2976933

Q1: Which of the following correctly describes predictive modeling?

1. Predictive models can only be derived from data.
2. Deep learning methods can build the best predictive models.
3. The best predictive model for a task always have the best predictive performance.
4. Predictive models can be constructed with supervised learning methods.

Correct answer:

1. Predictive models can be constructed with supervised learning methods.

Explanation:

A is incorrect. Predictive models can be derived from variety of sources including data and prior knowledge.

B is incorrect. The definition of best model differ from task to task. Deep learning methods can derive highly predictive models with large training data set, but the resulting models generally cannot be interpreted easily.

C is correct. The definition of best model differ from task to task. In some applications, one might prefer models that have lower predictive performance and lower deployment cost compared to models with higher predictive performance and higher deployment cost.

D is correct. Supervised learning methods are suitable for constructing predictive models.

Resource for additional study:

Duda RO, Hart PE, Stork DG. Pattern classification. John Wiley & Sons; 2012 Nov 9.

Statnikov, Alexander. *A gentle introduction to support vector machines in biomedicine: Theory and methods*. Vol. 1. world scientific, 2011.

Q2: Which of the following statement regarding predictive modeling is correct:

1. Data preprocessing is not a critical part of predictive modeling.
2. Cox modeling is a common method for modeling binary outcomes.
3. The variables that are predictive of the outcome are the direct causes of the outcome.
4. The Nested N-fold Cross Validation is an established state-of-the-art design for powerful model selection and unbiased error estimation.

Correct answer:

D. The Nested N-fold Cross Validation is an established state-of-the-art design for powerful model selection and unbiased error estimation.

Explanation:

A is incorrect. Data preprocessing is a critical part of predictive modeling.

B is incorrect. Cox modeling is a common method for modeling time-to-event outcome.

C is incorrect. Typically predictive models optimize productivity and do not guarantee causal interpretation, unless causal methods are used to derive the predictive model.

D is correct.

Resource for additional study:

Duda RO, Hart PE, Stork DG. Pattern classification. John Wiley & Sons; 2012 Nov 9.

Q3: Which of the following statement regarding modeling binary outcome is correct:

1. Decision trees can only be used to model binary outcomes.
2. Accuracy is a performance metric that is not sensitive to prevalence of positive vs. negative cases.
3. False positives are as bad as false negatives.
4. Sensitivity is defined as number of true positives divided by number of positive cases.

Correct answer:

D.Sensitivity is defined as number of true positives divided by number of positive cases.

Explanation:

A is incorrect. Decision trees can also be used to model continues outcomes (i.e. regression tree).

B is incorrect. Accuracy is sensitive to prevalence.

C is incorrect. The value of false positives and false negatives depends on the application/task.

D is correct.

Resource for additional study:

Duda RO, Hart PE, Stork DG. Pattern classification. John Wiley & Sons; 2012 Nov 9.

Q4: Which of the following statement regarding supervised learning methods is correct:

1. Supervised learning methods do not require labeled data.
2. The quality of data do not affect the performance of supervised learning methods.
3. Support vector machine is a supervised learning method.
4. Hierarchical clustering is a supervised learning method.

Correct answer:

1. Support vector machine is a supervised learning method.

Explanation:

A is incorrect. Supervised learning requires labeled data, unsupervised learning methods do not require labeled data.

B is incorrect. The quality of data can affect the performance.

C is correct.

D is incorrect. Hierarchical clustering is an unsupervised learning method.

Resource for additional study:

Duda RO, Hart PE, Stork DG. Pattern classification. John Wiley & Sons; 2012 Nov 9.

Q5: Which of the following statement regarding feature selection is correct:

1. Feature selection can only be done with data driven methods.
2. The feature set that includes the Markov boundary of the target will always result in models with optimal predictive performance.
3. The number of possible feature sets can be constructed from p features is 2p.
4. Overfitting is inevitable without explicit feature selection.

Correct answer:

1. The number of possible feature set can be constructed from p features is 2p

Explanation:

A is incorrect. Feature selection can also done according to domain knowledge.

B is incorrect. The classifier also contributes to whether optimal predictive performance can be achieved.

C is correct. A feature is either selected or not in to a feature set (i.e. 2 options), and there are p features, so overall it makes 2­p.

D is incorrect. Overfitting can also be handled with regularization or ensembling.

Resource for additional study:

Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of machine learning research. 2003;3(Mar):1157-82.

Q6: Which of the following statement regarding feature selection is correct:

1. Recursive feature elimination do not have hyper-parameters.
2. Lasso regression do not have hyper-parameters.
3. Causal feature selection methods usually result in a smaller feature set compared to non-causal feature selection methods.
4. Causal feature selection methods always return the Markov boundary of the target.

Correct answer:

1. Causal feature selection methods usually result in a smaller feature set compared to non-causal feature selection methods.

Explanation:

A is incorrect. One needs to specify the reduction rate for recursive feature selection.

B is incorrect. One needs to specify the lambda parameter which corresponds to the trade-off between the error and the parsimony of the model.

C is correct.

D is incorrect. If the assumption of the method is not met, or sample size is not sufficient, causal feature selection methods may not return the Markov boundary of the target. Also, some causal feature selection methods are designed to return the parents and children of the target.

