

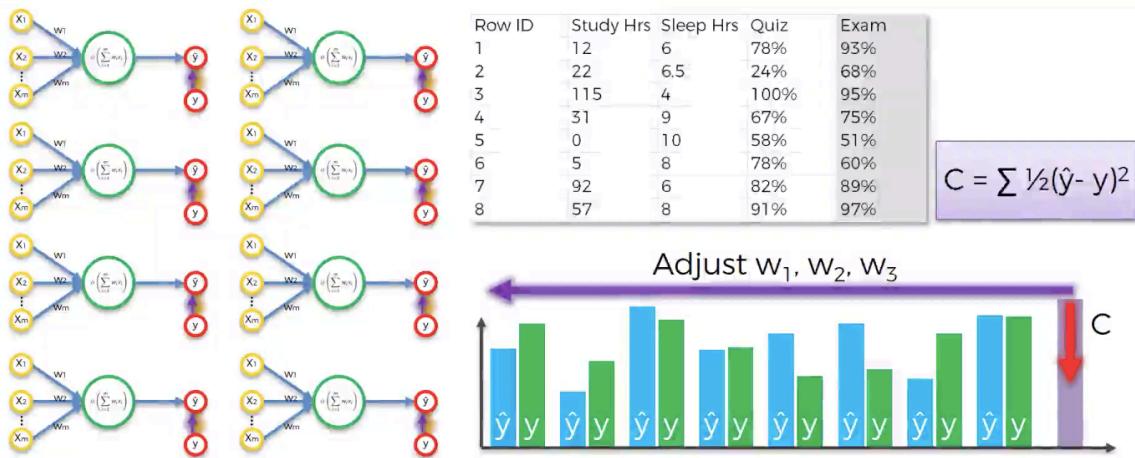
Gradient descent

Gradient Descent

Deep Learning A-Z

© SuperDataScience

Gradient Descent

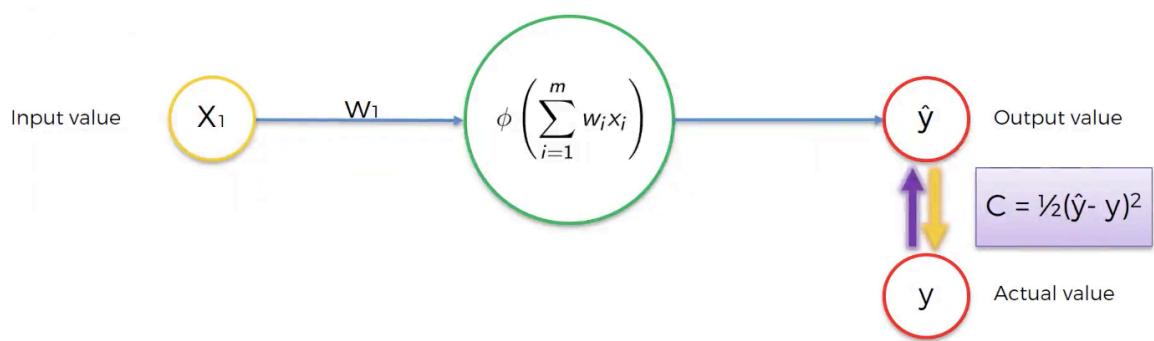


Deep Learning A-Z

© SuperDataScience

This is backpropagation method.

