

---

## Algorithm 1 Poisoning Example Generation

---

**Input:** target instance  $t$ , base instance  $b$ , learning rate  $\lambda$

Initialize  $\mathbf{x}$ :  $x_0 \leftarrow b$

Define:  $L_p(x) = \|f(\mathbf{x}) - f(\mathbf{t})\|^2$

**for**  $i = 1$  **to**  $maxIters$  **do**

    Forward step:  $\hat{x}_i = x_{i-1} - \lambda \nabla_x L_p(x_{i-1})$

    Backward step:  $x_i = (\hat{x}_i + \lambda \beta b) / (1 + \beta \lambda)$

**end for**

---