

ABSTRACT

A method for training a machine learning model is described, comprising receiving, for each perturbation of a plurality of perturbations of model parameters of a starting version of the machine learning model, a change of loss of the machine learning model caused by the perturbation for a set of training data determined by feeding the set of training data to one or more perturbed versions of the machine learning model, estimating a gradient of the loss of the machine learning model with respect to the model parameters from the determined changes of loss and updating the starting version of the machine learning model to an updated version of the machine learning model by changing the model parameters in a direction for which the estimated gradient indicates a reduction of loss.

FIG. 4