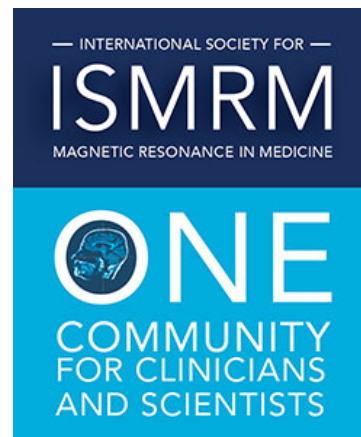


A RIDICULOUSLY FAST INTRODUCTION

MACHINE LEARNING FOR MR

Douglas Kelley, PhD, MIDS

douglas.kelley@med.ge.com



25TH Annual Meeting

& Exhibition • 22–27 April 2017

SMRT 26th Annual Meeting • 22–24 April 2017

HONOLULU, HI, USA

www.ismrm.org • www.smrt.org



Declaration of Financial Interests or Relationships

Speaker Name: Douglas Kelley

I have the following financial interest or relationship to disclose with regard to the subject matter of this presentation:

Company Name: General Electric

Type of Relationship: Employee

AGENDA

- ▶ What is Machine Learning?
- ▶ What kinds of problems can it solve?
 - ▶ Buzzword bingo
- ▶ How does it work?
- ▶ What is Deep Learning?
- ▶ What tools and resources are available?

WHAT IS MACHINE LEARNING?

- ▶ Machine learning is a set of techniques for constructing models of data using (mostly) just the data
- ▶ "All models are wrong; some models are useful." -- G Box
- ▶ Machine learning is the part of AI that works.
- ▶ Classification -- prediction of a categorical variable
- ▶ Regression -- prediction of a numerical variable
- ▶ Graph analysis (including search)

WHAT KINDS OF PROBLEMS CAN ML SOLVE?

- ▶ Classification: k-means, support vector machines, logistic regression, decision trees and random forests, topic modeling
- ▶ Regression: linear regression, regularization
- ▶ Graph analysis: search, PageRank, network analysis

BUZZWORD BINGO

- ▶ Supervised -- I have some previously analyzed examples
- ▶ Unsupervised -- I don't
- ▶ Training set -- the data I want to use to set up the model
- ▶ Test set -- other data I want to use to see how the model works
- ▶ Overfitting -- tailoring the model to accommodate the peculiarities of the training set while degrading the performance on the test set (and anywhere else)

BUZZWORD BINGO CONTINUED

- ▶ Precision -- I found all the needles in the haystack
- ▶ Recall -- everything I found is a needle
- ▶ Accuracy -- both have to be true -- **what does this mean for a population?**
An individual?
- ▶ Descriptive models -- I have no idea how it works
- ▶ Generative models -- this is what the model looks for
- ▶ Algorithmic bias -- believing something because the computer said so
- ▶ No Free Lunch Theorem – what works on the training set will not work any better on unobserved data....

HOW DOES MACHINE LEARNING WORK?

**TASK
PERFORMANCE MEASURE
EXPERIENCE**

- ▶ Data aggregation and feature extraction (T, E)
- ▶ Train-test split (P)
- ▶ Model construction and training (T, E)
- ▶ Model evaluation (P, E)
- ▶ Model correction (P, E)
- ▶ Model deployment (T, P, E)

WHAT IS DEEP LEARNING?

- ▶ Many functions can be represented by a multilayer perceptron – universal approximation theorem
- ▶ Training is lengthy and expensive but they run fast
- ▶ There are several mature frameworks for implementing deep learning pipelines

WHAT TOOLS AND RESOURCES ARE AVAILABLE?

- ▶ The Elements of Statistical Learning <https://statweb.stanford.edu/~tibs/ElemStatLearn/>
- ▶ Deep Learning <http://www.deeplearningbook.org>
- ▶ Scikit-learn <http://scikit-learn.org/stable/>
- ▶ Spark <https://spark.apache.org/docs/latest/ml-pipeline.html>
- ▶ Caffe <http://caffe.berkeleyvision.org>
- ▶ Tensorflow <https://www.tensorflow.org>

KEEPING UP WITH THIS

- ▶ GitHub repository:
- ▶ Blogs
- ▶ Podcasts
- ▶ Should we create a study group?

KEY QUESTIONS TO ASK

- ▶ What learning algorithm are you using?
- ▶ How did you train it? What was your objective function?
- ▶ How did you assess the accuracy? Test vs train?
- ▶ How do you serialize/featurize the input data?
- ▶ Is your model generative? Can you tie its performance to something that makes sense?
- ▶ How will you update the model if (your assumptions about) the data change?