Exam 2017

June 20, 2023

True or False:

1. Based on training data alone, there is no means of choosing which model is better: T
2. Complexity control is only necessary when data is high-dimensional: F
3. The empirical error in the training set is always smaller (or equal) than the empirical error in the test set: F (not guaranteed, but in most of the cases it is T)
4. Using a larger training data set reduces the chances to obtain an overfitted model: T
5. Regularization is intended to penalize models that are less complex than needed: F
6. Cross-validation is mainly used for model selection purposes: T
7. L2-regularization does not produce sparsity, as opposed to L1-regularization: T
8. Reducing the hypothesis (model) space is a way of controlling complexity: T
9. The VC dimension of a two-class linear classifier is a linear function of data dimension: T
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 numerator in Bayes formula is enough to perform classification, by taking simply the maximum over the classes: T
13. The Bayes classifier is the best possible classifier when the classes are Gaussian: F, unless we know the parameters of the distributions
14. The Bayes classifier is the best possible classifier when the true priors are known: T
15. For normally distributed classes, Bayesian classifiers turn out to be quadratic discriminant functions: T
16. For normally distributed classes, equal posterior probabilities yield linear discriminant functions: T
17. The Naive-Bayes classifier does not make assumptions about data distribution for continuous variables: F
18. The kNN classifier requires tuning of the number of neighbours, because we have a finite data sample: T
19. $\sum_a P(a|b) P(b) = P(b)$ where A, B are discrete random variables: F
20. $\sum_b P(a|b) = \sum_a P(b|a)$ where A, B are discrete random variables: F
21. The likelihood is a function of the parameters for a given choice of data sample: T
22. Logistic regression does not make assumptions about input data distribution: T
23. Linear regression assumes normally distributed outputs, conditioned on the inputs: T
24. 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

25. We can obtain an error function as the negative log-likelihood of a problem: T
26. The regression function is the best possible predictor, in the sense that it would achieve zero bias: F
27. The regression function is the best possible predictor, in the sense that it would achieve zero variance: F
28. The regression function is the best possible predictor, in the sense that it would achieve zero noise: F
29. In statistics, bias and variance are related concepts: they represent the distribution of errors in the training and test sets, respectively: F
30. The mean squared error is always preferred for GLM regression, because it is the only one that works in practice: F
31. An MLP needs no regularization, because backpropagation prevents arbitrary growth of the weights: F
32. We can convert a non-linear model into a linear one by giving values to the non-linear adaptive parameters: F
33. The backpropagation algorithm computes the partial derivatives of a given differentiable 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 neural network: T
35. The backpropagation algorithm is mainly used to compute the gradient vector of the error function at each step: T
36. The nature of the target variable dictates the activation function for the output neurons: T
37. The activation function for the hidden neurons could be a linear function to facilitate learning: F
38. Both RBF and MLP neural networks can have one or more hidden layers of neurons: T
39. An RBF neural network could in principle be trained with the backpropagation algorithm: T
40. Regularization makes little sense in neural networks, because they are non-linear models: F
41. The E-M algorithm refines a suboptimal solution obtained by k-means until a global optimum is found: F
42. A Gaussian mixture model assumes that the data has been generated by a "big" Gaussian that can be decomposed as a finite mixture: T
43. The k-means algorithm will discover the true clusters in the data, if given enough prototypes: F
44. Bagging methods are based on the fact that, for unstable learners, variance can be greatly reduced with little or no increase in bias: T
45. A Random Forest is "random" because decision trees are random learners (meaning that a single tree changes if we "execute" the algorithm again): F
46. A Random Forest is "random" because the data used in each decision node come from a different bootstrap resample: T
47. In Machine Learning, the lack of predictive variables can be compensated by more training data; in other words, there is no limit on the achievable predictive performance of a model, if we can gather enough data: F
48. We should optimize the number of folds in cross-validation, and separately for each modeling technique: F, we should decide this based on the size of the dataset at hand
49. In Machine Learning, better pre-processing can make a large impact on learning, and therefore on predictive performance: T
50. In a noiseless setting, at least theoretically speaking, there is no need for regularization: F, because we don't know the real generating function, so we can still be choosing a too complex model