)$

1

0

-1

u

A Rectified linear unit (ReLU) activation function. The value if either 0 or equal to to the input.

Input layer

Value of value for first variable (Size of vector = batch size)

$\bar{x}_1$

w_1

pairwise connections

Value of value for second variable (Size of vector = batch size)

$\bar{x}_2$

$w2$

Artifical neuron in a hidden layer

$\sigma(w_1 \bar{x}_1 + w_2 \bar{x}_2 + b)$

σ can be any function, even linear. It is genrally differentiable.

$f(u) = \max(0, u)$

1

0

-1

u

A Rectified linear unit (ReLU) activation function. The value if either 0 or equal to to the input.

Input layer

One step

800 samples

$\bar{x}_1 \ \bar{x}_2 \ \bar{x}_3$

One batch

$\bar{x}_1$

$\bar{x}_2$

$\bar{x}_3$

Input layer with one batch of size 800

Connections between each pair

First hidden layer

Connections between each pair

Second hidden layer

Connections between each pair

Third hidden layer

Connections between each pair

Fourth hidden layer

Connections between each pair

Fifth hidden layer

Connections between each pair

$\bar{x'}_1$

$\bar{x'}_2$

$\bar{x'}_3$

Output layer

$\bar{x'}_1 \ \bar{x'}_2 \ \bar{x'}_3$

800 samples

3 features

Reconstructed batch

One forward pass

Compute loss

$$\frac{(\bar{x}_1 - \bar{x'}_1)^2 + (\bar{x}_2 - \bar{x'}_2)^2 + (\bar{x}_3 - \bar{x'}_3)^2}{10}$$

Mean squared error for the batch. Can be any function (generally, differentiable)

Update network weights

Apply backpropogation with gradient descent and update weights (Frameworks to do automatic differentiation)

Backpropogation of loss

Encoder

Decoder

Bottleneck

(Also refered to as bottleneck code or simply, code)

800 samples

$\bar{x}_1 \ \bar{x}_2 \ \bar{x}_3$

One batch

$\bar{x}_1$

$\bar{x}_2$

$\bar{x}_3$

Input layer with one batch of size 800

Connections between each pair

First hidden layer

Connections between each pair

Second hidden layer

Connections between each pair

Third hidden layer

Connections between each pair

Fourth hidden layer

Connections between each pair

Fifth hidden layer

Connections between each pair

$\bar{x'}_1$

$\bar{x'}_2$

$\bar{x'}_3$

Output layer

$\bar{x'}_1 \ \bar{x'}_2 \ \bar{x'}_3$

800 samples

3 features

Reconstructed batch (Also refered to as autoencoded input)