We wish to minimise (our definition of SSE for mathematical convenience)

For finding gradient of towards SSE, differentiate SSE with respect to **a**

…. replacing

…. assuming *X* as constant

For finding gradient of towards SSE, differentiate SSE with respect to **b**

…. replacing

…. assuming *X* as constant

Update in **a**

Update in **b**

**Momentum (beta)**

**Updating a**

**Updating b**