

Multi-class Boosting

Start with a recent paper: <http://rob.schapire.net/papers/multiboost-journal.pdf>
and work backwards (look at references)

In particular, Multi-class AdaBoost:
<http://ww.web.stanford.edu/~hastie/Papers/SII-2-3-A8-Zhu.pdf>

Bandits

Survey by Bubeck and Cesa-Bianchi (basic model + regret):
<http://sbubeck.com/SurveyBCB12.pdf>

Bandits with Switching Costs: $T^{\{2/3\}}$ Regret: <https://arxiv.org/abs/1310.2997>

Active learning

Perhaps the first paper:
http://www.cs.northwestern.edu/~pardo/courses/mmml/papers/active_learning/improving_generalization_with_active_learning_ML94.pdf

Survey (theory inclined): <http://cseweb.ucsd.edu/~dasgupta/papers/twoface.pdf>

Online learning

Follow the leader and related strategies (chapter 2 of
<http://www.cs.huji.ac.il/~shais/papers/OLsurvey.pdf>)

Optimization

Neural networks

Power of depth: <https://arxiv.org/abs/1512.03965> and
<https://arxiv.org/abs/1610.09887> and also <https://arxiv.org/abs/1602.04485>

Dropout: a simple way to prevent neural networks from overfitting:
<http://dl.acm.org/citation.cfm?id=2670313>

On the importance of initialization and momentum in deep learning:
<http://jmlr.org/proceedings/papers/v28/sutskever13.html>

ImageNet Classification with Deep Convolutional Neural Networks:
<https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>

On the difficulty of training Recurrent Neural Networks:
<https://arxiv.org/pdf/1211.5063.pdf>

Long Short-Term Memory: <http://dl.acm.org/citation.cfm?id=1246450>

Unsupervised learning

Compressed sensing

<http://statweb.stanford.edu/~candes/papers/ExactRecovery.pdf>

<https://statweb.stanford.edu/~candes/papers/StableRecovery.pdf>

Survey: <http://statweb.stanford.edu/~candes/papers/CompressiveSampling.pdf>

SGD inspired algorithms for eigenvector computation

<https://arxiv.org/abs/1509.09002>

<https://arxiv.org/abs/1507.08788>

Stochastic block models

http://www.princeton.edu/~eabbe/publications/SBM2_exact.pdf