CS 229 Machine Learning
Project Final Report SUNet ID: josilos
Jose Ramon Silos Stanford ID #:06037574

Project: Local Degree of Freedom Ductility Demand

Project Category: Physical Sciences / Simulation

Abstract

Ductility demand during an earthquake has long been known to be a good measure of structural damage in buildings during earthquakes. For design purposes it is considered a priori according to some rules that may or may not apply to each building. Such ductility parameters are mostly based on the response of a single degree of freedom system to ground motions. However, it has been probed that these parameters are not accurate enough and might lead to underestimating risks to which a structure may be vulnerable to (e.g. Miranda and Bertero 1994, Esteva 1994, Santa-Ana and Miranda 2000). Application of probabilistic methods is recommended to evaluate vulnerability of structures, (ASCE 07-2010). And so, traditional probabilistic approaches (i.e. total probability theorem) are applied as part of the Performance Based Earthquake Engineering (PBEE) methodology. This project intends to evaluate the accuracy of Machine Learning algorithms when applied as an alternative to traditional methods. Linear Regression, Weighted linear regression and Neural Networks are evaluated by comparing deterministic results of nonlinear models with predictions resulting from these algorithms. Both Linear and Weighted linear regression resulted accurate enough but do not seem to capture the essence of the problem. In opposition, neural networks appear to capture the essence of the problem and further work may be promising.

Introduction

Earthquake resistant design is one of the most important issues to guarantee the correct behavior and serviceability of buildings and so, it is one of the major concerns for current Structural Engineering research and practice. Economy and life safety are the main goals for design, allowing inelastic behavior of materials during extreme earthquakes (i.e. design level earthquakes and maximum considered earthquake) is common for achieving these goals. However, if not designed properly, such type of behavior may lead to unstable behavior that may produce collapse. Such inelastic behavior during design process is measured (or assumed) by predefined ductility demand values pre-established in codes. Ductility is defined as the quotient of the maximum deformation of the structure during the earthquake divided by the limit of its elastic deformation.

In current codes the parameter ductility is based on single degree of freedom behavior. (e.g. single-story buildings). Commonly the reference point for measuring and understanding ductility is the top story of the building under design. However, this suggests a global measure of ductility and important information might be unnoticed by the designer. Most of the times, measuring ductility locally leads to a better understanding of the level of damage that a structure may present during an earthquake. (e.g. ductility at the first story, ductility at a certain beam or brace.)

Supervised machine learning techniques are evaluated as an alternative for predicting the maximum local ductility demands. The input for the model is formed by three types of scalar variables: 1) Ground Motion Characteristics, 2) Structure properties, and 3) Variables which are a combination of both. In PBEE, the first and third kind are both grouped in the name of intensity measures. However, the writer considered important to make the distinction for this work. The result of each model is a positive real value number which defines the ductility demand at a local story.

Even though probabilistic approaches are commonly used in performance based earthquake engineering for evaluating the seismic performance of different components of buildings, there is not a work known to the writer in which neural networks are used to predict the ductility demand or damage of a building for earthquake events. As part of this work, Linear Regression and Weighted Linear Regression are also applied with two purposes: 1) to improving the understanding of the behavior of machine learning techniques when applied to a problem which is more familiar to the author, and 2) to serve as benchmarks to know if a more complicated model is of use for this problem or if it would only result in "overcomplicating" the problem. The observation was that neural networks are indeed more effective for addressing this kind of problems.

Datasets and Features

As described above, three groups of features are considered, the ten following features were considered for the model:

a) Ground Motion Characteristics:

1. Peak Ground Acceleration (PGA): is the maximum acceleration from the ground motion record.
2. Peak Ground Velocity (PGV): is the maximum velocity computed when integrating the ground motion acceleration record over time.
3. Peak Ground Displacement (PGD): is the maximum displacement computed when integrating the ground motion acceleration record over time two times.
4. Strong GM Duration (90% Arias intensity): refers to the time-length in which the middle 90% of energy of the record is contained (i.e. leaving out 5% from the beginning of the record and 5% from the end). In this energy measure, the energy is computed as the integral of the squared accelerations of the record over time.

b) Building Properties:

5. Number of stories of the building
6. Fundamental period of vibration of the building (T1): is the period of vibration associated to the first mode of vibration of the building.
7. Design Ductility (μ): Is the expected ductility of the structure considering building design factors, it can be computed as the quotient of the seismic force reduction factor (R) and the overstrength factor (Ω) from the building code (ASCE 07-2010).

c) Composite Properties: result of the combination of properties from both the building and the ground motion.

