

# Adam AdamW

## Adam

- Initialization
  - Parameters:  $\eta$  (learning rate),  $\beta_1$  (momentum),  $\beta_2$  (variance),  $\epsilon$  (epsilon)
  - Initial values:  $m_0 = 0$ ,  $v_0 = 0$
  - Default values:  $\eta = 0.001$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ ,  $\epsilon = 1e-8$
- Update parameters
  - Input:  $\theta_t$  (current parameters),  $g_t$  (current gradients)
- Update moment
  - Formula:  $m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$
- Update variance
  - Formula:  $v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$
- Bias Correction
  - Formula:  $\hat{m}_t = m_t / (1 - \beta_1^t)$ ,  $\hat{v}_t = v_t / (1 - \beta_2^t)$
- Update parameters
  - Formula:  $\theta_{t+1} = \theta_t - \eta \cdot \hat{m}_t / \sqrt{\hat{v}_t + \epsilon}$

```
class AdamOptimizer:
    def __init__(self, learning_rate=0.01, beta1=0.9, beta2=0.999, epsilon=1e-8):
        self.lr = learning_rate
        self.beta1 = beta1
        self.beta2 = beta2
        self.epsilon = epsilon
        self.m = 0
        self.v = 0
        self.t = 0

    def update(self, params, grads):
        self.t += 1
        for param, grad in zip(params, grads):
            self.m[param] = self.beta1 * self.m[param] + (1 - self.beta1) * grad
            self.v[param] = self.beta2 * self.v[param] + (1 - self.beta2) * grad**2
            m_hat = self.m[param] / (1 - self.beta1**self.t)
```

```

v_hat = self.v[param] / (1 - self.beta2**self.t)

param = param - self.lr * m_hat / (np.sqrt(v_hat) + self.epsilon)

return params

```




## weight decay



```
w = w - lr * w.grad - lr * weight_decay * w
```





L2

```
final_loss = loss + wd * all_weights.pow(2).sum() / 2
```

**Adam**    L2 L2  


```
grad_w = grad_w + weight_decay * w
```

## AdamW

AdamW     L2   
 **grad\_w**

```

grad_w = grad_w

# update step

w = w - learning_rate * grad_w - learning_rate * weight_decay * w

```

Revision #3

Created 12 March 2025 08:09:34 by Colin

Updated 12 March 2025 09:33:51 by Colin