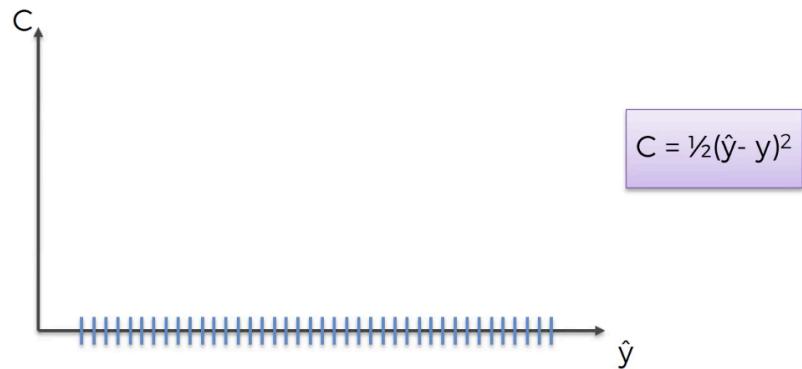
Gradient Descent



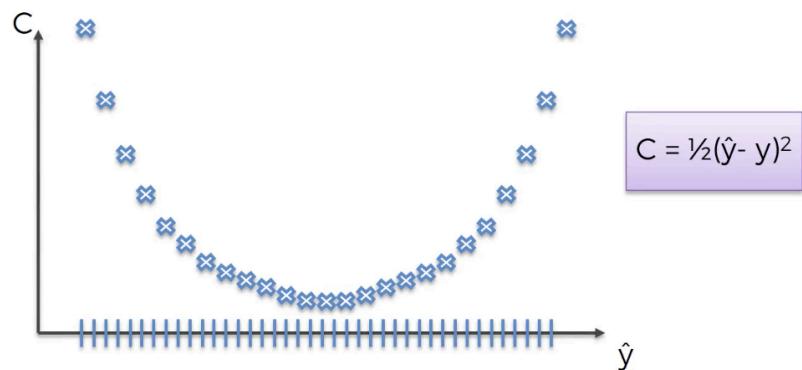
Gradient Descent



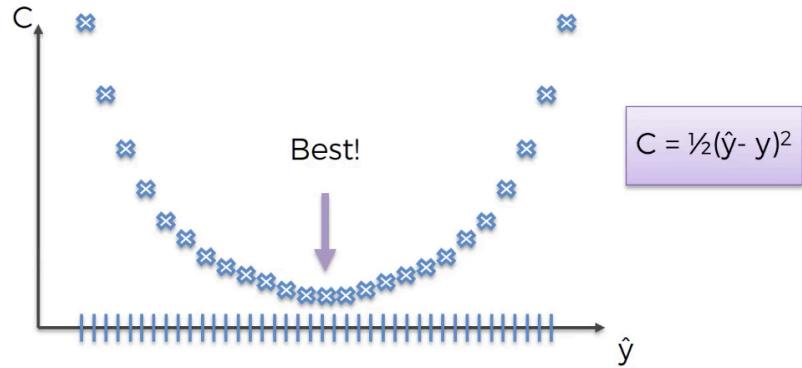
Gradient Descent



Gradient Descent



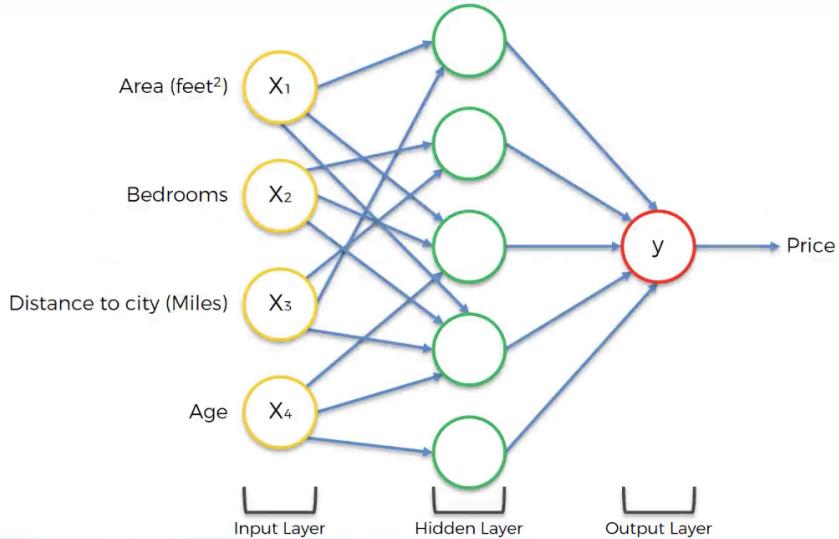
Gradient Descent



Gradient Descent

Curse of Dimensionality

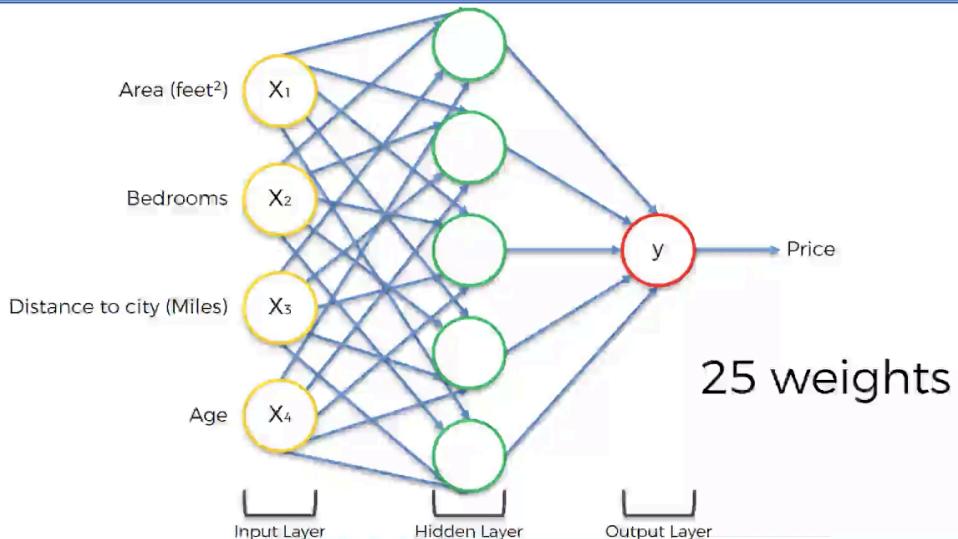
Gradient Descent



Deep Learning A-Z

© SuperDataScience

Gradient Descent



Deep Learning A-Z

© SuperDataScience

Gradient Descent

$1,000 \times 1,000 \times \dots \times 1,000 = 1,000^{25} = 10^{75}$ combinations

Deep Learning A-Z

© SuperDataScience



Gradient Descent

$1,000 \times 1,000 \times \dots \times 1,000 = 1,000^{25} = 10^{75}$ combinations

Sunway TaihuLight: World's fastest Super Computer

Deep Learning A-Z

© SuperDataScience



Gradient Descent

$1,000 \times 1,000 \times \dots \times 1,000 = 1,000^{25} = 10^{75}$ combinations

Sunway TaihuLight: World's fastest Super Computer



Deep Learning A-Z

© SuperDataScience

Gradient Descent

$1,000 \times 1,000 \times \dots \times 1,000 = 1,000^{25} = 10^{75}$ combinations

Sunway TaihuLight: World's fastest Super Computer



Deep Learning A-Z

© SuperDataScience

Gradient Descent

$1,000 \times 1,000 \times \dots \times 1,000 = 1,000^{25} = 10^{75}$ combinations

Sunway TaihuLight: World's fastest Super Computer

93 PFLOPS



Deep Learning A-Z

© SuperDataScience

A *petaflop* is the ability of a computer to do one quadrillion floating point operations per second (FLOPS).

