Glossary

1 Classification And Machine Learning Glossary

Attribute Machine learning term,
Synonym: *feature.*
See Also: Feature, Feature Space.

Bayes Decision Rule The function $f^* : \mathcal{X} \xrightarrow{f^*} \mathcal{Y}$ that assigns to each feature vector $\mathbf{x} \in \mathcal{X}$ a label $y \in \mathcal{Y}$ in a way that minimizes the probability of classification error (error rate).
The Bayes Decision Rule is not necessarily unique: if two or more classes are equiprobable on a subset of $\mathcal{X}$ having non-zero probability, there are multiple labeling functions that minimize the error rate.
The goal of supervised learning is to select the best approximation of the Bayes Decision Rule among the functions of the Hypothesis Space.
The Bayes Decision Rule is called *Target Function* in the machine learning literature.
See Also: Bayes Risk, Error Rate, Hypothesis Space, Risk, Target Function.

Bayes Risk The error rate of the Bayes Decision Rule, namely, its risk with respect to the 0-1 loss function.
See Also: Bayes Decision Rule, Loss Function, Risk.

Classification Function A mapping from $\mathcal{X}$ to $\mathcal{Y}$, namely, a function assigning a label to each feature vector in the feature space.
Synonym: *Labeling Function.*
See Also: Classifier, Classification Rule.

Classification Rule A sequence $\{f_n(x \mid (\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n))\}$ of classifiers (where by classifier here we mean random functions).
See Also: Classifier, Classification Function.

Classifier We use the term *classifier* to denote three related entities; in general, the context will determine what meaning should be assigned to the word. In general, we will use the first definition of the term.

1. A classifier is a random function $f_n(x \mid (\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n))$, where $\mathcal{X} \xrightarrow{f_n} \mathcal{Y}$;
In other words, a classifier is a mapping from the sample space supporting training sets of size n to a class of functions $\mathcal{F}_n$ (called *hypothesis space* in the machine learning literature) mapping the feature space $\mathcal{X}$ to the set of class labels $\mathcal{Y}$.
In this sense **a classifier can be thought of as a pair** $(\mathcal{F}_n, \mathcal{L}_n)$**,** where $\mathcal{F}_n$ is a *hypothesis space*, and $\mathcal{L}_n$ is a learning algorithm, that selects f_n from $\mathcal{F}_n$ using $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n))$.
Synonym: *Learning Algorithm + Hypothesis Space.*

2. Any function $f_n(x) \in \mathcal{F}_n$, namely, any mapping between the feature space $\mathcal{X}$ and the set of class labels $\mathcal{Y}$. Hence, any instance of a classifier, defined as a random function as above, will also be called a classifier.
   Synonym: *Hypothesis*.
3. the most general sense a classifier is also any measurable function from $\mathcal{X}$ to $\mathcal{Y}$.
   Synonyms: *Concept*, *Classification Function*.

See Also: Classification Rule, Hypothesis, Classification Function, Concept, Concept Space, Learning Algorithm.

Concept Term used in the machine learning literature. A member of the Concept Space, defined below.
Synonym: *Classification Function*, *Classifier*.
See Also: Target Concept.

Concept Space Term used in the machine learning literature. It denotes either a space $\mathcal{G}$ to which the Bayes Decision Rule (target function) belongs, or a space $\mathcal{F}$ to which the regression function belongs, depending on the context.
See Also: Bayes Decision Rule, Regression Function.

Cost Function See Loss Function.

Discrimination Rule Synonym: *Classification Rule*.

Error Rate The probability that a classifier incorrectly labels an observation $\mathbf{X}$.
Synonym: *Risk* (when using the 0-1 loss function).
See Also: Loss Function, Risk.

Feature A numeric or categorical quantity used as input to a classifier (e.g.: "the length of an object").
Synonym: *attribute*.
Rarely: the attribute-value specification of a numeric or categorical quantity used as input to a classifier (e.g.: length = 10).
See Also: Attribute, Feature Space, Feature Vector.

Feature Space The set $\mathcal{X}$ of feature vectors $\mathbf{x}$ that can be used as input to a classifier.

Feature Vector A vector of features, denoted by $\mathbf{x}$. In general, a classification function is a function defined on feature vectors and taking values in a set of class labels. set $\mathcal{Y}$.

Hypothesis: concept (i.e., classification function) belonging to the Hypothesis Space of a learning algorithm (see Hypothesis Space below).

