1. Gradient Descent

Gradient descent is a method to find minimum point of a function based on gradient of a function. Starting at any place on the function, calculate gradient and update new parameters

Example, . Initially, = -5 and = -2 ,

Epoch = 1:

0.2\*(-5) = -1

d = 4 = 4\*(-2) = -8

= = -5 -0.4\*(-1) = -4.6

= = -2 – 0.4\*(-8) = 1.2

Epoch = 2:

0.2\*(-4.6) = -0.92

d = 4 = 4\*(1.2) = 4.8

= = -5 – 0.4\*(-0.92) = -4.232

= = -2 – 0.4\*(4.8) = -0.72

2. Gradient Descent + Momentum

By combing with momentum, this method helps resolving the problem of local minimum point with the help of exponentially weighted average gradient.

Example, . Initially, = -5 and = -2 , ,

Epoch = 1:

0.2\*(-5) = -1

d = 4 = 4\*(-2) = -8

Epoch = 2:

0.2\*(-4.7) = -0.94

d = 4 = 4\*(0.4) = 1.6

3. RMSProp

Inspired by AdaGrad, use previous gradient to make learning rate becomes adaptive. RMSProp

Use exponentially weighted average squared gradient and then take a root of this value

Example, . Initially, = -5 and = -2 , ,

Epoch = 1:

0.2\*(-5) = -1

d = 4 = 4\*(-2) = -8

Epoch = 2:

0.2\*(-) = -0.8102

d = 4 = 4\*(-) = -4.204

4. Adam

This method is a combination between Momentum and RMSProp

Example, . Initially, = -5 and = -2 , , ,

Epoch = 1:

0.2\*(-5) = -1

d = 4 = 4\*(-2) = -8

Epoch = 2:

0.2\*(-4.8) = -0.96

d = 4 = 4\*(-1.8) = -7.2