Senior Thesis

Real Analysis
Dec 12 2016

William Cai[1]

1 Abstract

2 Introduction

2.1 Data setup

X is a n x T x d matrix where n is the number of patients, T is the number of measurements, and d is the dimensionality of the measurements.
Currently, in the synthetic data we have d x d matrices A and B which have 2-norm .9 and the data for a single patient is generated as follows:
P, the data for a single patient, is a T x d matrix. Let P_i be the measurement of that patient at time i.
P_0 is randomly initialized with all values mean 0 and variance 1 and normalized by $\sqrt{d}$.
Then, $P_{t+1} = (1_A A + (1 - 1_A)B)P_t + \sigma_w \epsilon$ where
σ_w is a constant (currently .2)
ϵ is d-dimensional white noise with all values mean 0 variance 1
1_A is an indicator variable which is 1 if there is no $i \leq t$ s.t. $P_i \cdot betaswitch > .5$

2.2 What are we trying to predict

There are two prediction problems we are interested in.
1) Given P_0 through P_t, predict P_{t+1} (regression).
2) Given P_0 through P_t, predict 1_A (binary classification)

3 Models (ignore this for now)

3.1 Predict next measurement from previous measurements

Currently only using a window of 20 measurements:
Try 1: Random Forest Regression (30 trees, 30 features) (DONE) Try 2: Neural Network (1 layer, 512 nodes) (DONE) Try 3: Deep Neural Network (2 layers, 512 nodes) (DONE) Try

[1] William Cai

4: Recurrent Neural Network, fed 20 at a time (DONE) Try 5: Recurrent Neural Network, fed 1 at a time
TODO: Vary the length of our window (not just 20)

3.2 Predict next value of flag (0/1) directly

Try 1: Random Forest Classifier (30 trees, 30 features) Try 2: Neural Network (1 layer, 512 nodes) Try 3: Deep Neural Network (2 layers, 512 nodes) Try 4: Recurrent Neural Network, fed 20 at a time Try 5: Recurrent Neural Network, fed 1 at a time

3.3 Predict next value of flag (0/1) from prediction of the model which predicts the next measurement

Question: Should we even do this, since all our predictions just look like they're guessing 0? Try 1: Random forest Classifier Try 2: Neural Network Try 3: Deep Neural Network

4 Results

5 Discussion