

# Machine Learning Exam

Mario Martin

June 21, 2022

## Instructions:

1. **Enter your name** in the box at the beginning of the quiz.
2. You have to open this file using Acrobat Reader in order to fill the questionnaire. Click on ‘**Begin Quiz**’ to initialize the quiz. When finished, click on ‘**End Quiz**’ **and save the file**. Opening the saved file at any moment **should** maintain your previous answers (try this in your software before starting to work hard on the quiz). You can send me this file with answers filled (check always, before sending the file, that your answers are there!). *I recommend you always have a copy on paper of your solutions because errors can happen at any moment and you could lose all the work done.*

Alternatively, you can edit the PDF file using your favourite PDF editor, but make sure that the answers you mark are visible.

If all else fails, you can send me a file containing for each number of question a list of the options checked.

3. Each question in a box has a value of 0.6 points. Each option rightly guessed in a quiz question has a value of 0.14 points. Each quiz question has an unknown number of correct answers and each **incorrect answer will discount** accordingly to the probability of choosing it randomly so the expectation of a randomized answer is close to 0. Choose carefully.

Enter your name here:

Begin Quiz Answer each of the following.

1. Explain with your own (few) words the following concepts in ML:

- Loss function
- Regularized error

2. Select the true sentences:

a, c, f

- (a) Based only on the error in the training set, there is no way to choose which model is better
- (b) Regularization is designed to correct for a lack of sufficient complexity to get a good model
- (c) Using a larger validation data set reduces the chances to select an overfitted model **true**
- (d) Complexity control is necessary in non-linear methods only because linear methods are simple enough
- (e) In machine learning, it is convenient to represent in one column categorical variables as positive correlative numbers
- (f) Cross-validation can be used for model selection **overfitting  
multicollinearity  
heteroscedasticity**

3. Cite three problems we can detect in the result of applying linear regression to a dataset for the prediction of a real number using Maximum Likelihood Estimation (MLE):

4. Select the true sentences:

e,

- (a) The bias and the variance are related concepts: they represent the distribution of errors in the train and test sets, respectively
- (b) In linear regressions, the mean squared error is always preferred for optimization in MLE, because it is the more theoretically sound
- (c) Linear regression assumes normally distributed input features with possible errors also normally distributed.
- (d) The likelihood is a function of the data sample for a given choice of parameters
- (e) L1-regularization produces sparsity, as opposed to L2-regularization
- (f) Full bayesian view of linear regression allows a better knowledge of uncertainties because it not assumes same variance for each point

5. Explain with your words the additional assumptions taken in Ridge Regression with respect to the Maximum Likelihood Estimation (MLE) in Linear regression

6. Select the true sentences:

- (a) The k-means algorithm converges to a global optimum as the number of iterations goes to infinity
- (b) We cannot estimate beforehand the k parameter for the k-means algorithm. We can only try several values of k to find the best one using cross-validation.
- (c) A Gaussian mixture model assumes that the data has been generated by some finite mixture of Gaussians
- (d) If all the components in a mixture of Gaussians have equal covariance matrices, then E-M behaves as a linear model
- (e) The E-M algorithm approximates the probability of belonging to a cluster as the proportion of data that ends up being assigned to that cluster
- (f) Maximization step in EM algorithm over a mixture of Gaussians consist in computing the parameter of each Gaussian in the mixture given an assignment of point to clusters
- (g) The k-means algorithm can be used to initialize a Gaussian mixture model

7. Select the true sentences: b, f

- (a) 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
- (b) The Bayes classifier is the best possible classifier when the prior and class-conditional distributions are known
- (c) The Naive-Bayes classifier does not make assumptions about the distribution of data for continuous variables, apart from class-conditional independence FALSE
- (d) Bayesian classifiers are called generative because one could in principle use them to generate new data, once the model has been fitted true
- (e) For normally distributed classes, equal prior probabilities yield linear discriminant functions
- (f) QDA will fail (numerical error) if there is an insufficient number of data points in some class
- (g) Logistic regression assumes Gaussianity of the output data given the fitted data
- (h) Logistic regression is a linear method that can be used to predict an arbitrary numerical quantity

8. Select the true sentences: c

- (a) In general, when training a classifier using the *k-nn algorithm* on an unbalanced training dataset, the best choice for *k* is to use high values. false
- (b) In order to use the *k-nn* method is enough to have a clean dataset without missing values and containing only numerical attributes. false
- (c) In the *k-nn* algorithm, the distance-weighted vote is more relevant when *k* is large than when *k* is low.

- (d) In general, the larger the value of  $k$ , the better the accuracy because we have more a more robust estimator.
- (e) Decision Trees will overfit the data easily if we don't apply regularization techniques **true**
- (f) To reduce overfitting of a Decision Tree we can increase the max depth of tree
- (g) To reduce overfitting of a Decision Tree we can set a threshold on the minimum information gain to split a node **true**

9. Assume you have 3 independent classifiers, each of them with an accuracy of 0.7. Compute which is the accuracy for the Majority Vote algorithm for those 3 classifiers.

10. Select the true sentences: **a, c,**

- (a) In bagging, one maximizes the diversity of the individual models by using different bootstrap resamples for each model
- (b) In a Random Forest (RF), every tree in the forest first chooses its own set of variables, which is then used for all operations needed to grow that tree
- (c) The OOB error in a RF is a valid estimation of generalization error
- (d) A Random Forest is "random" partly because the variables used in each decision tree are optimized amongst a randomly chosen subset

11. Select the true sentences: **a, c, e**

- (a) The backpropagation algorithm computes the partial derivatives of a given differentiable error function with respect to the network weights
- (b) The backpropagation algorithm converges to a global optimum of the error function, if given a sufficiently large number of iterations
- (c) The activation function for the output neurons is dictated by the nature of the target variable
- (d) The activation function for the output neurons can never be the identity function
- (e) The complexity in a neural network can be controlled by the judicious choice of the number of hidden neurons
- (f) Regularization makes little sense in neural networks, because regularization is only useful in linear models

**b,c,d, e, g**

12. Select the true sentences:

- (a) In SVMs, the Lagrange coefficients  $\alpha_i$  are positive or zero for the support vectors and negative for the rest (the non support vectors) **false**
- (b) The cost parameter (C) in a SVM acts as a regularizer of the solution
- (c) Increasing the value of C in a SVM, the number of training errors cannot increase
- (d) By choosing a valid kernel, we get an inner product in some Hilbert space, which is our new feature space
- (e) In order to kernelize a learning algorithm, this must be based on Euclidean distances or inner products of the data

- (f) The kernel function defines kernel matrices whose elements are always positive
- (g) The product of an odd number of kernel functions is a kernel function

End Quiz