8. Spectral Acceleration at T1 (SaT1): the maximum acceleration response to the considered ground motion of an elastic single degree of freedom system having the same period as the fundamental period of the building.
9. Geometric Mean of Spectral accelerations at T1 (SaGmean): Geometric Mean of maximum response accelerations of single degree of freedom systems with periods in the vicinity of T1 (i.e. from 0.2T1 to 3.0 T1 spaced at 0.1 T1).
10. Average Spectral acceleration at T1 (SaAve): Mean of maximum response accelerations of single degree of freedom systems with periods in the vicinity of T1 (i.e. from 0.2T1 to 3.0 T1 spaced at 0.1 T1).

Since there were not published databases of such analysis containing all these variables, simplified models' analyses were conducted to generate the applied dataset. The variations to generate the database were as follows:

Ground Motions	Number of Stories	μ	Number of periods (T1) for each number of stories	Total analyzed samples
270	1-7	1,2,3,4	5	37800

A post-process was required to evaluate the stability of the simplified model for resulting in representative values. As a result, 27 ground motions were removed and a resulting database of 34,020 samples was used to perform this project. The ground motions were considered as representative according to Eads (2013).

Methods

Three machine learning methods were applied for this project to improve the understanding of such methods, and to develop sensibility of how variations within these Machine Learning methods may affect the outcome when applied to this type of problems.

1. Linear Regression

This is the simplest method in Machine Learning, its goal is to find a vector, the size of the number of features and predict the outcome of the problem as the result of a dot product operation of a vector containing the features' values and the linear regression vector. The regression vector can be obtained with the following closed form formula:

$$\theta := (X^T X)^{-1} X^T \vec{y}$$

Where θ is the linear regression vector, X is the matrix containing the feature values of the training samples, and $\vec{y}$ is the vector containing the result or outcome of each training sample.

2. Weighted Linear Regression

Just a step further of linear regression, weighted linear regression intends to obtain the regression vector by making the samples closer to the intended prediction to have more influence in the outcome. This is achieved by adding weights to each sample of the training set. In this project, the weights were computed as follows:

$$w_{ii} = \exp(-(x^{(i)} - x)^T {\Sigma_x}^{-1} (x^{(i)} - x)/2)$$

Where w_{ii} is the weight value for the training sample i, $x^{(i)}$ is the vector containing the features' values of the i[th] training sample, x is the vector containing the features' values of the intended prediction, and Σ_x is the covariance matrix for the features of the training data set.

The regression vector is obtained with the following closed form formula:

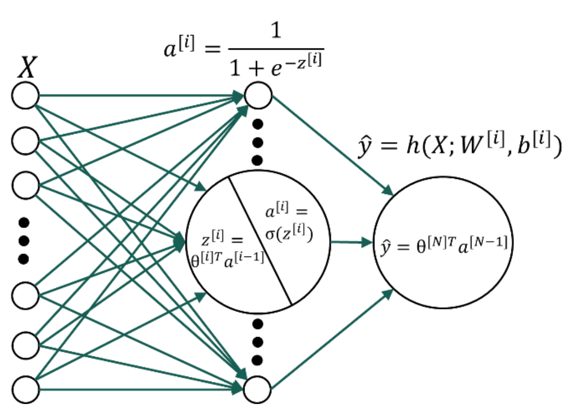
$$\theta := (X^T W X)^{-1} X^T W \vec{y}$$

Where θ, X, and $\vec{y}$ are as defined for linear regression and W is a diagonal matrix containing the weights w_{ii}.

