

JSM 2018 Poster

Activity Number: 181 - Contributed Poster Presentations: Section on Statistical Education

Type: Contributed

Date/Time: Monday, July 30, 2018 : 10:30 AM to 12:20 PM

Sponsor: Section on Statistical Education

Title: Classroom Demonstration: Deep Learning for Classification and Prediction, Introduction to GPU Computing

Author(s): Eric Sues*
*

Companies: CSU East Bay

Keywords: Artificial Neural Networks ; Deep Learning ; Classification ; Prediction ; parallel processing ; GPU computing

Abstract: We present examples of the use of basic Artificial Neural Networks (ANNs) for introductory Statistics classes at the undergraduate, major and first year graduate classes. Because of the available packages in R, ANNs are easily included in the discussion of Statistics classes as alternative methods to logistic regression and linear regression.

With the increases in computational power (parallel computation on CPUs, parallel computation on GPUs, TPUs, and NPUs, and with increases in RAM) Deep Learning has become possible. With the newer packages in R to connect to h2O, tensorflow, and keras, implementing Deep Learning is possible.

We present examples for running ANNs and Deep Learning in Statistics classes with discussion of the similarities and differences between traditional Statistical Methods and Deep Learning.

Outline:

1. Introduction to Prediction in Statistics

In Statistics the idea of Prediction is primarily discussed with linear models and predictions are estimated by plugging specific values of the input variables where a prediction of Y is to be made. Confidence Intervals are computed for the mean prediction or and Prediction Intervals are computed for a future value.

Intro Statistics: Prediction is the estimation of an output variable from an input variable or variables.

The variables are both *numeric* and usually *continuous*, but many examples use *discrete* values and ignore this fact.

$$Y = \beta_0 + \beta_1 X$$

Statistics Majors: Prediction is the estimation of the dependent variable Y , call it \hat{Y} , from the independent variables X_1, X_2, \dots, X_p .

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_p X_p$$

The variables are all assumed to *numeric* and usually *continuous* or *discrete*. The use of indicator variables are introduced for the inclusion of categorical variables as independent variables.

The linear model that is estimated is

$$\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$$

where $\epsilon \sim N(0, \sigma_\epsilon)$.

First Year Graduate Students: Prediction is the estimation of the dependent variable Y from the independent variables X_1, X_2, \dots, X_p

$$\hat{Y} = \hat{f}(X_1, X_2, \dots, X_p)$$

At this level, linear and non-linear functions of the data are considered or transformations of the Y variable, as in Generalized Linear Models (GLMs).

The variables used in linear models are assumed to be *numeric* and usually *continuous* or *discrete*. Indicator variables are used for categorical independent variables. Functions of the independent variables are introduced for nonlinearities and interaction terms. Link functions are used for GLMs.

- Example
 1. House Prices

1. Introduction to Classification in Statistics

In Statistics the idea of Classification is presented is discussed with categorical variables.

Intro Statistics: Which subgroup in a two-by-two table has the highest probability?

Statistics Majors: Which conditional subgroup in a two-by-two table has the highest conditional probability? Maybe Logistic Regression is introduced.

First Year Graduate Students: Logistic Regression is introduced and predicted probabilities are used to make predictions of a binary dependent variable $Y = 0$ or $Y = 1$. Maybe Multinomial Regression and/or Poisson Regression is introduced in a course on Categorical Variables.

- Examples:
 1. Iris Data
 2. Titanic

1. Introduction to Prediction and Classification in Machine Learning

In Machine Learning prediction and classification are discussed in terms of the algorithms used and presented using the Holdout Method, using Training, Validation, and Testing datasets. So thinking in terms of measuring how well the model will perform on future data.

Prediction algorithms: Linear Regression, Decision Tree, Random Forests, SVMs

Prediction evaluation: MSE, MAE, R^2 , adjusted R^2

Classification algorithms: kNN, Naive Bayes, Logistic Regression, Decision Trees, Random Forests, SVMs

Classification evaluation: Accuracy, Sensitivity, Specificity, ROC, Recall, Precision, F1

1. Connections between the use of Linear Regression in Statistics and Prediction in Machine Learning

At all levels of Classical Statistics instruction, Linear Regression is the primary example that is discussed when prediction is presented. Prediction is mostly discussed in terms of Prediction Intervals. The estimation of a new observation within the range of the data, otherwise, Forecasting is discussed.