Resource for additional study:

Guyon I, Elisseeff A. An introduction to variable and feature selection. Journal of machine learning research. 2003;3(Mar):1157-82.

Guyon I, Aliferis C. Causal feature selection. InComputational methods of feature selection 2007 Oct 29 (pp. 75-97). Chapman and Hall/CRC.

Q7: Which of the following statement regarding causality is correct:

1. If X and Y are correlated, either X causes Y, or Y causes X.
2. If X and Y are correlated, and X happens before Y, then X causes Y.
3. Experimentation is the only way to determine causation.
4. Causation can be discovered to a certain degree with observational data.

Correct answer:

1. Causation can be discovered to a certain degree with observational data.

Explanation:

A is incorrect. Correlation do not imply causation.

B is incorrect. There could be a variable Z that cause both X and Y.

C is incorrect. Causation can be discovered to a certain degree with observational data.

D is correct.

Resource for additional study:

Spirtes P, Glymour CN, Scheines R, Heckerman D, Meek C, Cooper G, Richardson T. Causation, prediction, and search. MIT press; 2000.

Q8: Which of the following statement regarding causal feature selection is correct:

1. PC algorithm is a causal feature selection algorithm.
2. PC algorithm could be modified and used as a causal feature selection algorithm.
3. PC algorithm is based on optimizing the BIC score.
4. PC algorithm return the correct causal graph all the time.

Correct answer:

B. PC algorithm could be modified and used as a causal feature selection algorithm.

Explanation:

A is incorrect. PC algorithm is a causal structure learning algorithm.

B is correct.

C is incorrect. PC algorithm is based on conditional independency test.

D is incorrect. PC algorithm returns the correct causal graph when its assumptions are met.

Resource for additional study:

Spirtes P, Glymour CN, Scheines R, Heckerman D, Meek C, Cooper G, Richardson T. Causation, prediction, and search. MIT press; 2000.

Q9: Which of the following statement regarding biomedical data is correct:

1. Gene expression data do not require preprocessing.
2. Missing values are rare.
3. EHR data is not suitable for knowledge discovery.
4. Supervised learning methods can help construct diagnosis tests.

Correct answer:

D. Supervised learning methods can help construct diagnosis tests.

Explanation:

A is incorrect. Gene expression data requires preprocessing.

B is incorrect. Missing values are common with certain type of biomedical data such as the EHR data.

C is incorrect. EHR data can be leveraged for knowledge discovery.

D is correct. Supervised learning methods can help construct diagnosis tests.

Resource for additional study:

Statnikov, Alexander. *A gentle introduction to support vector machines in biomedicine: Theory and methods*. Vol. 1. world scientific, 2011.

Q10: Which of the following statement regarding time-to-event outcome is correct:

1. Cox model is the only method for modeling time-to-event outcome.
2. Censored observations do not require special treatment.
3. Accuracy is a metric for evaluating model performance for time-to-event models.
4. The change of the value of a feature over time can be incorporated into time-to-event models.

Correct answer:

D. The change of the value of a feature over time can be incorporated into time-to-event models.

Explanation:

A is incorrect. There are several alternatives, including accelerated failure time model.

B is incorrect. Proper treatment of censored observations are critical for time-to-event models.

C is incorrect. Accuracy is a metric for evaluating model performance when the outcome is binary.

D is correct. The change of the value of a feature over time can be incorporated as time varying covariate.

Resource for additional study:

Cox DR. Analysis of survival data. Routledge; 2018 Feb 19.

Q11: Which of the following statement regarding causal feature selection is correct:

1. Causal feature selection only applies to modeling binary outcomes.
2. Recursive feature elimination is a causal feature selection method.
3. Different data distributions require different conditional independency tests.
4. Causal feature selection methods miss variables that have nonlinear relationships with the target.

Correct answer:

C. Different data distributions require different conditional independency tests.

Explanation:

A is incorrect. With appropriate statistical tests, causal feature selection can select features for outcomes from various distributions.

B is incorrect. Recursive feature selection is not a causal feature selection method.

C is correct.

D is incorrect. With appropriate statistical test, causal feature selection can detect nonlinear relationships.

Resource for additional study:

Guyon I, Aliferis C. Causal feature selection. InComputational methods of feature selection 2007 Oct 29 (pp. 75-97). Chapman and Hall/CRC.

Q12: Which of the following statement regarding causal feature selection is correct:

1. Model built with features selected by causal feature selection do not need validation.
2. Causal feature selection methods always outperform non-causal feature selection methods.
3. Causal feature selection methods require a lot of computational resources.
4. Causal feature selection methods are powerful for predictive tasks and they should be considered when building predictive models.

Correct answer:

D. Causal feature selection methods are powerful for predictive tasks and they should be considered when building predictive models.

Explanation:

A is incorrect. Predictive models should be validated.

B is incorrect. Causal feature selection methods do not always outperform non-causal feature selection methods.

C is incorrect. Good implementations of causal feature selection methods take computational efficiency into account.

D is correct. And that is why you are here.

Resource for additional study:

Guyon I, Aliferis C. Causal feature selection. InComputational methods of feature selection 2007 Oct 29 (pp. 75-97). Chapman and Hall/CRC.