

How to Break Into Data Science and Machine Learning

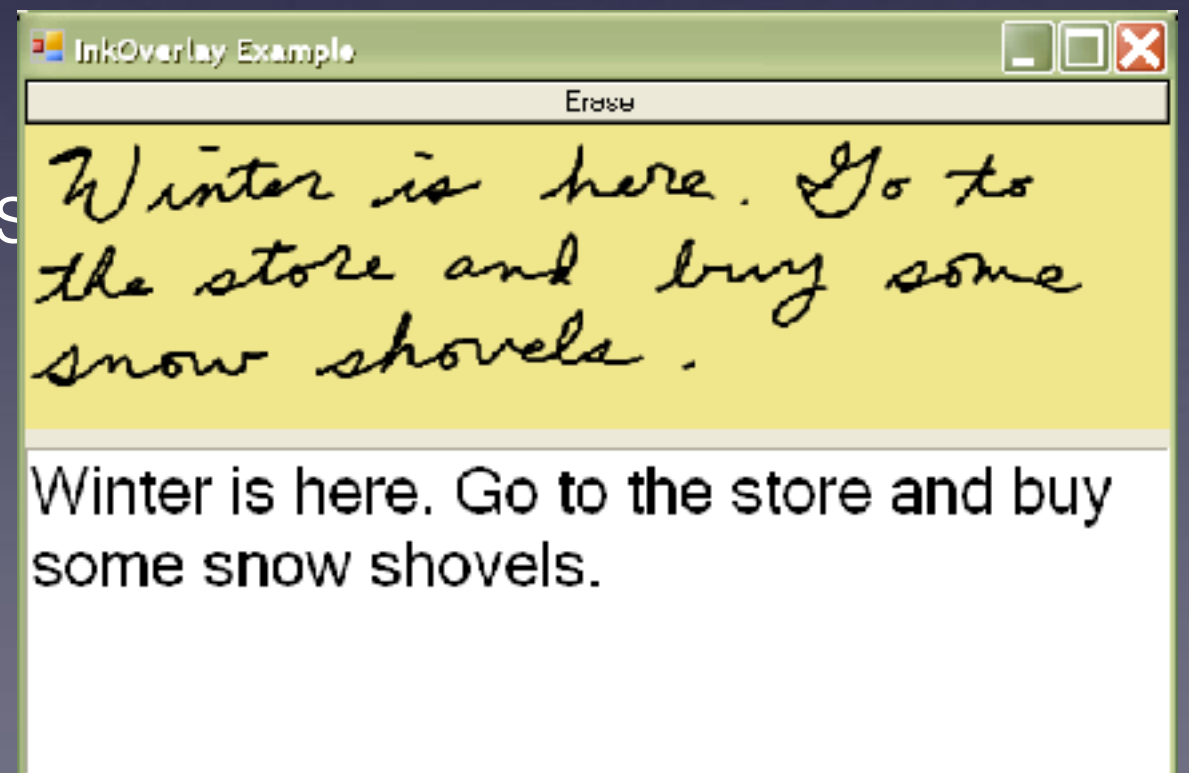
Joe Silverstein

My Background

- Former PhD student at Penn
- Have worked as data scientist at Nielsen, SeedInvest, and AI Labs
- Have a lot of experience with applying and interviewing for data science jobs

Things You Can Do With Machine Learning / Statistics

- spam filtering
- targeted advertising
- product recognition
- speech recognition
- object recognition in images
- handwriting recognition
- process automation



Things You Can Do With Machine Learning / Statistics

- financial forecasting
- fraud detection
- delivery logistics
- self-driving cars
- disease prediction
- artificial intelligence
- robotics



Skills Needed For Data Science

- Statistics / Machine learning
- Programming
- Some data science jobs are more programming heavy, whereas others are more ML/stats heavy. They all have some combination of both though.
- Depending on the type of data science job, a strong mathematical background could help

Course Offering

- Planning on offering data science bootcamp
- Python-based
- I put together a preliminary syllabus

Overview of Syllabus

- Week 1: Computer preliminaries
 - basic programming, etc.
- Weeks 2-3: Math, stats, and computer science preliminaries
 - calculus, basic linear algebra, probability, statistics, Bayes' Rule, correlation vs. causation, bias/variance tradeoff, etc.
- Weeks 4-5: Intro to supervised ML algorithms
 - scikit-learn, classification, regression
 - decision trees, random forests, tree boosting, support vector machines, logistic regression, naive Bayes, neural networks, etc.

Overview of Syllabus

- Week 5: Unsupervised learning
 - principal component analysis, k-means clustering, etc.
- Week 6: natural language processing
 - bag-of-words, n-gram, topic modeling, etc.
- Week 7: Time series analysis
 - ARIMA, GARCH, VARIMA, etc.
- Week 8: Working with large datasets
 - Hadoop, MapReduce, etc.

Overview of Syllabus

- Weeks 9-11: Special Topics
 - density estimation, efficiently optimizing hyperparameters, computer vision, causal inference, R programming, etc.
- Capstone Project Presentations and Critiques

Applying and Interviewing for Data Science Jobs

- For getting interviews, keywords are everything
 - keywords: machine learning (ML), natural language processing (NLP), Python, big data, scikit-learn, etc.
 - Looks good if you've done well in Kaggle competitions, but it's not necessary
- Typically there will be a data science take home assignment