For both, linear and weighted linear regression, the prediction or hypothesis is obtained with the following equation:

$$h := \theta^T x$$

3. Neural Network



Neural networks are a more robust method in Machine learning. They are formed by layers in which both, linear operators and nonlinear activation functions are applied. Each layer is formed by a defined number of units, in each unit a linear dot product operator is applied to the vector of values resulting from the previous layer and a nonlinear activation function is applied to the result of that dot product. The prediction is the result of the last layer activation function, this activation function may be linear or nonlinear. For the present project, a 10-units single hidden layer network having the structure from the figure above.

To obtain the weights and the bias terms, gradient descent is applied to minimize the loss function because there is no close form solution. Both no regularization and L2 regularization approaches were used. 50 epochs of 1000-size batches batch gradient descent was used. Two different ways to define the learning rates were used, one was a constant value of five, the second of a constantly descending learning rate for each epoch from 5.0 to 0.1.

For the three previous methods, the following loss function was considered:

$$L = \frac{1}{m}\sum\left(\frac{(\hat{y} - y)}{y}\right)^2$$

Where *m* is the number of training samples, $\hat{y}$ is the prediction resulting of each model, and y is the actual value of the training sample. This loss function is normalized by the actual local degree of freedom ductility demand to maintain a physical meaning which is easily related to the problem (i.e. how big is the expected error with respect to the expected result).

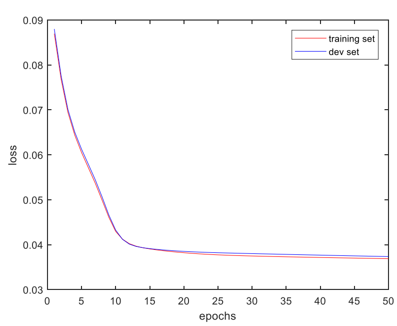
Discussion

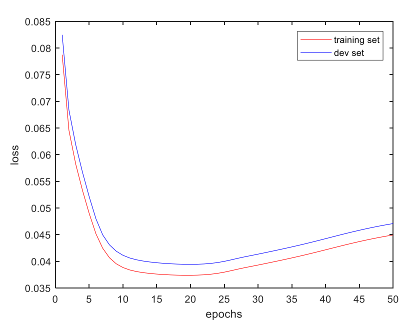
For practicality, a 30,000 – 4,020 hold out training was preferred to be able to easily compare the outcomes of the three methods. The immediate comparison that can be made of the three methods is their accuracy. The resulting losses of each method are summarized in the following table:

Model	L(θ)	(L(θ))^0.5
Linear Regression	0.30	0.55
Weighted Linear Regression	0.10	0.33
Neural Network	0.06	0.24
L2 Regularized Neural Network	0.035	0.18

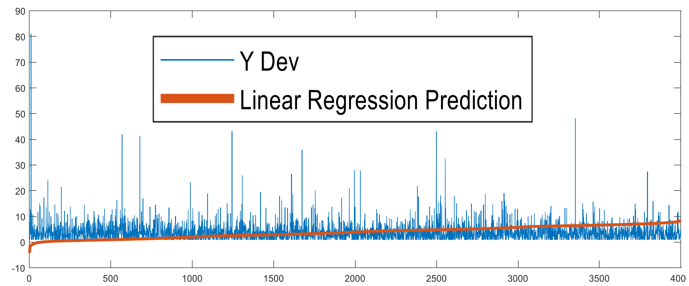
Since the Loss function contains the sum of the squared errors, the third column represents the actual physical interpretation of the loss function, which is the error normalized by the expected outcome. For all four cases, this error is acceptable when compared to the accuracy of current Structural Engineering practice methods for computing the actual ductility demand in a local manner. However, it is important to remark that the gain in accuracy with weighted linear regression may not be sufficient to justify how expensive this process becomes. And that Neural Networks were found to be highly accurate and efficient for computation. The value of forward and backward propagation was clearly noticed in the computation time of this project.

The following figures represent the loss with constant learning rates without regularization (left) and with L2 regularization (right). The need of annealing was noticed when observing this plots in which the effect of regularization being negative after certain number of epochs was observed.

