Q2. Supervised learning aims at find a function that maps inputs into targets. Unsupervised learning is target not given.

Q3. Reinforcement learning is learning from interaction with an environment, from the consequences of action rather than from explicit teaching. The consequences of action may be viewed as supervision on the learning algorithms. So in the framework of statistical machine learning, reinforcement learning may be considered a kind of supervised learning.

Q4. Overfitting: model fits noise instead of generalisability. For linear regression models: more data, regularisation.

Q5. No. Linear in model parameters.

Q6.

Q7.

Q8.disease, positive

Q9.

Q10.

Q11. . The value can be approximated by importance sampling.

Q12.

Q13.

Q14. A set of linearly independent vectors that spans a vector space. Basis defines a coordinate system of the spanned vector space.

Q15. Sum of the diagonal entries

Q16. A projection is a linear transformation from vector space to itself such that

Q17. An orthogonal projection is a projection whose range and kernel are orthogonal subspaces.

Q18. Assuming has full rank,

Q19. Number of columns that are independent

Q20.

Q39.

Q40.

where are activation functions.

Q41.

Q42.

1) Linear regression/classification: Number and parameters of basis functions are fixed before learning starts.

NN: Number of basis functions is fixed before learning starts

2) NN is capable to perform non-linear tasks due to the nonlinearity introduced by the activation functions.

Q43.

The weights map the input space to the output space. Assume is used as the activation function. . Changing the sign of all inputs to a hidden unit and outputs of this hidden unit results in the same mapping.

Q44. Back-propagation. A feed forward NN can be represented as a directed graph. BP is the directed graph with reversed directions of FP. Mathematically, BP is an implementation of chain rule.

Q45. regularisation and early stopping

Q46.

Q47.

Substitute into to eliminate , we have the dual formulation

Q48. Kernel function defines an inner product of two vectors of basis functions

Advantages:

1) Map input space into higher dimensional space which is more linear separable

2) Kernel can be constructed to define the notion of distance for graphs, sets, strings and other complicated objects.

Q49. Addition and multiplication

Q50. If there are inequality constraints, one needs to solve the dual problem before one can solve the primal.

Q51.

Separable:

s.t.

Inseparable:

s.t.

Idea of SVM: Maximise the minimum margin. Margin is the distance between a point and the decision boundary

Q52.

Objective function

Idea of SVM

Algorithm to solve SVM (QP) (Sparse kernel machines pp.13)

Dual formulation and kernel trick

Sparsity – only support vectors count towards classification decision

Q53.

Q54. If , is conditionally independent of given

How to determine conditional independence from a Bayesian network:

equivalence between conditional independence and probability factorisation

Q55. MRF is defined over a graph with undirected edges. MRF allows for different conditional independence statements than Bayesian networks

Q56. a subset of the nodes which are fully connected

Q57. A bipartite graph is a graph whose vertices can be divided into two disjoint sets U and V such that every edge connects a vertex in U to one in V.

Q58.