Gradient Descent

$1,000 \times 1,000 \times \dots \times 1,000 = 1,000^{25} = 10^{75}$ combinations

Sunway TaihuLight: World's fastest Super Computer

93 PFLOPS

93×10^{15}



Deep Learning A-Z

© SuperDataScience

Gradient Descent

$1,000 \times 1,000 \times \dots \times 1,000 = 1,000^{25} = 10^{75}$ combinations

Sunway TaihuLight: World's fastest Super Computer

93 PFLOPS

93×10^{15}

$10^{75} / (93 \times 10^{15})$



Deep Learning A-Z

© SuperDataScience

Gradient Descent

$1,000 \times 1,000 \times \dots \times 1,000 = 1,000^{25} = 10^{75}$ combinations

Sunway TaihuLight: World's fastest Super Computer

93 PFLOPS

93×10^{15}

$10^{75} / (93 \times 10^{15})$

$= 1.08 \times 10^{58}$ seconds



Deep Learning A-Z

© SuperDataScience

Gradient Descent

$1,000 \times 1,000 \times \dots \times 1,000 = 1,000^{25} = 10^{75}$ combinations

Sunway TaihuLight: World's fastest Super Computer

93 PFLOPS

93×10^{15}

$10^{75} / (93 \times 10^{15})$

$= 1.08 \times 10^{58}$ seconds

$= 3.42 \times 10^{50}$ years



Deep Learning A-Z

© SuperDataScience

Gradient Descent

$1,000 \times 1,000 \times \dots \times 1,000 = 1,000^{25} = 10^{75}$ combinations

Sunway TaihuLight: World's fastest Super Computer

93 PFLOPS

93×10^{15}

$10^{75} / (93 \times 10^{15})$

$= 1.08 \times 10^{58}$ seconds

$= 3.42 \times 10^{50}$ years



Deep Learning A-Z

© SuperDataScience

As we can see, we cannot run these combinations even in the fastest super computer.

