Chapter 1: The Machine Learning Landscape

* Supervised vs. unsupervised
* Online vs. batch
* Instance vs. model based
* ML great for
  + Exisiting solutions which require lots of tuning or many rules
  + Complex problems without a good existing solution
  + Changing environments
  + Gaining insights about complex problems and large data sets
* Supervised vs. unsupervised
  + Supervised
  + Unsupervised
    - Clustering
      * Hierarchical
      * Visualization
    - Dimensionality reduction
    - Anomaly detection
    - Association rule learning
  + Semi-supervised
  + Reinforcement Learning
    - Agent
    - Penalties/rewards
* Batch vs. online
  + Batch: trained all at once, offline learning
  + Online: train incrementally, sequential, mini-batches
* Instance vs. model based
  + Instance: similarity to memorized casess
  + Model: system for analyzing new instance and making prediction
    - Success measures: utility or cost functions
* Challenges: most have to do with data
  + Not enough
    - “Unreasonable Effectiveness of Data
  + Not representative
    - Sampling noise
    - Sampling bias
  + Poor quality
  + Irrelevant
    - Feature selection
    - Feature extraction
  + Oveffiting: bias
    - Regularization
  + Underfitting: variance
* Testing and validation
* Hyperparameter Tuning and Model selection
* Data mismatch
  + No Free Lunch Theorem: A priori model preference