Exam 2015

June 21, 2023

1. If we keep complexity low, we do not need to care about training error: F
2. Training error is always smaller (or equal) than test error: T (in general, unless very scarce data and lucky shots)
3. Supplying more training data reduces the chances to obtain an overfitted model: T, in general
4. Regularization penalizes models that are either simpler or more complex than needed: F, only more complex
5. Training error is enough to perform model selection: F
6. The VC dimension for a two-class classifier penalizes training sample size: ?
7. The VC dimension for a two-class classifier is the maximum number of linear separations that the classifier can perform: ?
8. In order to check that the VC dimension is (at least) some integer k, we just need to find k points that can be shattered: ?
9. Checking that the VC dimension is infinite requires an infinite number of checks: ?
10. A two-class classifier with infinite VC dimension must have an infinite (or very large) number of parameters: ?
11. The Bayes formula converts prior distributions into posterior distributions: T
12. The Bayes formula is of theoretical importance, but can never be used in practice: F
13. The numerator in Bayes formula is enough to perform classification: T
14. The Bayes classifier is the best possible classifier when the prior and posterior distributions are known: T (the prior and the class-conditional are needed. But this is to compute the posterior, so if you already know the posterior, I guess it should also work)
15. For normally distributed classes, Bayesian classifiers turn out to be quadratic discriminant functions (QDA): T
16. For normally distributed classes, statistical independence among all variables yields linear discriminant functions (LDA): F
17. For normally distributed classes, Bayesian classifiers are minimum-distance classifiers: T
18. The Naive-Bayes classifier assumes statistical independence among all variables: T
19. The kNN classifier can be explained as a Bayesian classifier: F
20. The kNN classifier works better with more neighbours, although it is computationally more costly: F
21. The likelihood of a sample is its density for a given choice of parameters: T (kinda)
22. The likelihood of a sample is a function of the sample: F

23. Logistic regression is a generative linear classifier: F (it is discriminative)
24. Logistic regression assumes normally distributed classes: F
25. In a Generalized Linear Model, the prediction is the logistic function applied to a linear function of the predictors: F
26. The solution for Logistic regression can be found analytically by minimizing the log-likelihood: F
27. In Poisson regression, we are interested in predicting integer outcomes, which are equally likely: ?
28. In a Generalized Linear Model, we always find the logistic function in one way or another (so it is called the link function): F
29. In statistics, bias and variance are opposite concepts: increasing one must decrease the other: F
30. Variance always decreases with increasing sample size; however, bias can increase or stay the same: T
31. The regression function is the best possible model in regression, and achieves zero error on the training data: F
32. The risk is equal to the sum of the (squared) bias, the variance and the noise variance: T
33. The theoretical MSE does not depend on the regression function: F
34. Models that are "more complex than needed" will tend to have a large bias and large variance: F
35. Models that are "less complex than needed" will tend to have a small bias and small variance: F
36. A linear combination of non-linear (fixed) functions of the inputs make a linear model: F
37. Ridge regression adds a penalty term to a linear model such that the new model is non-linear: F
38. Non-linear functions of the data can be estimated by using linear fitting techniques: T
39. Both RBFs and MLPs create non-linear models by learning adaptive regressors (regressors with parameters): ?
40. Regularization allows the specification of models that are more complex than needed; it also helps numerically: F
41. The k-means algorithm converges to a global optimum if the number of iterations goes to infinity: F
42. A Gaussian mixture model assumes normality of the training data: F
43. The k-means algorithm is used to fine-tune a Gaussian mixture model after the latter has converged: F
44. The backpropagation algorithm computes the partial derivatives of the error function with respect to the network weights: T
45. The backpropagation algorithm must be coupled with an optimization method (update rule) to make it a learning algorithm for a MLP: T
46. A MLP requires the specification of the number of hidden neurons; this is best done by trial-and-error, monitoring the fitting error of the network: T
47. RBF neural networks are a particular case of MLP networks: ?
48. In a RBF neural network, regularization does not make sense, because it is based on Euclidean distance: ?
49. Feature selection can never increase the practical performance of a learning method, it only reduces learning time: F
50. Feature selection can be performed after feature extraction, using the extracted features as new variables for selection: T