Gradient Descent

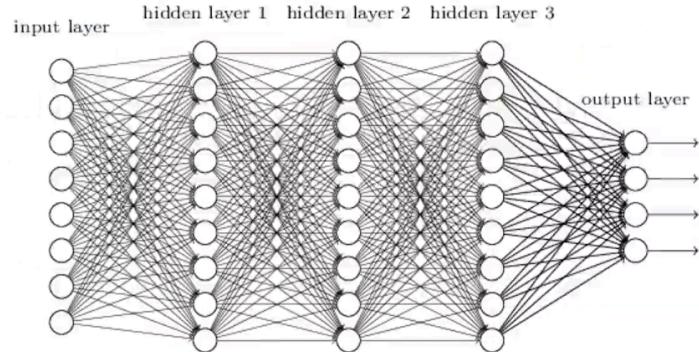


Image Source: neuralnetworksanddeeplearning.com

Deep Learning A-Z

© SuperDataScience

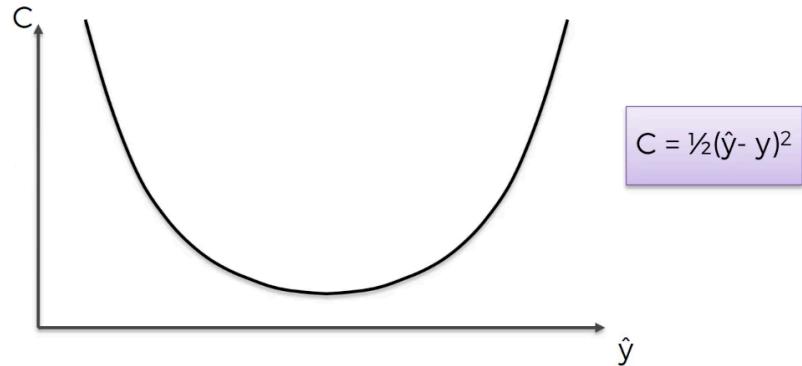
Gradient Descent

Gradient Descent

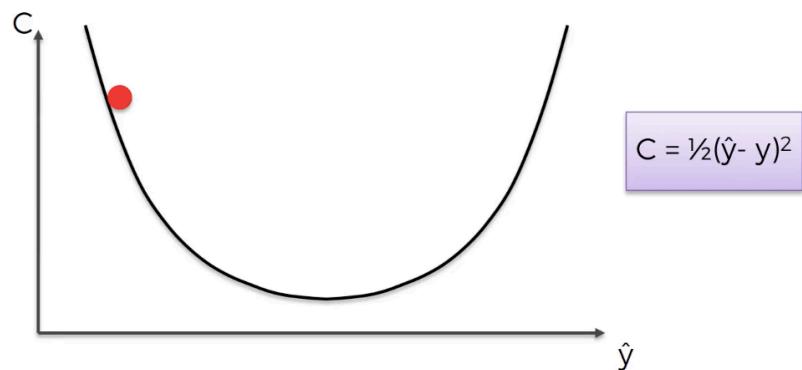
Deep Learning A-Z

© SuperDataScience

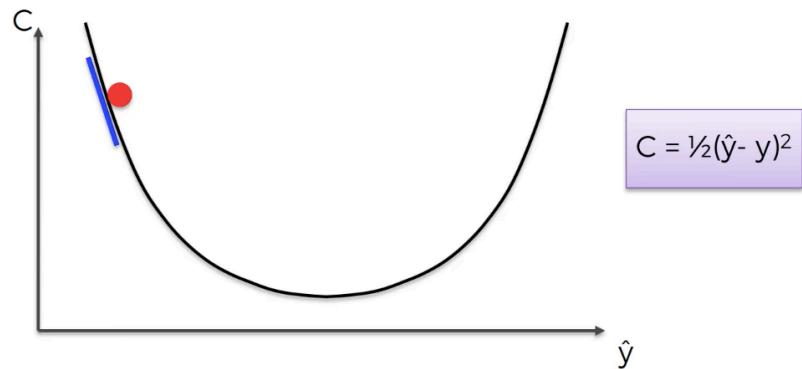
Gradient Descent



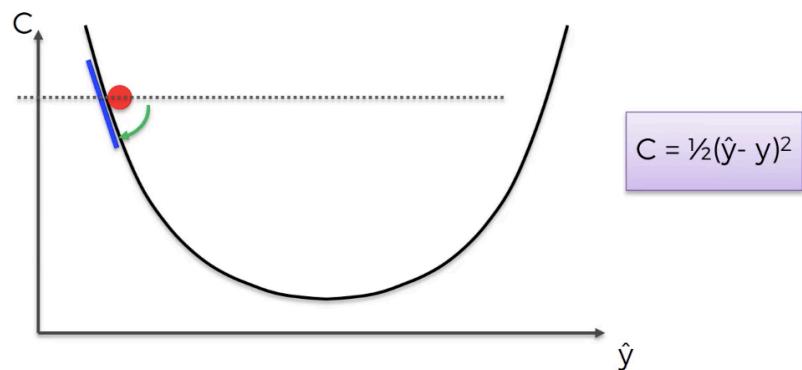
