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 Experiment

4.1 Regression where $A = B$

4.1.1 Introduction

Given a sequence of measurements from time 0 through t, our goal is to predict the measurement at time t + 1.
As a sanity check, we check to see our models are working when A = B, or in other words, the patient measurements are generated by:

$$P_{t+1} = AP_t + \sigma_w \epsilon$$

Then, in this case a perfect model would predict $P_{t+1} = AP_t$.
All of these models were trained using the window method, where each training example is a sequence of 10 measurements and the result is the next measurement. The dataset was split 80-20 between training/validation.

4.1.2 Results

Models: All Zero Regressor: 0.168840895246 Oracle Regressor (knows the value of A): 0.159880463551 Random Forest Regressor 0.164463405344 Neural Net, 512 nodes 0.161696706556 Deep Neural Net, 2 layers of 512 nodes 0.160718555103 Recurrent Neural Net 0.160057111654

4.2 Regression where $A \neq B$

4.2.1 Introduction

Given a sequence of measurements from time 0 to 1, our goal is to predict the measurement at time t + 1. Now that $A \neq B$, our data is generated by

$$P_{t+1} = (1_A A + (1 - 1_A)B)P_t + \sigma_w \epsilon$$

which means our models have to deal with the fact that sometimes the data is being generated with A and sometimes with B.
All of these models were trained using the window method, where each training example is a sequence of 10 measurements and the result is the next measurement. The dataset was split 80-20 between training/validation.

4.2.2 Results

Models:
Zero predictor (just guess all-zeros): MSE: 0.168869027809 Random Forest MSE: 0.17058158425 Neural Net, 512 nodes MSE: 0.161847476813 Deep Neural Net, 2 layers of 512 nodes MSE: 0.160740647684 Recurrent Neural Net MSE: 0.161418953358
The error of an oracle predictor which knows the value of A, B and flag should be similar to the error from the previous section but wasn't implemented because it's hard to figure out the value of flag for a given training example.

4.3 Classification where $A \neq B$

4.3.1 Introduction

Given a sequence of measurements from time 0 through t, our goal is to predict the value of flag at time t + 1.
All of these models were trained using the window method, where each training example is a sequence of 10 measurements and the result is the next value of flag. The dataset was split 80-20 between training/validation.

4.3.2 Results

Random Forest
[[5104 28]
[1271 197]]
Neural Net, 512 nodes
[[5009 123]
[196 1272]]
Deep Neural Net, 2 layers of 512 nodes
[[5011 121]
[212 1256]]
Recurrent Neural Net
[[4965 167]
[148 1320]]

5 Discussion