Hypothesis Space Term used in the machine learning literature. It denotes the space $\mathcal{F}_n$ of classifiers, or the space $\mathcal{G}_n$ of conditional probabilities, from which the learning algorithm selects a hypothesis.

Learning Algorithm An algorithm to select a hypothesis (classification function) from the hypothesis space using a training set.
Synonyms: *Training Algorithm*, *Classifier*.
See Also: Classifier, Classification Function, Hypothesis Space, Training Set, Hypothesis Space.

Loss Function

1. In the classification context, a function $l(\hat{y}, y) : \mathcal{Y} \times \mathcal{Y} \stackrel{l(\cdot,\cdot)}{\rightarrow} \mathbb{R}^+$, described the cost of assigning the label $\hat{y}$ to a sample of class y.
2. In the density estimation/regression estimation context, a function $l(\hat{y}, y)$: $\mathbb{R} \times \mathbb{R} \stackrel{l(\cdot,\cdot)}{\rightarrow} \mathbb{R}^+$, described the cost of assigning the value $\hat{y}$ to the density/regression function evaluated at $\mathbf{x}$, where it takes value y.

The simplest loss function for classification is the 0-1 loss, having value 0 iff (i.e., if and only if) $y == \hat{y}$, and value 1 otherwise.
The typical loss function for regression estimation is the squared error.
The expected value of the loss is called *Risk*.
Synonym: *Cost Function*, term less frequently used in the classification literature.
Related quantities are the *Utility Function* and the *Payoff Function*, used in optimization.
See Also: Bayes Risk, Error Rate, Risk.

Pruning Set A collection $(\mathbf{X}''_1, Y_1), \ldots, (\mathbf{X}''_p, Y_p)$, used to prune a tree classifier.
See Also: Training Set, Test Set.

Regression Function In the statistical classification literature, the term is used to denote $g_y(\mathbf{x}) = p(y \mid \mathbf{x})$, the conditional probability of the class label y given the observation $\mathbf{x}$, considered AS A FUNCTION of $\mathbf{x}$.
In the machine learning literature, the regression function called target function.
See Also: Concept Space, Target Function.

Risk Given a loss function $l(\hat(Y)(\mathbf{X}), Y)$, the risk R is the expected value of the loss with respect to the probability measure generating the pair $(\mathbf{X}, Y)$:

$$R = E_{P(\mathbf{X},Y)} l\left(\hat(Y)(\mathbf{X}), Y\right).$$

If the loss function is the 0-1 loss, then Risk is synonym of *error rate*.
See Also: Bayes Risk, Loss function, Error Rate.

Target Concept Term used in the machine learning literature to denote the Bayes decision rule, or the regression function, depending on the context. The target concept is a member of the concept space.
Synonyms: *Bayes Decision Rule* in classification, *Regression Function* in regression.
See Also: Bayes Decision Rule, Concept Space, Regression Function.

Target Function Synonym: *Target Concept.*

Test Set A collection $(\mathbf{X}'_1, Y_1), \ldots, (\mathbf{X}'_m, Y_m)$, used to validate the performance of a classifier.

See Also: Training Set, Pruning Set

Training Set A collection $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$, where the $(\mathbf{X}_i, Y_i)$ are labeled examples (very commonly called labeled samples), used for training.

A few authors use the term *Training Sample* to denote the training set, which is confusing since the examples are commonly called samples.

See Also: Training Set, Test Set

2 A Short Probability Glossary

Algebra An algebra of sets is a collection Σ of sets which contains the sure event, is closed with respect to the operations of complement and of finite union:

- $\Omega \in \Sigma$;
- $F \in \Sigma \quad \Rightarrow \quad F^c \in \Sigma$;
- $F \in \Sigma,\ G \in \Sigma \quad \Rightarrow \quad F \bigcup G \in \Sigma$.

See Also: σ-algebra, Sample Space, Measurable Space, Measure Space, Probability Space.

Borel Set A Borel Set is an element of a Borel σ-algebra. (Sorry, a lame definition ...)
See Also: Algebra, Borel σ-algebra, Measurable Space, Measure, Measure Space, Probability Measure, Probability Space, Sample Space, σ-algebra.

Borel Set If Ω is a topological space, the Borel σ-algebra $\mathcal{B}(\Omega)$, on Ω is the σ-algebra generated by the family of open subsets of Ω. [1]
See Also: Algebra, Borel Set, Measurable Space, Measure, Measure Space, Probability Measure, Probability Space, Sample Space, σ-algebra.