Gradient Descent



Gradient Descent

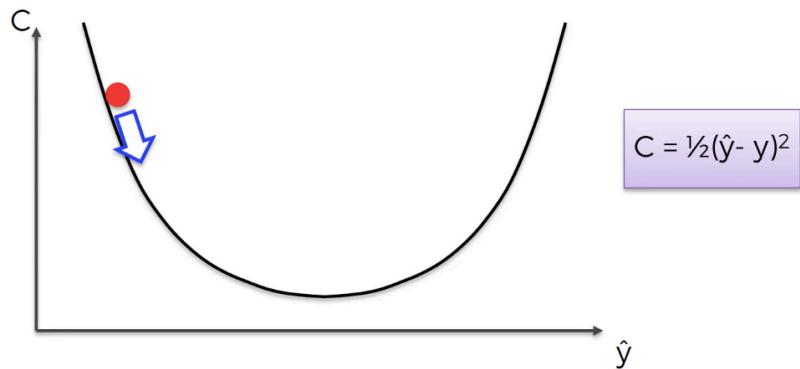


Gradient Descent



In here we get the slope, if it's negative then we go downward otherwise upward.

Gradient Descent

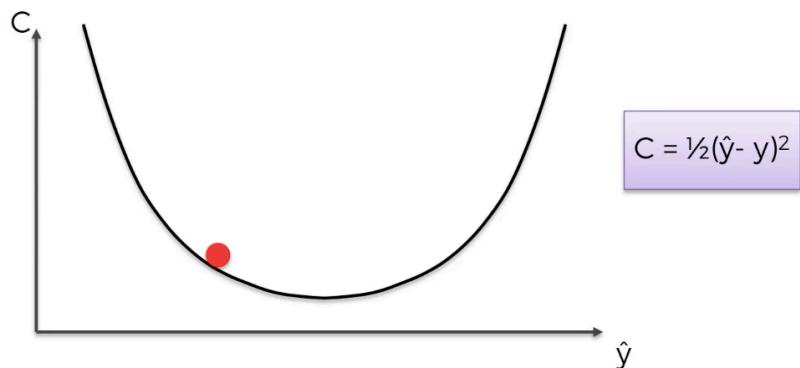


Deep Learning A-Z

© SuperDataScience

Imagine like a ball that is rolling. Of course in reality it's not rolling and it has a zig zag shape.

