Pattern Classification and Machine Learning

FEN3202

Notes for Lecture 1

Saikat Chatterjee

I. Discussion Agenda

1) What is pattern recognition? Page 1 and 2 of the text book.
2) What a pattern recognizer should have? Page 2 and 3 of the text book.
   Discussion points: Generalization, pre-processing (feature extraction), supervised learning, classification, regression, unsupervised learning, clustering, density estimation, dimension reduction, reinforcement learning.
3) Tools for machine learning: Probability theory, Decision theory, and Information theory.
   Note: We will cover required Probability Theory
4) Example: Polynomial curve fitting Setion 1.1 (Page 4-12) of the text book.
   Discussion points: Example of sinusoidal data in noise and polynomial fit, training set, this simple example has a high scope with reality, Goal of the task, equation 1.1, linear and non-linear qualities in eqation 1.1, error cost and equation 1.2, unique solution exists for the chosen cost, model selection problem, over-fitting, equation 1.3, $M = 9$ case and exact fit, paradox in polynomial fit, what are the value of optimal weights (Table 1.1), Size of dataset, choice of model complexity, connection exists between least-squares and maximum likelihood, over-fitting problem occurs and Bayesian view is important, limited data and regularization, equation 1.4, regularization effect in Figure 1.8, validation set
5) Probability Theory Setion 1.2 (Page 12-32) of the text book.
   Discussion Points: Sum rule and product rule (equation 1.10 and 1.11), Bayes's theorem: equation 1.12, 1.13, probability density: equation 1.24, 1.25, 1.26, Expectations and covariances: equation 1.33, 1.34, 1.36, 1.37, 1.38, to 1.42, Section 1.2.3: Bayesian probabilities: frequesntist and Bayesian views (we read the whole paragraph carefully), Inference: equation 1.43, likelihood and posterior, equation 1.44, A frequenstist estimator: maximum likelihood, bootstrap, a criticism of Bayesian view, non-informative prior, Computational limitation of full Bayesian treatement, MCMC, variational Bayesian approach, expectation propagation, Gaussian distributions: equation 1.46, 1.49-1.52, Parameter estimation: equation 1.53, log likelihood, equation 1.54, sample mean and variance (in ML sense), equation 1.55 and 1.56, decoupling of parameters for Gaussian, Limitation of ML approach, bias, equation 1.57 and 1.58, Bishop says they are straight forward to prove (but are they

so easy?), Curve fitting re-visited: towards a full Bayesian treatement, equation 1.60, Figure 1.16, equation 1.61, 1.62, relation between sum-of-squarer cost and maximum likelihood (noise is Gaussian), equation 1.63, predictive distribution, equation 1.64, a step more towards full Bayesian, equation 1.65, Hyperparameters, MAP approach, equation 1.66-1.67, relation between MAP and regularization.