

---

**Algorithm 1** : Adam Optimization

---

**Require:**  $\alpha$  (learning rate),

$\beta_1$  (first moment decay rate),

$\beta_2$  (second raw moment decay rate),

$\mathbf{w}_0$  (initial guess for weights vector  $\mathbf{w}$ )

**Ensure:** Updates the current weights vector  $\mathbf{w}$  for an optimization step

- 1:  $m_0 \leftarrow 0$  (1st moment vector)
  - 2:  $v_0 \leftarrow 0$  (2nd raw moment vector)
  - 3:  $t \leftarrow 0$  (initialize timestep)
  - 4: **while**  $\mathbf{w}$  not converged **do**
  - 5:    $g_t \leftarrow \nabla L(\mathbf{w}_t)$  {Get current loss function gradient}
  - 6:    $m_t \leftarrow \beta_1 \cdot m_t + (1 - \beta_1) \cdot g_t$  {Update 1st moment vector}
  - 7:    $v_t \leftarrow \beta_2 \cdot v_t + (1 - \beta_2) \cdot g_t^2$  {Update 2nd raw moment vector}
  - 8:    $\hat{m}_t \leftarrow \frac{m_t}{(1 - \beta_1^t)}$  {Compute bias-corrected 1st moment est.}
  - 9:    $\hat{v}_t \leftarrow \frac{v_t}{(1 - \beta_2^t)}$  {Compute bias-corrected 2nd raw moment est.}
  - 10:    $\mathbf{w}_t \leftarrow \mathbf{w}_t - \alpha \cdot \frac{\hat{m}_t}{(\sqrt{\hat{v}_t} + \epsilon)}$  {Update current  $\mathbf{w}$ }
  - 11: **end while**
  - 12: **return**  $\mathbf{w}$  {Optimized weights vector}
-