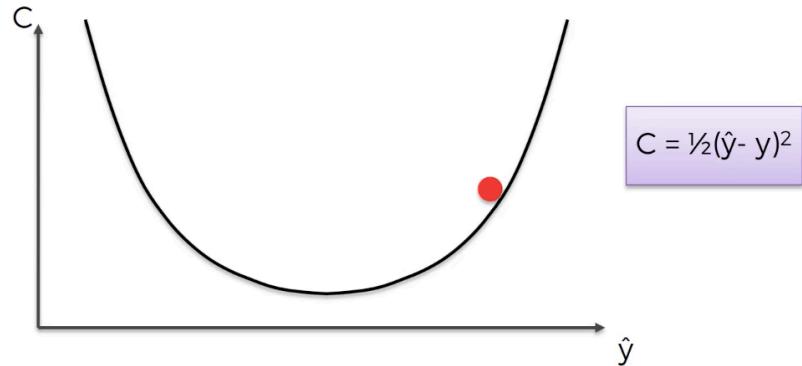
Gradient Descent



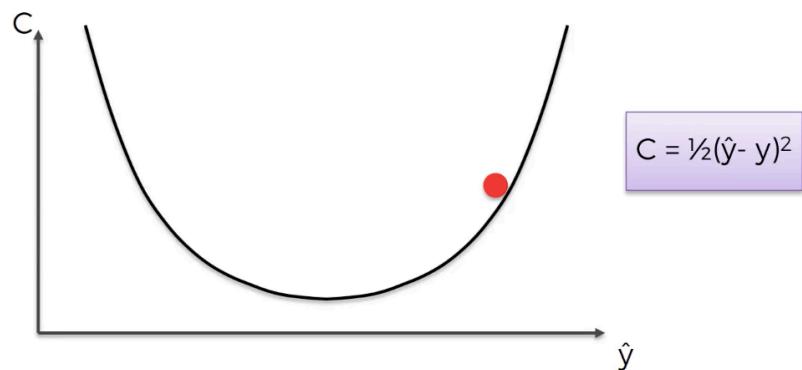
Deep Learning A-Z

© SuperDataScience

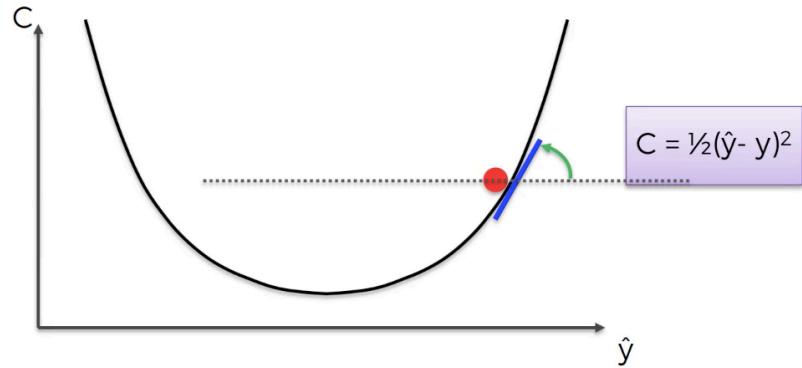
Gradient Descent



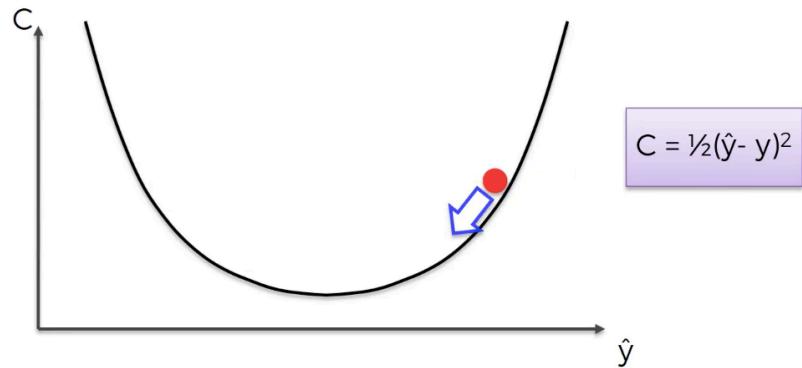
Gradient Descent



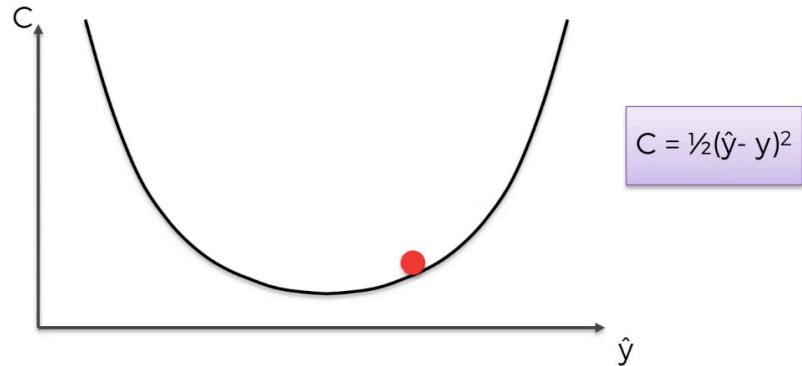
Gradient Descent



Gradient Descent



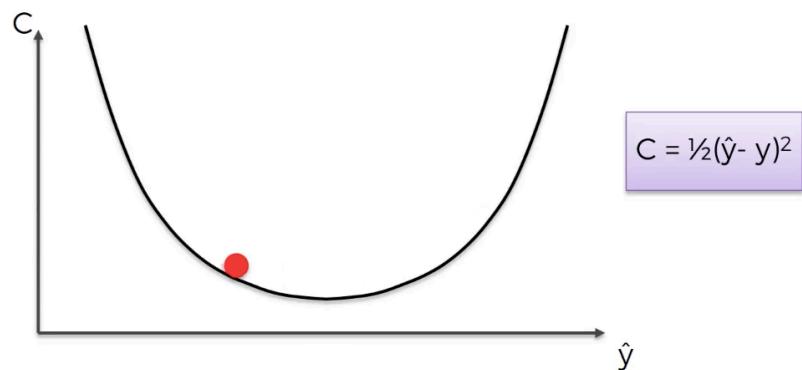
Gradient Descent



Deep Learning A-Z

© SuperDataScience

