
Algorithm 1: Train a model to generate NFA

Require: Learning rate: $\alpha > 0$.

Require: Loss function $L(f(\mathbf{x}; \boldsymbol{\theta}), \mathbf{y}) = \|f(\mathbf{x}; \boldsymbol{\theta}) - \mathbf{y}\|$, the combination of norms.

Initial parameters $\boldsymbol{\theta}$;

while *the loss function L stop decrease* **do**

 Sample a minibatch of N examples from the training set $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}$ with corresponding targets $\mathbf{y}^{(i)}$, such as $\mathbf{x}^{(i)} = (101010, T)$, $\mathbf{y}^{(i)} = (Q, \Sigma, \delta, q, F)$;

 Apply update:

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \alpha \frac{1}{N} \sum_{i=1}^N \nabla_{\boldsymbol{\theta}} L(f(\mathbf{x}^{(i)}; \boldsymbol{\theta}), \mathbf{y}^{(i)});$$