

## Final examination (EIT digital), B

*Documents and calculators forbidden. Write the answers on the examination copy, without justification (this is a quiz). One answer per question.*

*Give back the subject with your copy (+0.5 points!). One point for a correct answer (zero point otherwise). Duration: 2h.*

- (1) You are working on a project that involves clustering together images of different dogs. You take image and identify it as your centroid image. What type machine learning algorithm are you using?
  - (a) K-nearest neighbour.
  - (b) Centroid reinforcement.
  - (c) Binary classification.
  - (d) K-means clustering.
- (2) When would you reduce dimensions in your data?
  - (a) When the data comes from sensors.
  - (b) When your data set is larger than 500 GB.
  - (c) When you are using a Linux machine.
  - (d) When you have a large set of features with similar characteristics.
- (3) What is overfitting?
  - (a) When you perform hyper-parameter tuning and performance degrades.
  - (b) When a predictive model is accurate but takes too long to run.
  - (c) When you apply a powerful deep learning algorithm to a simple machine learning problem.
  - (d) When the model learns specifics of the training data that can't be generalized to a larger data set.
- (4) What kind of table compares classifications predicted by the model with the actual class labels?
  - (a) Confusion matrix.
  - (b) Chaos table.
  - (c) Prediction plot.
  - (d) Residual plot.
- (5) What type of machine learning algorithm makes predictions when you have a set of input data and you know the possible responses?
  - (a) Deep learning.
  - (b) Supervised learning.
  - (c) Supervisory logic.
  - (d) Unsupervised learning.
- (6) Application of Machine learning is \_\_\_\_\_.
  - (a) E-mail filtering.
  - (b) Face recognition.
  - (c) Sentimental analysis.
  - (d) All of the above.
- (7) \_\_\_\_\_ is the machine learning algorithms that can be used with labeled data.
  - (a) Regression algorithms.
  - (b) Clustering algorithms.
  - (c) None of the above.
- (8) \_\_\_\_\_ is the machine learning algorithms that can be used with unlabeled data.
  - (a) Clustering algorithms.

- (b) Regression algorithms.
  - (c) Instance-based algorithms.
  - (d) All of the above.
- (9) \_\_\_\_\_ is a disadvantage of decision trees?
- (a) Decision trees are robust to outliers.
  - (b) Decision trees are prone to be overfit.
  - (c) Both A and B.
  - (d) None of the above.
- (10) Machine Learning can automate many tasks, especially the ones that only humans can perform with their innate intelligence.
- (a) False.
  - (b) True.
- (11) \_\_\_\_\_ looks at the relationship between predictors and your outcome.
- (a) Regression analysis.
  - (b) K-means clustering.
  - (c) Big data.
  - (d) Unsupervised learning.
- (12) What is principal component analysis?
- (a) A feature selection technique that adds or removes features to optimize prediction accuracy.
  - (b) A clustering technique that partitions data into mutually exclusive clusters.
  - (c) A linear feature transformation technique for reducing data dimensionality.
  - (d) A predictive technique that identifies a better set of parameters.
- (13) What is an example of a commercial application for a machine learning system?
- (a) A data entry system.
  - (b) A data warehouse system.
  - (c) A product recommendation system.
  - (d) A massive data repository.
- (14) You work for a power company that owns hundreds of thousands of electric meters. These meters are connected to the internet and transmit energy usage data in real-time. Your supervisor asks you to direct project to use machine learning to analyze this usage data. Why are machine learning algorithms ideal in this scenario?
- (a) The algorithms would help your organization see patterns of the data.
  - (b) The algorithms would help the meters access the internet.
  - (c) The algorithms will improve the wireless connectivity.
- (15) To predict a quantity value. use \_\_\_\_\_.
- (a) Regression.
  - (b) Clustering.
  - (c) Classification.
  - (d) Dimensionality reduction.
- (16) Why is naive Bayes called naive?
- (a) It naively assumes that you will have no data.
  - (b) It does not even try to create accurate predictions.
  - (c) It naively assumes that all the predictors depend on one another.
  - (d) It naively assumes that the predictors are independent from one another.
- (17) You work for an insurance company. Which machine learning project would add the most value for the company?
- (a) Create an artificial neural network that would host the company directory.
  - (b) Use machine learning to better predict risk.
  - (c) Create an algorithm that consolidates all of your Excel spreadsheets into one data lake.
  - (d) Use machine learning and big data to research salary requirements.
- (18) What is one reason not to use the same data for both your training set and your testing set?

- (a) You will almost certainly under-fit the model.
  - (b) You will almost certainly over-fit the model.
  - (c) You will pick the wrong algorithm.
  - (d) You might not have enough data for both.
- (19) What type of machine learning algorithm makes predictions when you have a set of input data and you know the possible responses?
- (a) Supervisory logic.
  - (b) Unsupervised learning.
  - (c) Supervised learning.
  - (d) Data prophecy.
- (20) Which of the following is an example of a deterministic algorithm?
- (a) K-means.
  - (b) PCA.