The self-assessment exam is available at:  <http://cs.ucla.edu/~ameet/self_assessment.pdf>.  Note that the self-assessment exam also includes pointers to review material.

1. Selected Research Papers
   * [Spark: Cluster Computing with Working Sets](http://people.csail.mit.edu/matei/papers/2010/hotcloud_spark.pdf),   
     Matei Zaharia, Mosharaf Chowdhury, Michael J. Franklin, Scott Shenker, Ion Stoica.   
     USENIX HotCloud (2010)  
     .
   * [*Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing*](http://usenix.org/system/files/conference/nsdi12/nsdi12-final138.pdf)*,*  
     Matei Zaharia, Mosharaf Chowdhury, Tathagata Das,   
     Ankur Dave, Justin Ma, Murphy McCauley, Michael J. Franklin,   
     Scott Shenker, Ion Stoica  
     . NSDI (2012)
   * [MLlib: Machine Learning in Apache Spark](http://arxiv.org/pdf/1505.06807.pdf), X. Meng, J. Bradley, B. Yuvaz, E. Sparks, S. Venkataraman, D. Liu, J. Freeman, D. Tsai, M. Amde, S. Owen, D. Xin, R. Xin, M. Franklin, R. Zadeh, M. Zaharia, A. Talwalkar. Preprint (2015).

If we evaluate our model on the test set and then retrain the model to obtain a better result, it is possible that we will overfit our model to the data in the test set.  This means that our subsequent evaluation of the model could provide an overly optimistic view of the model's performance on future data.  Later in the course we will discuss how we can use another hold out set (called a validation set) to circumvent this issue.