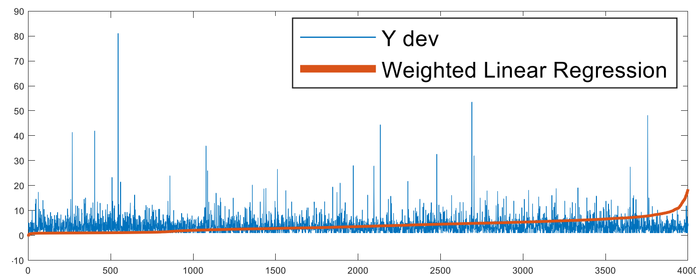




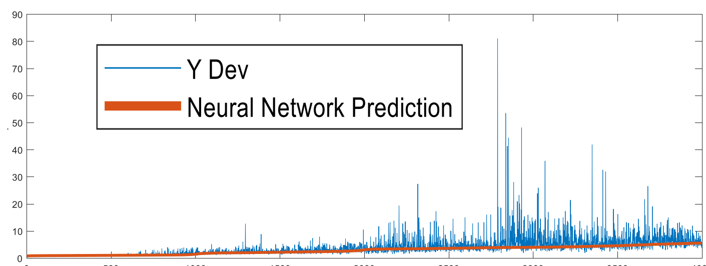
To further investigate the outcome and precision of each method, the hypothesis for the dev set were ordered in an ascending manner and compared to the expected values. This process is plotted in the following three figures. A major finding is that even though the accuracies may be acceptable for LR and WLR, they don't seem to correctly capture the real model. Their plots result in the superposition of an ascending line with random values. However, when looking at the Neural Network predictions, the random values seem to align to the ascending prediction and their dispersion appears to be growing in the same manner. It is observed that for all cases, extreme values are not easily predicted and represent most of the loss.



a) Liner Regression Prediction and actual Local Ductility Demand



b) Weighted Liner Regression Prediction and actual Local Ductility Demand



c) 10-unit, single-hidden-layer neural network, 50-epochs, 30-batches training, varying learning rate from 5.0 to 0.1, L2 regularization

Conclusion and Future Work

The use of neural networks in Performance Based Earthquake Engineering may result in a powerful addition to the current methodology. Its ability to capture this kind of highly nonlinear problems might result very useful. However, it is important to push this work further to improve the outcomes since some downsides are observed. The most noticeable is the fact that the predictions approximate the damage measure (i.e. local degree of freedom ductility demand) from below. In plot c), most of the actual values result above the prediction. Since the most valuable of Earthquake Engineering is to protect people's lives, it would be desired to predict a higher damage than the real one when estimating losses to stay in the conservative side.

Including more layers, trying ReLu activation function and including more complex input features such as Maximum Incremental Velocity, Sa Ratios and Shape Factors are suggested as the following steps to this work. Also, extrapolating this work to predicting the combination of effects in three-dimensional buildings considering the three components of an earthquake ground motion is of personal interest.

References

1. Eads, L. (2013). *Seismic collapse risk assessment of buildings: effects of intensity measure selection and computational approach* (Doctoral dissertation, Stanford University).
2. ASCE. (2010). "Minimum design loads for buildings and other structures." ASCE/SEI 7-10 including Supplement No.1, Reston, VA
3. Diaz, O., Mendoza, E., & Esteva, L. (1994). Seismic ductility demands predicted by alternate models of building frames. *Earthquake Spectra*, *10*(3), 465-487.
4. Esteva, L., & Ruiz, S. E. (1989). Seismic failure rates of multistory frames. *Journal of Structural Engineering*, *115*(2), 268-284.
5. Santa-Ana, P. R., & Miranda, E. (2000, January). Strength reduction factors for multi-degree-of-freedom systems. In *Proceedings of the 12th world conference on Earthquake Engineering* (Vol. 1446). Auckland, New Zealand.
6. Miranda, E., & Bertero, V. V. (1994). Evaluation of strength reduction factors for earthquake-resistant design. *Earthquake spectra*, *10*(2), 357-379.