In Bayesian Statistics Prediction is discussed in terms of Predictive Distributions and the associated credible intervals.

In Machine Learning numeric prediction is performed using Decision Trees and Random Forests, Support Vector Machines (SVMs). Closeness of the predicted values to the holdout values is evaluated using a Loss Function such as MSE.

1. Connections between Logistic Regression in Statistics and Classification in Machine Learning

Logistic Regression is usually not discussed in Introductory Statistics. This is unfortunate because of the results of applying logistic regression might be better appreciated by those students than hypothesis testing for proportions.

1. What is an Artificial Neural Network?

How to explain aNNs for the students in different levels of Statistics Education?

First, it is important for students of Statistics to be aware of and be able to use aNNs. The reason is because so many of their devices (cell phones, computers, smart speakers) now use these methods. From facial recognition to translation to voice recognition.

Second, it should be pointed out that aNNs can be used to perform similar tasks as they are learning in Statistics for prediction and classification.

The idea of prediction is well presented in Statistics Education, but classification is not so well presented. And it is classification that is one of the main uses of Neural Networks. Topics including spam filtering and digit classification could become examples in Introductory Statistics courses though major courses and Graduate classes. These are modern extensions of the classification of Fisher's iris data.

Intro Statistics: Modern Machine Learning technique. A black-box

Statistics Majors: Feed-Forward Neural Network

First Year Graduate Students: Feed-Forward Neural Network, Convolutional Neural Networks, and beyond. The discussion of the methods for estimating the weights on the network connections between the neurons can be discussed, including back propagation and dropout.

1. What is Deep Learning?

There is a lot of discussion about Deep Learning these days from companies like Google. It was Google's Chief Economist Hal Varian who has said, **I keep saying the sexy job in the next ten years will be statisticians. People think I'm joking, but who would've guessed that computer engineers would've been the sexy job of the 1990s?** Our ten years are about up and we have not added some of these new examples from computer science. We should to keep students interested. Don't forget that Fisher what working on Biology problems and Tukey had a background in Chemistry. Many of the advances in our field have come from people working in other areas.

1. Packages in R: neuralnet and nnet, h2o, tensorflow, keras

There are packages written in R that can be used, these are the neuralnet and nnet packages.

There is a package that can be installed to use h2o and others for tensorflow and keras. The advantage of these are that the software that R is connecting to are parallelized for use on multicore computers or clusters and they can be used with GPUs.

- Examples:
 1. Iris Data using keras and tensorflow
 2. Titanic
 3. cats and dogs

1. With increased Ram (32 gigabytes, 64 gigabytes, 128 gigabytes) larger datasets can be run.

Getting more Ram can really help when trying to run bigger Deep Learning models.

1. With GPUs the processing of deep learning models is much fast.

Figuring out how to use a nVidia GPU can be much faster on the same computer than using a CPU.

1. New TPU's from Google and NPU's are allow the running of very large Deep Learning models.

Finally, cloud computing can be used to access GPUs and Google TPU's. Google has introduced Python Notebooks with GPUs in it's [colab] which lets you save your Python notebooks to Google Docs.

1. Conclusions:

Should be introduce Neural Networks in the Statistics curriculum?

I would say yes and I would encourage they be introduced in Introductory books. The value of linear regression and logistic regression as first steps toward working with Big Data and Deep Learning is high. As students realize that linear regression and logistic regression can do the same things overall interest in Statistics will continue to grow.

References:

- American Statistician Prediction Limits for a Univariate Normal Observation by G. A. Whitmore
- R Journal neuralnet: Training a Neural Networks by Frauke Günther and Stefan Fritsch
- packt R Deep Learning Essentials by Dr. Joshua F. Wiley *h2o*
- Deep Learning with R by Francois Chollet, J. J. Allaire *tensorflow* and *keras*
- Deep Learning with Python by François Chollet *tensorflow* and *keras*
- Deep Learning by Ian Goodfellow, Yoshua Bengi, and Aaron Courville