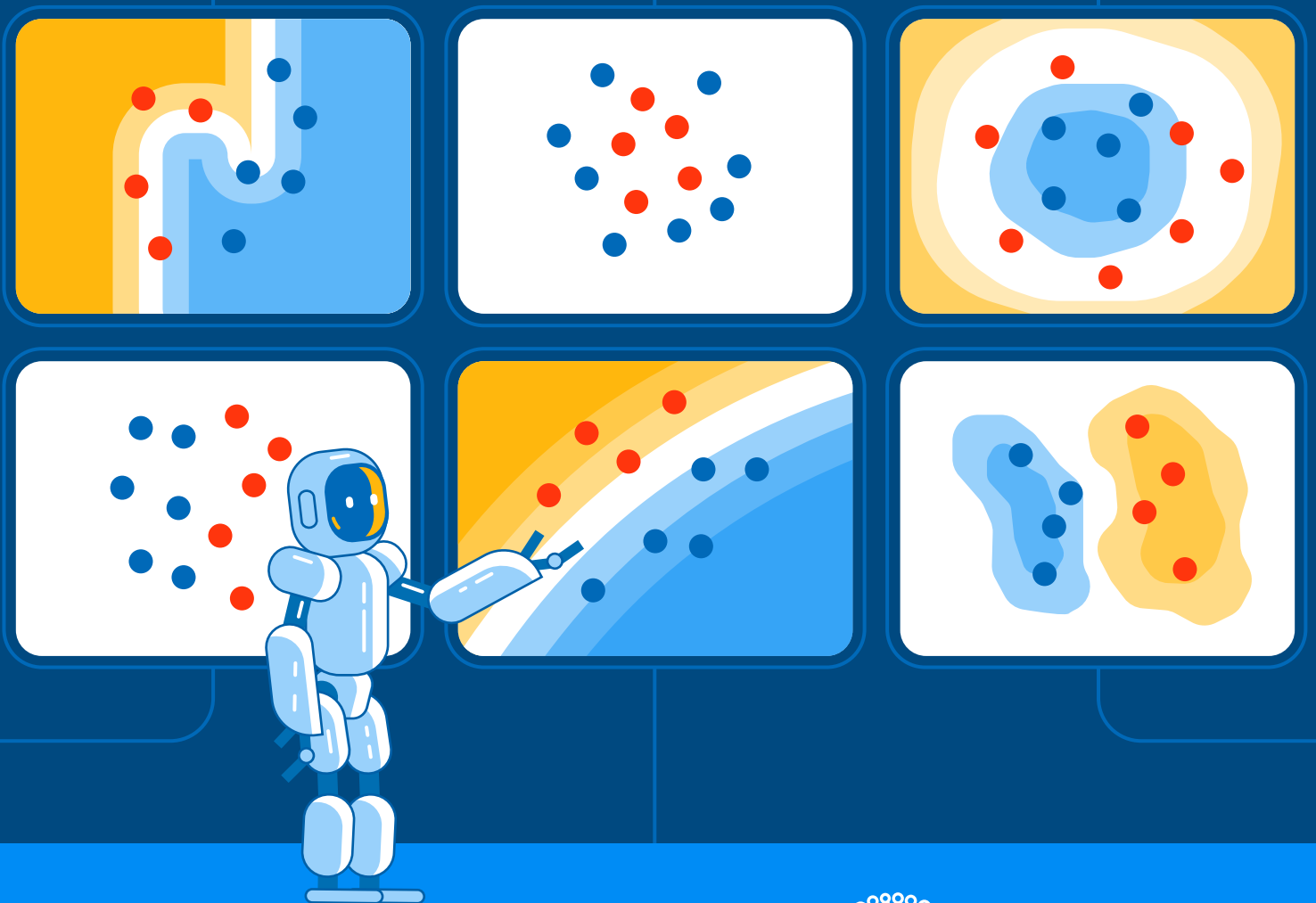
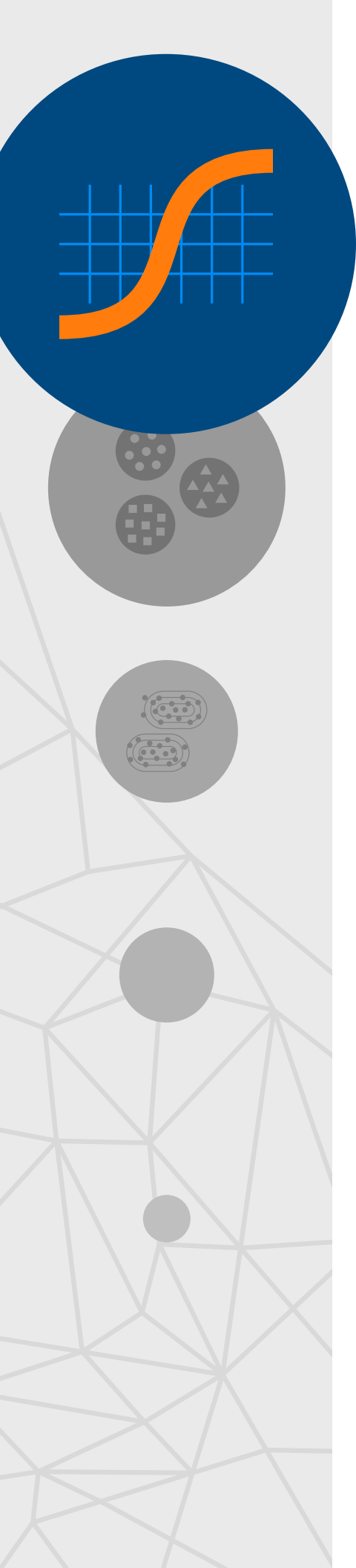


