

# Questions

1. What is Machine learning?
  - a) statistics on large-scale data sets
  - b) any optimization algorithm
  - c) a computer program that writes other programs
  - d) automatic parameter fitting on a statistical model
  
2. What is a test dataset used for?
  - a) to estimate the error of a model
  - b) to estimate data quality
  - c) to estimate model quality in production
  - d) to estimate the business impact of a model
  
3. Which is **not** a good strategy against overfitting?
  - a) obtain more training data
  - b) tune regularization parameters
  - c) increase the number of input features
  - d) use a simpler model
  
4. Which is a classification method?
  - a) Logistic Regression
  - b) Gradient Descent
  - c) Cross-Validation
  - d) Principal Component Analysis
  
5. Which is an advantage of Logistic Regression?
  - a) It does not overfit
  - b) It can solve classification and regression problems
  - c) It calculates probabilities
  - d) It does not require feature engineering
  
6. Which is a disadvantage of a Random Forest model?
  - a) it requires manual feature selection
  - b) it is sensitive to transformation of the data
  - c) it cannot handle non-scalar data
  - d) it requires scaled data

7. Which is false about Explorative Data Analysis (EDA)?

- a) EDA means plotting the data in different ways
- b) EDA is usually done before training a model
- c) EDA can be fully replaced by machine learning
- d) EDA helps to build expert knowledge on your data

8. What is Principal Component Analysis (PCA) primarily used for?

- a) to reduce the number of features in data
- b) to reduce overfitting
- c) to fill gaps in data
- d) to build recommender systems

9. Feature Engineering **does not** improve:

- a) the accuracy of a machine learning model
- b) the quality of the training data
- c) the interpretability of a machine learning model
- d) the execution speed of a machine learning model

10. Which is **not** an unsupervised learning method?

- a) clustering
- b) anomaly detection
- c) Principal Component Analysis
- d) k-Nearest Neighbors

11. Which is true about Artificial Neural Networks?

- a) there exist pretrained models for complex data such as images and text
- b) they are generally better than "traditional" Machine Learning
- c) they require huge servers to train
- d) you need at least 100,000 data points

© 2022 Dr. Kristian Rother

Distributed under the conditions of the Creative Commons Attribution License 4.0. See <https://creativecommons.org/licenses/by/4.0/> for details.