(Questions collected by Wén Shao)

COMP4670/8600 Exam Revision Questions

These questions may help in your preparation for the final exam. Sit down with a friend or by yourself and practice talking about each topic out loud. Please note that Jan and I have absolutely no access to exam questions, and as such these questions are not an exhaustive list of topics, nor are they necessarily related to the questions you may be asked in the actual exam. Use them to help practice for the oral exam but make sure you study material from the lectures, tutorials, and textbook as well.

2. What is supervised learning and what is unsupervised learning?

3. What is reinforcement learning and how it is different from statistical machine learning? Or explain your reason if you find that statistical machine learning fits into reinforcement learning framework.

4. What is overfitting in general? For a linear regression problem, how we can avoid overfitting?

5. Consider a linear regression model, is the model linear in the training data X?

6. Given a set of training data $(\mathbf{X},\mathbf{t})$, write down the linear regression model with regularisation.

7. Write down the D dimensional Gaussian distribution.

8. Assume the prevalence of a certain disease in the general population is 1%. Assume there exists a quite reliable test for the disease, say, the test on a diseased/healthy person is positive/negative with 99% probability. If the test (on some randomly selected person) is positive, what is the chance that (s)he has the disease? Explain the result.

9. Calculate the mean of a single variable Gaussian distribution $\mathcal{N}(x|\mu,\sigma^2)$.

10. Calculate the mean of a Bernoulli distribution $Bern(\theta)$

11. Suppose that x is distribution according to a continuous distribution $p(x)$, write down the weighed average of a function $f(x)$ under this particular distribution. How to approximate this value numerically?

12. Verify that for an arbitrary function $f(x)$, $var[f] = \mathbb{E}[(f(x)-\mathbb{E}[f(x)])^2] = \mathbb{E}[f(x)^2]-\mathbb{E}[f(x)]^2$.

13. Verify that the covariance of two random variables $x \in I\!R$ and $y \in I\!R$ can be calculated by $cov[x,y]=\mathbb{E}_{x,y}[(x-\mathbb{E}[x])(y-E[y])]=\mathbb{E}_{x,y}[xy]-\mathbb{E}[x]\mathbb{E}[y]$.

14. What is called a basis in a vector space?

15. What is the trace of a square matrix A?

16. What is a projection?

17. What is an orthogonal projection?

18. Given a matrix $A \in I\!R^{n\times p}$ and a vector $\mathbf{x}$. What is the closest point $\hat{\mathbf{x}}$ to $\mathbf{x}$ in the column space of A?

19. what is the rank of a matrix A.

20. Calculate the determinate of the following matrix.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

21. What is the singular value decomposition of a matrix A.

22. Calculate the gradient of the following function

$$f(X)=tr\{X^T CX\}$$

where $X \in \mathbb{R}^{n\times p}$ and $C \in \mathbb{R}^{n\times n}$

23. Explain the difference of frequentist's approach and Bayesian approach.

24. What is S-fold cross validation? Why we use it?

25. Derive a least squares solution for linear regression model (with regularisation).

26. What is stochastic gradient descent? How to apply it to linear regression problems specifically with the sum-of-squares error function?

27. What is the bias-variance decomposition? Demonstrate the bias-variance decomposition where the the error is measured by the mean squared error. What you can deduce from the result?

28. What is a conjugate prior? Why we normally use conjugate prior? What is the conjugate prior for Gaussian distribution.

29. What are the limitations of linear basis function models?

30. What is the curse of dimensionality? Why it can be a problem?

31. What are the three models for decision problems? How they are different?

32. What are the deficiencies of the least squares approach in linear classification?

33. What's the idea of Fisher's linear discriminant? Derive the fisher's linear discriminant that projects $x \in \mathbb{R}^D$ to $x \in \mathbb{R}^{D'}$ where $D > D'$.

34. What is the perceptron algorithm? Describe it in detail.

35. What is the probabilistic generative model? Describe it in detail.

36. What is logistic regression? Describe it in detail?

37. What is the feature mapping in classification problems? Why we need this feature mapping sometimes?

38. What is Laplace approximation? Why we need to do Laplace approximation sometimes? Describe how we do Laplace Approximation in detail.

39. Describe Bayesian Logistic Regression in detail.

40. Derive the formula for a feed-forward three layers neural network.

41. Why if all activation functions in a neural network are linear then there exists an equivalent network without hidden units?

42. What are the advantages of neural network compared with linear regression (classification) model?

43. Describe the weight-space symmetries in a neural network.

44. What are the methods to optimise the parameters in neural network? Describe it (them) in as much detail as you can.

45. What are the methods that are used to avoid overfitting in training a neural network?

46. Describe Bayesian Neural Network in as much detail as you can.

47. Derive the dual representation of a linear regression model with regularised sum-of-squared error.

48. What is the kernel function? What are the advantages of kernel method.

49. What are the operations that can preserve the validity of a kernel function (kernel functions)?

50. Describe Lagrange multipliers in detail. Especially consider that we have inequality constraints.

51. What is the objective of SVM? (What is the idea behind SVM?)

52. Describe SVM in as much detail as you can.

53. Describe relevance vector machine in as much detail as you can.

54. What is conditionally independency? How to determine conditionally independency in a given Baysian network? (In the exam, you will be probably given a Baysian network and Chris will ask questions around it)

55. What is a Markov random field.

56. What is a clique?

57. What is a bipartite graph?

58. What is the sum-product algorithm? What is the basic idea behind it? Describe it in detail.

59. Describe K-means Clustering.

60. Describe EM algorithm for mixture of Gaussians in detail.

61. Show that Gaussian distribution belongs to exponential family.

62. How do we sample from standard distributions?

63. Describe the method you would use to sample from the following distribution:

$$p(x=i)=\frac{1}{i}-\frac{1}{i+1},\ i=1,2,3...$$

64. Describe rejection sampling in detail.

65. Describe importance sampling in detail.

66. Describe Metropolis-Hasting algorithm in detail.

67. What is the idea behind principle component analysis?

68. What is a non-stationary distribution?

69. Use a Baysian network to represent a second-order Markov chain.

70. What is the difference between principle component analysis and Fisher's discriminant.

71. Prove that independence of two random variables implies uncorrelatedness. Give counter example to show that uncorrelatedness does not imply independence.

72. Assuming K different states for each variable x, how many parameters does a M-order Markov chain have?

73. What is a homogeneous hidden Markov model?

74. Describe Viterbi algorithm in detail.

75. What are the motivations for combining models?