

Date: _____

Name: _____

1. (2 pts) Let the observation be described by 2 real-valued features x and y . Let's introduce a new feature z , defined as $z = 2x + y$. Each observation may come from one of 2 classes. Which of the following is true? More than one may be correct.

- If the data are linearly separable in space spanned by x and y , they are also linearly separable in space spanned by x, y, z .
- If the data are linearly separable in space spanned by x, y, z , they are also linearly separable in space spanned by x, y .

2. (4 pts) Which of the following holds for a decision tree: (more than one may be correct)

- Classification of a new observation is done by passing the observation through the tree from root to certain leaf node. The resulting class is determined by the majority of training instances belonging to the final leaf node.
- There are infinitely many decision trees which are equivalent regarding the classification of certain data set.
- The depth (or height, or number of levels) of the tree is not directly dependent on the number of input attributes.
- Number of leaves is equal to the number of classes to which we classify.

3. (3 pts) What is the main difference between regression and classification?

4. (3 pts) Choose all correct answers. If we perform basis expansion (feature space straightening), we

- select a subset of input features which most influence the dependent variable.
- derive new features hoping that at least some of them will be more useful for prediction than the original features.
- increase the dimension of the input space, and thus also the risk of model overfitting.

5. (3 pts) Choose all correct answers. Feature selection of type *filter*:

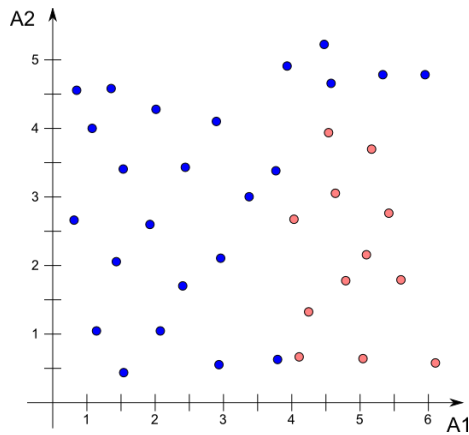
- There is a single particular method of feature filtering that provides the best results.
- Sorting features according the strength of their relation with the dependent variable is used very often.
- It can be classified as a data preprocessing method since it is not dependent on the used (classification or regression) model.

6. (2 pts) Predict the value of continuous dependent variable y for observation $x=5.2$ using the method of 3 nearest neighbors. The training *multiset* $\{(x,y)\}$ is:

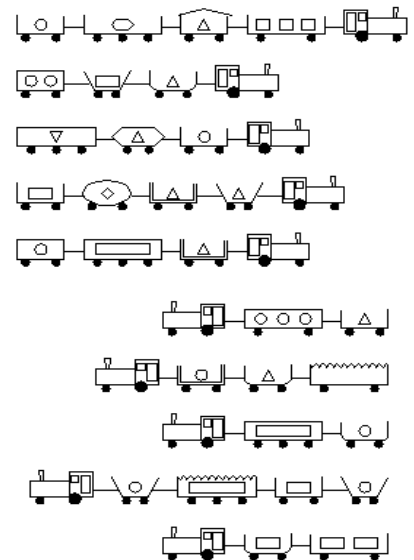
$\{(2, 1.0), (3, 5.0), (3, 4.0), (4, 4.0), (5, 1.0), (6, 1.0), (6, 2.0), (7, 6.0)\}$

- 1
- 4/3
- 2
- 7/3

7. (4 pts) Draw a decision tree that classifies the data in the picture without any error. In the picture, draw the splits of the space done by individual tree nodes. Explain why you chose feature A1 or A2 as the one used for the test in the tree root.



8. (3 pts) Consider 10 data examples in the picture and the association rule:
train has 3 wagons => train goes to east.
 Compute the support and confidence of this association rule.



(Association rules are not taught in this course for several years now; of course, questions related to them will not be part of the exam.)

9. (4 pts) Explain the terms “false positives” and “true negatives”.

10. (2 pts) Check all correct answers. We can say about boosting:

- It builds many weak models and by combining them it creates a stronger model.
- Individual weak models are independent of each other, thus can be created in parallel in such a way that each is trained on a different subset of training data.
- Individual weak models are not independent of each other, they must be created sequentially.
- All training examples have equal weights during the whole learning process.