

Reading: Ch 1, Ch 2 except sec 2.8

1. (Supervised and unsupervised learning)
 - (a) Describe the differences between supervised and unsupervised learning, for example, the data structure, the probability model, task, etc.
 - (b) Which category do *classification* and *regression* belong to? What is the difference between them?
 - (c) For unsupervised learning, read the introduction to clustering and dimension reduction (especially, PCA).
2. (Basic concepts in ML)
 - (a) What is the curse of dimensionality
 - (b) Describe the probabilistic models for linear regression and logistic regression.
 - (c) What rule do we use for model selection? In practice, we often partition the dataset into training set and testing (validation) set. What problem may occur if we just select, say, 80% of the data to train and the rest for validation? How to fix it?
3. (Fundamentals in probability)
 - (a) Some basic concepts (probability and some properties, CDF, PDF for continuous RV, PMF for discrete RV, joint distribution, etc.)
 - (b) conditional probability, Bayes formula
 - (c) independence
 - (d) quantiles, mean, moments, variance, covariance and correlations
 - (e) Some common distributions