Exam 2016

June 20, 2023

True or False:

1. Complexity control is necessary in non-linear methods only: F
2. The empirical error in the training set is always smaller (or equal) than the empirical error in the test set: T (in general)
3. The empirical error in the training set is always smaller (or equal) than the empirical error in the validation set: F (it depends, for example in LOOCV this does not generally holds)
4. Using a larger validation data set reduces the chances to select an overfitted model: T
5. Regularization usually penalizes models that are more complex than needed: T
6. Regularization may penalize models that are simpler than needed: F
7. Cross-validation guarantees that our model does not overfit the data: F
8. L2-regularization produces sparsity, as opposed to L1-regularization: F
9. The VC dimension of a two-class classifier is independent of data dimension: F
10. The VC dimension of a two-class classifier is always a finite integer: T
11. The Bayes formula transforms prior distributions into posterior distributions: T
12. The denominator in Bayes formula is enough to perform classification, by taking simply the maximum over the classes: F
13. The Bayes classifier is the best possible classifier when the prior and class-conditional distributions are known: T
14. For normally distributed classes, Bayesian classifiers turn out to be quadratic discriminant functions: T
15. For normally distributed classes, equal prior probabilities yield linear discriminant functions: F
16. The Naive-Bayes classifier can only be used with discrete random variables, because it assumes statistical independence among all variables, given the class: F
17. The kNN classifier needs no tuning of the number of neighbours, because in the limit of infinite data it is a Bayesian classifier: F
18. $\sum_b P(a|b) P(b) = 1$ where A, B are discrete random variables: F
19. $\sum_b P(b|a) = 1$ where A, B are discrete random variables: T
20. $\sum_b P(a|b) = 1$ where A, B are discrete random variables: F
21. The likelihood is a function of the data sample for a given choice of parameters: F
22. The negative log-likelihood sometimes yields different results than the likelihood: F
23. Logistic regression is a linear method that can be used to predict an arbitrary numerical quantity: F

24. Linear regression assumes normally distributed inputs and outputs: F
25. In a GLM, the model tries to predict the expected value of the target using a linear function of the predictors and a suitable interface function: T
26. The solution for a GLM can be found analytically by minimizing the log-likelihood, or iteratively using Newton-Raphson: T
27. In Poisson regression, we are interested in predicting positive outcomes that represent counts: T
28. The regression function is the best possible predictor, and would achieve zero error on the population: T
29. In statistics, bias and variance are related concepts: increasing one must increase the other, and viceversa (decreasing one must decrease the other): F
30. The mean squared error is always preferred for optimization, because it is the more theoretically sound: F
31. Non-linear functions of the data can be estimated by using linear fitting techniques: T
32. A linear combination of non-linear functions with adaptive parameters is a linear model: F
33. The backpropagation algorithm computes the partial derivatives of the given error function with respect to the network weights: T
34. The backpropagation algorithm must be coupled with an optimization method (update rule) to make it a learning algorithm for a MLP: T
35. Even if we fix the initial weights, a MLP is a non-deterministic method: F
36. The activation function for the output neurons is dictated by the nature of the target variable: T
37. A MLP requires the specification of the number of hidden neurons, which can be done in a variety of ways: T
38. RBF and MLP neural networks can be seen as a particular case of the same class of neural networks: ?
39. In a RBF neural network there is no regularization, because they are based on Euclidean distances instead of inner products: ?
40. Regularization does not make sense in neural networks, because they learn adaptive regressors (regressors with parameters): F
41. The k-means algorithm converges to a global optimum as the number of iterations goes to infinity: F
42. A Gaussian mixture model assumes that the data has been generated by some finite mixture of Gaussians: T
43. The k-means algorithm can be used to initialize a Gaussian mixture model: T
44. A Random Forest is "random" partly because the variables used in each decision tree are optimized amongst a randomly chosen subset: F
45. A Random Forest is "random" partly because the variables used in each decision node are optimized amongst a randomly chosen subset: T
46. A Random Forest is "random" partly because the data used in each decision tree come from a different bootstrap resample: T
47. A Random Forest is "random" partly because the data used in each decision node come from a different bootstrap resample: F
48. In Machine Learning, there is no limit on the achievable predictive performance of a model, it is just a matter of choosing the correct method, and tuning the parameters: F
49. In Machine Learning, pre-processing can make a large impact on learning, and therefore on predictive performance: T
50. A system (living or not) learns when it uses past experience to improve future performance: T