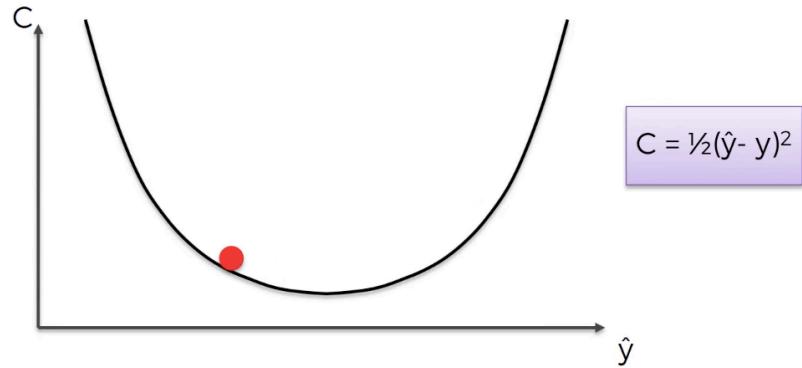
Gradient Descent



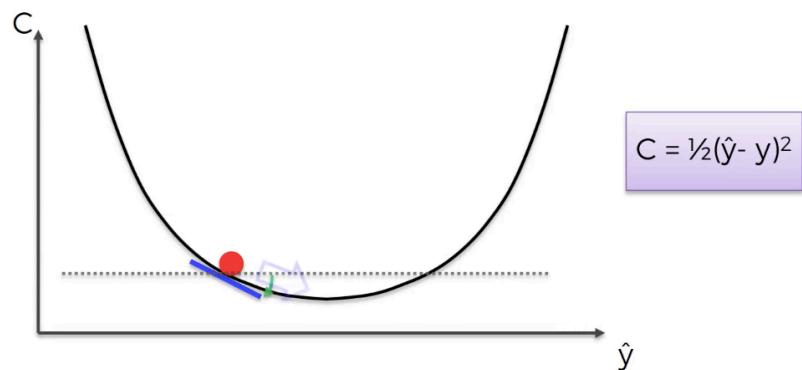
Deep Learning A-Z

© SuperDataScience

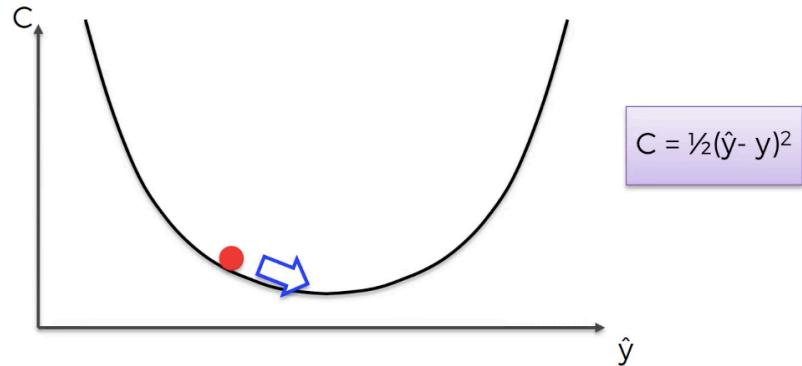
Gradient Descent



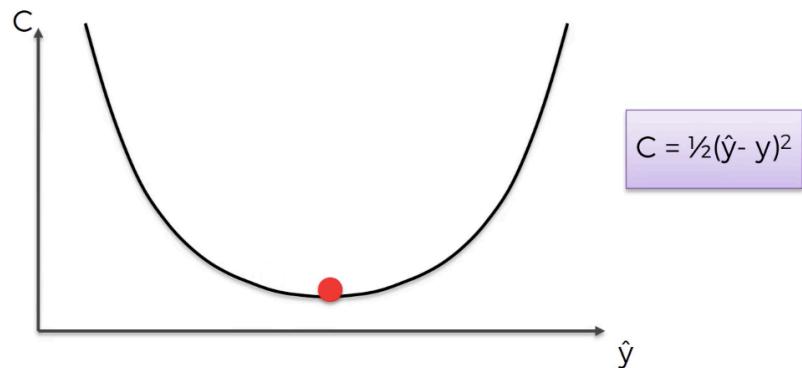
Gradient Descent



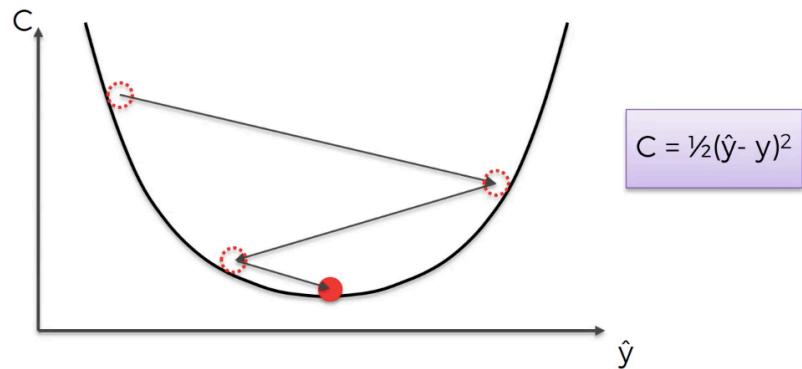
Gradient Descent



Gradient Descent



Gradient Descent

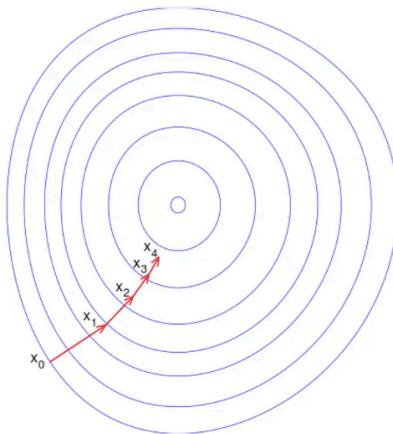


Deep Learning A-Z

© SuperDataScience