25 MACHINE LEARNING CLASSIFICATION QUESTIONS

WITH ANSWERS!

TEST YOUR KNOWLEDGE ON MOST POPULAR ML
CLASSIFICATION ALGORITHMS





1. Is Logistic Regression a linear or non-linear model?

It is a linear model.

2. What are the Logistic Regression assumptions?

First, binary logistic regression requires the dependent variable to be binary and ordinal logistic regression requires the dependent variable to be ordinal.

Second, logistic regression requires the observations to be independent of each other. In other words, the observations should not come from repeated measurements or matched data.

Third, logistic regression requires there to be little or no multicollinearity among the independent variables. This means that the independent variables should not be too highly correlated with each other.

Fourth, logistic regression assumes linearity of independent variables and log odds. Although this analysis does not require the dependent and independent variables to be related linearly, it requires that the independent variables are linearly related to the log odds.

3. Can Logistic Regression be used for many independent variables as well?

Yes, Logistic Regression can be used for as many independent variables as you want. However be aware that you won't be able to visualize the results in more than 3 dimensions.

4. What are odds?

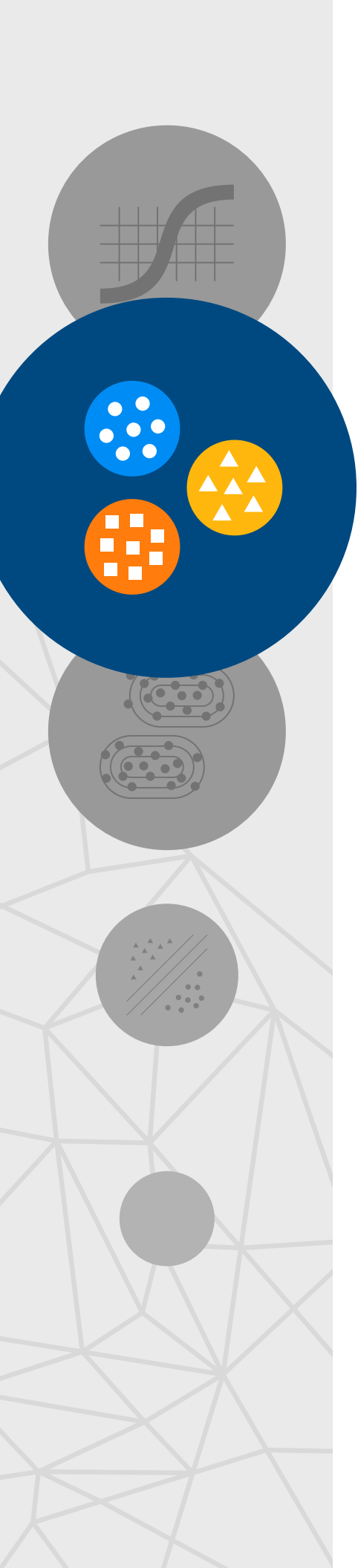
It is the ratio of the probability of an event occurring to the probability of the event not occurring. For example, let's assume that the probability of winning a lottery is 0.01.

Then, the probability of not winning is $1 - 0.01 = 0.99$

The odds of winning the lottery = (Probability of winning)/(probability of not winning)

The odds of winning the lottery = $0.01/0.99$

The odds of winning the lottery is 1 to 99, and the odds of not winning the lottery is 99 to 1.

A decorative sidebar on the left side of the page. It features a large dark blue circle containing three smaller circles: a blue one with white dots, an orange one with white squares, and a yellow one with white triangles. Above this is a grey circle with a white sine wave. Below are two grey circles with white geometric patterns (one with dots, one with lines). At the bottom is a grey circle with a white geometric pattern. The background of the sidebar is a light grey with a white geometric pattern of lines and circles.

