Université Nice Sophia-Antipolis
Statistical leaarning, 2020-2021
https://math.unice.fr/~rubentha/cours.html

Final examination, $(\mathrm{MPA} \cup \mathrm{MATHMODS})^{\complement}$, C

Documents and calculators forbidden. Send answers by e-mail, without justification (this is a quiz). One answer per question.

Give back the subject with your copy. Two points for a correct answer (zero point otherwise).

(1) What is the purpose of performing cross-validation?
   (a) To assess the predictive performance of the models .
   (b) To judge how the trained model performs outside the sample on test data.
   (c) Both 1 and 2.

(2) What do you mean by generalization error in terms of the SVM?
   (a) The threshold amount of error in an SVM
   (b) How accurately the SVM can predict outcomes for unseen data.
   (c) How far the hyperplane is from the support vectors.

(3) We do a linear regression. Which of the following is true in the context of regularization/penalization?
   (a) Ridge regression can reduce the parameters to zero.
   (b) Ridge regression is also called $L1$ regularization.
   (c) Lasso regression can reduce the parameters to zero.
   (d) None of the above.

(4) Which of the following is NOT supervised learning?
   (a) Principal Component Analysis.
   (b) Naïve Bayes Classifier.
   (c) Linear Regression.

(5) It is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. It is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. This clustering only works with numeric data. Which clustering has been discussed here?
   (a) K-nearest neighbours
   (b) K-clustering
   (c) K-means

(6) To find the minimum or the maximum of a function, we set the gradient to zero because
   (a) The value of the gradient at extrema of a function is always zero.
   (b) Depends on the type of problem.
   (c) Both (a) and (b).

(7) What does the naïve Bayes classifier assume?
   (a) Some features are independent of each other.
   (b) No features are independent of each other.
   (c) Most features are independent of each other.
   (d) All features are independent of each other.

(8) Assuming that you have a very large training set, which of the following algorithms do you think can be parallelized using map-reduce and splitting the training set across different machines?
   (a) A neural network trained using batch gradient descent.
   (b) Linear regression trained using batch gradient descent.
   (c) Logistic regression trained using stochastic gradient descent.

(9) Which of the following statements about regularization is correct? (We add a penalization/regularization term multiplied by λ.)
   (a) Using too large a value of lambda can cause your model to under-fit the data.

(b) Using too large a value of lambda can cause your model to over-fit the data.
(c) Using a very large value of lambda cannot hurt the performance of your model.

(10) Imagine, you are solving a classification problems with highly imbalanced class. The majority class is observed 99% of times in the training data. Your model has 97% accuracy after taking the predictions on test data. Which of the following is true in such a case?
(a) My model is good.
(b) My model is bad.
(c) Both (a) and (b).