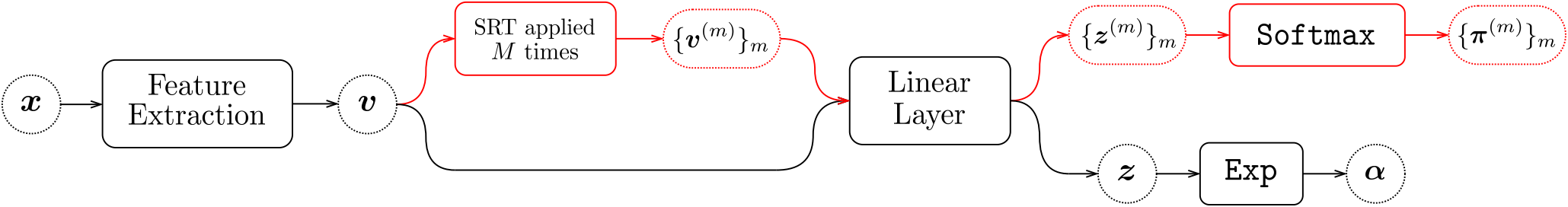
$\boldsymbol{x}$
Feature Extraction
$\boldsymbol{v}$
SRT applied M times
$\{\boldsymbol{v}^{(m)}\}_m$
Linear Layer
$\{\boldsymbol{z}^{(m)}\}_m$
Softmax
$\{\boldsymbol{\pi}^{(m)}\}_m$
$\boldsymbol{z}$
Exp
$\boldsymbol{\alpha}$