5. Why is accuracy not a good measure for classification problems?

Accuracy is not a good measure for classification problems because it gives equal importance to both false positives and false negatives. However, this may not be the case in most business problems. For example, in the case of cancer prediction, declaring cancer as benign is more serious than wrongly informing the patient that he is suffering from cancer. Accuracy gives equal importance to both cases and cannot differentiate between them.

6. What does coefficient mean in logistic regression?

A regression coefficient describes the size and direction of the relationship between a predictor and the response variable. Coefficients are the numbers by which the values of the term are multiplied in a regression equation.

7. What is the intercept in logistic regression?

The intercept (often labeled the constant) is the expected mean value of Y when all $X=0$. Start with a regression equation with one predictor, X . If X sometimes equals 0, the intercept is simply the expected mean value of Y at that value.

8. What is the difference between “likelihood” and “probability”?

In non-technical parlance, "likelihood" is usually a synonym for "probability," but in statistical usage, there is a clear distinction in perspective: the number that is the probability of some observed outcomes given a set of parameter values is regarded as the likelihood of the set of parameter values given the observed outcomes.

9. What is K-Nearest Neighbors?

K-Nearest Neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition already at the beginning of 1970's as a non-parametric technique.

10. Tell the pros and cons of KNN?

Pros:

No assumptions about data — useful, for example, for nonlinear data
Simple algorithm — to explain and understand/interpret
High accuracy (relatively) — it is pretty high but not competitive in comparison to better-supervised learning models
Versatile — useful for classification or regression

Cons:

Computationally expensive — because the algorithm stores all of the training data
High memory requirement
Stores all (or almost all) of the training data
Prediction stage might be slow (with big N)
Sensitive to irrelevant features and the scale of the data

11. How does K-Nearest Neighbors work?

K-Nearest Neighbors algorithm is straightforward. It works based on minimum distance from the query instance to the training samples to determine the K-nearest neighbors. The data for the KNN algorithm consists of several multivariate attributes name that will be used to classify.

12. What is the difference between K means and KNN?

In machine learning, people often confused with k-means (k-means clustering) and KNN (k-Nearest Neighbors). K-means is an unsupervised learning algorithm used for clustering problem whereas KNN is a supervised learning algorithm used for classification and regression problem.

13. What K means in KNN?

'K' in KNN is the number of nearest neighbors used to classify or (predict in case of continuous variable/regression) a test sample

14. What is a Naive Bayes classifier algorithm?

It is a classification technique based on Bayes Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature.

15. What is “naive” in a Naive Bayes classifier?

In simple terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 4" in diameter. Even if these features depend on each other or upon the existence of the other features, a naive Bayes classifier considers all of these properties to independently contribute to the probability that this fruit is an apple.

16. Why we use Bayes Theorem?

The Bayes theorem describes the probability of an event based on the prior knowledge of the conditions that might be related to the event. If we know the conditional probability, we can use the Bayes rule to find out the reverse probabilities. The above statement is the general representation of the Bayes rule.

17. What is SVM?

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two dimensional space, this hyperplane is a line dividing a plane into two parts wherein each class lay in either side.

18. When SVM is used and which technique is using?

SVM is a supervised machine learning algorithm which can be used for classification or regression problems. It uses a technique called the kernel trick to transform your data and then based on these transformations it finds an optimal boundary between the possible outputs.

19. What is the SVM kernel?

SVM algorithms use a set of mathematical functions that are defined as the kernel. The function of the kernel is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions.

20. What is a hyperplane in SVM?

A Support Vector Machine (SVM) performs classification by finding the hyperplane that maximizes the margin between the two classes. The vectors (cases) that define the hyperplane are the support vectors.

21. What is C and gamma in SVM?

C and Gamma are the parameters for a nonlinear support vector machine (SVM) with a Gaussian radial basis function kernel. ... C is the parameter for the soft margin cost function, which controls the influence of each support vector; this process involves trading error penalty for stability.

22. What is Random Forest?

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set.

23. Is Random Forest bagging or boosting?

Random forest is a bagging technique and not a boosting. In boosting as the name suggests, one is learning from other which in turn boosts the learning. The trees in random forests are run in parallel. There is no interaction between these trees while building the trees.

24. What is the difference between a random forest and decision tree?

Random Forests is an ensemble classifier which uses many decision tree models to predict the result. A different subset of training data is selected, with replacement to train each tree. ... The fundamental difference is that Random Forest (RF) is a collection or ensemble model of numerous Decision Trees (DT).

25. Is random forest better than decision tree?

Random forest is a collection of decision trees. Random forests are a strong modeling technique and much more robust than a single decision tree. They aggregate many decision trees to limit overfitting as well as error due to bias and therefore yield useful results.

