

Homework 0

1. Definitions and terms: the first column gives the definition of terms; the second column gives the terms. Please write down the corresponding term for each definition.

1.1 The rate of correct predictions made by the model over a data set. (k)	a. Unsupervised learning
1.2 The rate of incorrect predictions made by the model over a data set. ()	b. Association learning
1.3 learned rules that uncover relationships and structure in data. ()	c. Missing value
1.4 A quantity describing an instance (single observation, perhaps multidimensional) of data. ()	d. Categorical data
1.5 A finite number of discrete values. ()	e. Error rate
1.6 Categorical data without an implied ordering. ()	f. Supervised learning
1.7 Categorical data with an explicit ordering. ()	g. Continuous data
1.8 Data that lives within an interval of the real number line. ()	h. Regressor
1.9 A mapping from unlabeled instances to (discrete) classes. ()	i. Specificity
1.10 The proportion of data which is correctly classified. ()	j. Cross-validation
1.11 A method for estimating the accuracy (or error) of a method. ()	k. Accuracy
1.12 The process of improving the quality of the data by modifying its form or content, for example by removing or correcting data values that are incorrect. ()	l. Response Variable
1.13 A set of independent and identically distributed instances. ()	m. Sensitivity
1.14 The value for an attribute is not known or does not exist. ()	n. Predictive coverage
1.15 Independent data which is related to matched dependent data. ()	o. Nominal data
1.16 The "explained" data which depend on a regressor. ()	p. i.i.d. sample
1.17 True positive rate. ()	q. Data cleaning /cleansing

1.18 True negative rate. ()	r. Classifier
1.19 Techniques used to learn the relationship between independent attributes and a corresponding dependent attribute (a labeling). ()	s. Ordinal data
1.20 Learning techniques that group instances without a pre-specified dependent attribute. ()	t. Attribute (field, variable, feature)

2. Statistical coverage is defined as the proportion of intervals (confidence, credible, etc.) that cover the true parameter. Is this the same as Predictive coverage, as defined above? Why or why not?

3. Consider the 2×2 matrix:

actual \ predicted	-	+
-	a	b
+	c	d

Define the following terms mathematically using the 2×2 confusion matrix above.

For example: Accuracy $\frac{a+d}{a+b+c+d}$

True positive rate (Recall, Sensitivity) _____

True negative rate (Specificity) _____

Precision _____

False positive rate _____

False negative rate _____