Measurable Space A measurable space is a pair $(\Omega, \mathcal{F})$ where Ω is a sample space (the collection of the possible outcomes of an experiment) and $\mathcal{F}$ is a σ-algebra of subsets of Ω.
See Also: Algebra, Borel Set, Borel σ-algebra, Measure, Measurable Space, Measure Space, Probability Measure, Probability Space, Sample Space, σ-algebra.

Measure A measure μ is a countably additive set function from an algebra Σ to $\mathbb{R}^+$, namely, it satisfies: [2]

- $\mu(\emptyset) = 0$,
- $\mu(F \bigcup G) = \mu(F) + \mu(G), \ \ \forall F, G \in \Sigma | F \bigcap G = \emptyset$,
- if $\{S_n\}$ is a sequence of sets in Σ, and $\bigcup_n S_n \in \Sigma$ (Note: Σ is an algebra, not a σ-algebra!), such that the sets S_n are disjoint, then

$$f(\bigcup_n S_n) = \sum_n f(S_n).$$

If $\mu(F)$ is not constrained to be nonnegative, it is called a *signed measure*. Rejoice: we will not use signed measures.
See Also: Algebra, Borel Set, Borel σ-algebra, Measurable Space, Measure, Measure Space, Probability Measure, Probability Space, Sample Space, σ-algebra.

[1] For instance, every subset of $\mathbb{R}$ that we commonly use is a Borel set, One cannot construct non-Borel sets $\in \mathbb{R}$ without the help of the of the axiom of choice. All the sets we will see are Borel Sets.

[2] The first and second condition without the third condition define an additive set function.

Measure Space A measure space is a triple $(\Omega, \mathcal{F}, \mu)$, where $(\Omega, \mathcal{F})$ is a measurable space and μ is a measure.

See Also: Algebra, Borel Set, Borel σ-algebra,Measurable Space, Measure, Measure Space, Probability Measure, Probability Space, Sample Space, σ-algebra.

Probability Measure A probability measure is a measure satisfying the additional constraint that $\mu(\Omega) = 1$.

See Also: Algebra, Borel Set, Borel σ-algebra, Measurable Space, Measure, Measure Space, Probability Measure, Probability Space, Sample Space, σ-algebra.

Random Variable A random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is an $\mathcal{F}-$measurable function.

Categorical Random Variable A numerical random variable is a random variable having categorical range. Note that the only natural metrics to measure the distance between two categorical values is the zero-one metric (d(X,Y) = 1 if X$\neq$ Y, = 0 if X=Y), although other metrics can be constructed. Categorical Random Variables have

Numerical Random Variable A numerical random variable is a random variable having numerical range. The numerical range can be the real line, the complex plane, d-dimensional Euclidean spaces, etc. A variety of metrics can be used to measure the distance between numerical random variables. Numerical random variables have

See Also: measurable function, probability law, expectation

Sample Space The sample space Ω, also called the *sure event*, is defined, for the purpose of this course, as the collection of all possible **outcomes** of an experiment. [3]

See Also: Algebra, Borel Set, Borel σ-algebra, Measurable Space, Measure, Measure Space, Probability Measure, Probability Space, Sample Space, σ-algebra.

σ-algebra A σ-algebra is a collection $\mathcal{F}$ of sets which contains the sure event, is closed with respect to the operations of complement and of countable union:

See Also: algebra, Sample Space, Measurable Space, Measure Space, Probability Space.

- $\Omega \in \mathcal{F}$;
- $F \in \mathcal{F} \quad \Rightarrow \quad F^c \in \mathcal{F}$;
- $F_1, F_2, \ldots \in \mathcal{F}, \quad \Rightarrow \quad \bigcup_i F_i \in \mathcal{F}$.

[3] Note that *possible* is not a probabilistic concept: an outcome is possible if it can occur, impossible if it cannot occur. For example, if the experiment is the measure of the voltage between two points of a circuit, Ω can be identified with the set of real numbers. Not knowing a priori what the circuit is, we cannot bound the maximum value of the voltage, so we will say that any real number is a possible outcome. However, the result of the experiment will not be a letter of the English alphabet, and no letter is a possible outcome of the experiment.

See Also: Algebra, Borel Set, Borel σ-algebra, Measurable Space, Measure, Measure Space, Probability Measure, Probability Space, Sample Space, σ-algebra.

Sure Event See: Sample Space.