Being data smart in the 21st century
An Introduction to Machine Learning

Elizabeth Traiger PhD MSc
02 October 2017
Data Big Bang Explosion

- Δ Sources
- Δ Resolutions
- Δ Formats
- Δ Accuracy
- Δ Rates
- Δ Measurements
- Δ Content
- Constant Δ
Smart Data, Not Big Data Hype

Being Data Smart is being smart about...

Data Value, Benefits and Limitations

Combining Data

What to Collect

Data Hardware & Software Architecture Utilization

Assumptions from Physics

Analysis & Algorithms
Statistics versus Machine Learning

Statistics

- Primary Data Analysis
- ‘Top Down’
- Hypothesis Testing
- Model Driven
- Confirmatory Analysis

Machine Learning

- Include Secondary Observational Data
- Hypothesis Generation
- Data Driven
- Knowledge Discovery
Machine Learning Basics

- **Supervised**
  - Classification
  - Regression

- **Unsupervised**
  - Clustering
  - Dimension Reduction

- Training Set / Validation Set

(Plus Reinforcement Learning)
Common Machine Learning Algorithms

- Regression Algorithms
- Instance-based Algorithms
- Regularization Algorithms
- Clustering Algorithms
- Decision Tree Algorithms
- Bayesian Algorithms
- Artificial Neural Network Algorithms
- Ensemble Algorithms
Supervised Learning in minutes
Linear Discriminants
Linear Discriminants - Classes
Linear Discriminants – Poor Classification (Conservative Bias Pink)
Linear Discriminants – Poor Classification (Bias Blue)

4 wrong
Logistic Regression – Better Classification via Probabilities
Support Vector Machines (SVM) – Buffer Planes
Support Vector Machines – Attempt 1
Support Vector Machines – Attempt 2
Support Vector Machines – The Kernel Trick
Support Vector Machines – Attempt 3
Support Vector Machines – Advanced
K-Nearest Neighbours - Clustering
K-Nearest Neighbours

$k=1$
K-Nearest Neighbours
K-Nearest Neighbours – Conflict Case
K-Nearest Neighbours – Conflict Case

$k=1$
K-Nearest Neighbours – Conflict Case

$k=2$
K-Nearest Neighbours – Conflict Case
K-Nearest Neighbours – Advanced
Decision Tree Learners – Flow Chart
Decision Tree Learners – Flow Chart
Decision Tree Learners
Ensemble Models – Use Multiple Learners
ML Applications within Resource Assessment

- Power to move beyond 3-dimensions and limitations of complex physics
- In the near future
  - Specification of Power Curves
  - Assimilating Satellite Data
  - Use of Observational Data
  - Incorporating Orography and Roughness
  - Treatment of Stability
  - Time Series Predictions
  - Portfolio Preconstructions
  - Detrending Series
  - Verification
  - New Types of Probabilistic Predictions
Thank You

Elizabeth Traiger PhD MSc
Elizabeth.traiger@dnvgl.com

Image Credits

Brian Lange
Jason Brownlee
Noun Project
scikit-learn documentation

www.dnvgl.com

SAFER, SMARTER, GREENER