Abstract

In this thesis we introduce the Recurrent Neural Network (RNN) model starting from the more common Feed Forward Neural Networks showing how to use such models to define optimization problems for machine learning. We pay particular attention to the gradient structure of such models because it helps explaining the "infamous" *exploding/vanishing* gradient problem which constitutes the most important difficulty in training RNNs. We show how this problem affects the learning process and we review the solutions that have been developed over time. We then introduce a novel approach based on stochastic gradient descent which deals with the *exploding/vanishing* gradient problem through a careful initialization and by using a descent direction different from the usual anti-gradient. Finally we evaluate our approach both on synthetic and real datasets.