

CS599-Advanced Topics in Neural Computation and Statistical Learning

Questions for Jordan&Bishop Chapter 3

1. How could one characterize the difference between a probability theorist and a statistician?
2. What is the difference between a Bayesian and a frequentist approach to statistical learning?
3. Which three core classes of statistical learning exist?
4. What is Bayes Rule? How is it derived?
5. What are the definitions and applications of the following distributions:
 - a) Bernoulli distribution
 - b) Normal distribution
 - c) Binomial distribution
 - d) Multinomial distribution
 - e) Dirichlet distribution
 - f) Gamma distribution
 - g) Beta distribution
 - h) Uniform distribution
6. What is the bias and variance of an estimator?
7. What is maximum likelihood estimation?
8. What is the difference between Maximum Posterior and Mean Posterior estimators?
9. Why does a uniform prior over the parameters correspond to maximum likelihood estimation?
10. What is penalized maximum likelihood estimation?
11. Why does applying a monotonic function to a given function NOT change the optimum of this function?
12. Is maximum likelihood estimation a Bayesian estimation method?
13. What is the meaning of "conditional independence"?
14. How do Bayesian and Frequentist statistical learning differ in their predictions?
15. What is the difference between density estimation, regression, and classification?
16. What does it mean to have "identically distributed" random variables?
17. What does it mean to have "independently" distributed random variables?
18. What is the SAMPLE mean and SAMPLE variance?
19. Outline the procedure for maximum likelihood density estimation.
20. Outline the procedure for Bayesian density estimation.
21. How does one choose priors in Bayesian estimation?
22. What are hyperparameters in Bayesian estimation?
23. What is "completing the square"?

24. How many level of random parameters are usually assumed in Bayesian inference?
25. What is a convex combination of two variables?
26. What is a plate in Graphical Models?
27. What is the method of Lagrange multipliers?
28. Where is the application of mixture models useful?
29. What is the exponential family?
30. What is a convex function?
31. What is a marginal distribution?
32. What is nonparametric density estimation?
33. What is a kernel function, and what examples do you know?
34. Why is max likelihood not suitable to determine the smoothing parameters of kernels?
35. What is a mixture sieve?
36. What is the standard statistical model (the data generating model) for regression analysis?
37. What is nonparametric regression?
38. What is a conditional mixture model?
39. What is kernel regression?
40. What is the difference between classification and regression?
41. What is the difference between a generative and a discriminative approach to classification?