This is the real shape of movement which is zigzag. And as you can see, instead of going through all combinations we took so much less steps.

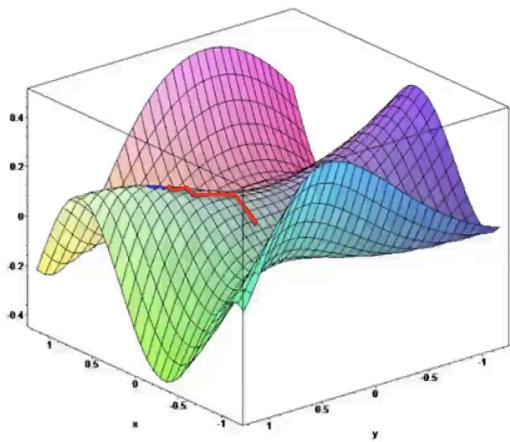
Gradient Descent



Deep Learning A-Z

© SuperDataScience

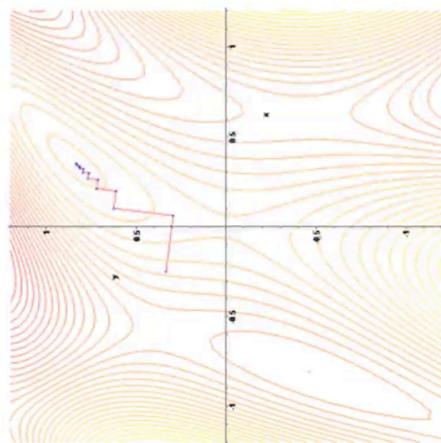
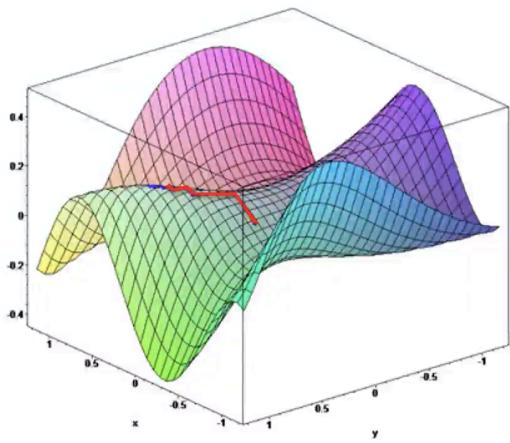
Gradient Descent



Deep Learning A-Z

© SuperDataScience

Gradient Descent



Deep Learning A-Z

© SuperDataScience