This page:

https://cutt.ly/gcc20-mlsec

virtualbox image:

https://drive.google.com/file/d/1KQ0gFTGuJkXBn5tT-v_r_osWEwXxCCN6

Slides:

https://docs.google.com/presentation/d/1PMPLjRLEvoiUQKD_xE2l5MAj35wCgfFTvibpAsLAaUQ/edit?usp=sharing

Github repo:

https://github.com/cchio/qcon19-mlswe

Recommended resources

Halevy, Norvig, Pereira, The unreasonable effectiveness of data
Feynman, R. (1974, June). Cargo cult science. Engineering and Science 37(7).
Domingos, A few useful things to know about machine learning
Hawkins, The problem of overfitting
Paul Graham on Naive Bayes (2002)
http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Recommended online courses:

https://www.coursera.org/learn/machine-learning
https://www.coursera.org/lecture/neural-networks-deep-learning/geoffrey-hinton-interview-dcm5r

email:

mail@cchio.org

Books:

Elements of Statistical Learning - really deep textbook, the "bible" of ML
https://web.stanford.edu/~hastie/Papers/ESLII.pdf
Deep Learning Book
https://github.com/janishar/mit-deep-learning-book-pdf

Day 1:

Scikit-learn
https://scikit-learn.org/stable/index.html
Visual Introduction to Decision Trees
http://www.r2d3.us/visual-intro-to-machine-learning-part-1/
PCA visualization
http://setosa.io/ev/principal-component-analysis/
K-means clustering
https://www.naftaliharris.com/blog/visualizing-k-means-clustering/

~/Desktop/notebooks/deep-learning/mnist
LSTMs
https://colah.github.io/posts/2015-08-Understanding-LSTMs/
Jupyter notebook (new version of iPython)
https://jupyter.org/
pyspark -> to launch spark powered ipython notebook
TREC Spam Corpus
https://plg.uwaterloo.ca/~gvcormac/treccorpus07/

Parallelizing Decision Trees (an example)
http://www.jmlr.org/papers/volume11/ben-haim10a/ben-haim10a.pdf
Automatic Patch Generation
https://groups.csail.mit.edu/pac/patchgen/