Machine Learning — Statistical Methods for Machine Learning

Quiz list for the written test

Instructor: *Nicolò Cesa-Bianchi* version of June 8, 2023

A variable subset of this quiz list will be used to create the written test in each exam session. The list will be expanded as the course nears its end. Bonus quizzes (not in this list) will be added to each test for extra points.

1. Write the formulas for the square loss, the zero-one loss, and the logarithmic loss.
2. What does a learning algorithm receive in input? And what does it produce in output?
3. Write the mathematical formula defining the training error of a predictor h.
4. Write the mathematical formula defining the ERM algorithm over a class $\mathcal{H}$ of predictors. Define the main quantities occurring in the formula.
5. Explain in words how overfitting and underfitting are defined in terms of behavior of an algorithm on training and test set.
6. Name and describe three reasons why labels may be noisy.
7. Is k-NN more likely to overfit when k is large or small?
8. Write a short pseudo-code for building a tree classifier based on a training set S.
9. What is the property of a splitting criterion ψ ensuring that the training error of a tree classifier does not increase after a split?
10. Write the formulas for at least two splitting criteria ψ used in practice to build tree classifiers.
11. Write the formula for the statistical risk of a predictor h with respect to a generic loss function and data distribution.
12. Write the formula for the Bayes optimal predictor for a generic loss function and data distribution.
13. Write the formula for Bayes optimal predictor and Bayes risk for the zero-one loss.
14. Can the Bayes risk for the zero-one loss be zero? If yes, then explain how.
15. Write the formula for Bayes optimal predictor and Bayes risk for the square loss.
16. Explain in mathematical terms the relationship between test error and statistical risk.
17. State the Chernoff-Hoeffding bounds.
18. Write the bias-variance decomposition for a generic learning algorithm A and associate the resulting components to overfitting and underfitting.

19. Write the upper bound on the estimation error of ERM run on a finite class $\mathcal{H}$ of predictors.
20. Write the upper bound on the estimation error of ERM run on a the class of complete binary tree predictors with at most N nodes on d binary features.
21. Write the bound on the difference between risk and training error for an arbitrary complete binary tree classifier h on d binary features in terms of its number N_h of nodes. Bonus points if you provide a short explanation on how this bound is obtained.
22. Write the formula for the K-fold cross validation estimate. Explain the main quantities occurring in the formula.
23. Write the pseudo-code for computing the nested cross validation estimate.
24. Write the mathematical definition of consistency for an algorithm A.
25. Write the statement of the no-free-lunch theorem.
26. Write the mathematical definition of nonparametric learning algorithm. Define the main quantities occurring in the formula.
27. Name one nonparametric learning algorithm and one parametric learning algorithm.
28. Write the mathematical conditions on k ensuring consistency for the k-NN algorithm.
29. Write the formula for the Lipschitz condition in a binary classification problem. Define the main quantities occurring in the formula.
30. Write the rate at which the risk of a consistent learning algorithm for binary classification vanish as a function of the training set size m and the dimension d under Lipschitz assumptions.
31. Explain the curse of dimensionality.
32. Write the bound on the risk of the 1-NN binary classifier under Lipschitz assumptions.
33. Can the ERM over linear classifiers be computed efficiently? Can it be approximated efficiently? Motivate your answers.
34. Write the system of linear inequalities stating the condition of linear separability for a training set in binary classfication.
35. Write the pseudo-code for the Perceptron algorithm.
36. Write the statement of the Perceptron convergence theorem.
37. Write the closed-form formula (i.e., not the argmin definition) for the Ridge Regression predictor. Define the main quantities occurring in the formula.
38. Write the pseudo-code for the projected online gradient descent algorithm.
39. Write the upper bound on the regret of projected online gradient descent on convex functions. Define the main quantities occurring in the bound.

40. Write the upper bound on the regret of online gradient descent on σ-strongly convex functions. Define the main quantities occurring in the bound.

41. Write the formula for the hinge loss.

42. Write the mistake bound for the Perceptron run on an arbitrary data stream for binary classification. Define the main quantities occurring in the bound.

43. Write the formula for the polynomial kernel of degree n.

44. Write the formula for the Gaussian kernel with parameter γ.

45. Write the pseudo-code for the kernel Perceptron algorithm.

46. Write the mathematical definition of the linear space $\mathcal{H}_K$ of functions induced by a kernel K.

47. Let f be an element of the linear space $\mathcal{H}_K$ induced by a kernel K. Write $f(\boldsymbol{x})$ in terms of K.

48. Write the mistake bound of the Perceptron convergence theorem when the Perceptron is run with a kernel K. Define the main quantities occurring in the bound.

49. Write the mistake bound for the kernel Perceptron run on an arbitrary data stream for binary classification. Define the main quantities occurring in the bound.

50. Write the closed-form formula (i.e., not the argmin definition) of the kernel version of the Ridge Regression predictor.

51. Write the convex optimization problem with linear constraints that defines the SVM hyperplane in the linearly separable case.

52. Write the unconstrained optimization problem whose solution defines the SVM hyperplane when the training set is not necessarily linearly separable.

53. Write the bound on the expected value of the SVM objective function achieved by Pegasos. Provide also a bound on the expected squared norm of the loss gradient.

54. Write the definition of ε-stability for a learning algorithm.

55. Write the value of ε for which SVM is known to be stable. The value depends on the radius X of the ball where the training datapoints live, the training set size m, and the regularization coefficient λ.

56. Write the mathematical conditions on the regularization coefficient λ ensuring consistency for the SVM algorithm wih Gaussian kernel.

57. Consider the class $\mathcal{F}_d$ of all functions of the form $f : \{-1,1\}^d \to \{-1,1\}$. Let $\mathcal{F}_{G,\mathrm{sgn}}$ be the class of functions computed by a feedforward neural networks with the sgn activation function and graph $G = (V,E)$. Provide asymptotic upper and lower bounds on $|V|$ such that $\mathcal{F}_d \subseteq \mathcal{F}_{G,\mathrm{sgn}}$.

58. Define a class of neural networks for which the ERM problem with the square loss is probably NP-hard.
59. Write the update line of the stochastic gradient descent algorithm. Explain the main quantities.
60. Write the definition of logistic loss for logistic regression with linear models.
61. Write the definition of consistency for surrogate losses.
62. Write a sufficient condition for consistency of a surrogate loss.
63. Write the formula for Bayes optimal predictor and Bayes risk for the logistic loss.