Tangent Prop – A formalism for specifying selected invariances in an adaptive network

Patrice Simard  
AT&T Bell Laboratories  
101 Crawford Corner Rd  
Holmdel, NJ 07733

Bernard Victorri  
Université de Caen  
Caen 14032 Cedex  
France

Yann Le Cun  
AT&T Bell Laboratories  
101 Crawford Corner Rd  
Holmdel, NJ 07733

John Denker  
AT&T Bell Laboratories  
101 Crawford Corner Rd  
Holmdel, NJ 07733

Abstract

In many machine learning applications, one has access, not only to training data, but also to some high-level a priori knowledge about the desired behavior of the system. For example, it is known in advance that the output of a character recognizer should be invariant with respect to small spatial distortions of the input images (translations, rotations, scale changes, etcetera).

We have implemented a scheme that allows a network to learn the derivative of its outputs with respect to distortion operators of our choosing. This not only reduces the learning time and the amount of training data, but also provides a powerful language for specifying what generalizations we wish the network to perform.

1 INTRODUCTION

In machine learning, one very often knows more about the function to be learned than just the training data. An interesting case is when certain directional derivatives of the desired function are known at certain points. For example, an image