1. Which statement about Logistic Regression is TRUE?
   1. Logistic Regression is a generalized linear model.
2. (True/False) Logistic regression is similar to a linear regression, except that it uses a logistic function to estimate probabilities of an observation belonging to a certain class or category.
   1. True
3. Usually the first step to fit a logistic regression model using scikit learn is to:
   1. import logistic regression from the sklearn.linear\_model module

e.g. from sklearn.linear\_model import LogisticRegression

# Module 1 Graded Quiz Logisitic Regression

1. The output of a logistic regression model applied to a data sample \_\_\_\_\_\_\_\_\_\_\_\_\_.
   1. is the probability of the sample being in a certain class.
2. Describe how any binary classification model can be extended from its basic form on two classes, to work on multiple classes.
   1. Use a one-versus all technique, where for each class you fit a binary classifier to that class versus all of the other classes.
3. Which tool is most appropriate for measuring the performance of a classifier on unbalanced classes?
   1. The precision-recall curve.
4. (True/False) One of the requirements of logistic regression is that you need a variable with two classes.
   1. False
5. (True/False) The shape of ROC curves are the leading indicator of an overfitted logistic regression.
   1. False
6. **You are evaluating a binary classifier. There are 50 positive outcomes in the test data, and 100 observations. Using a 50% threshold, the classifier predicts 40 positive outcomes, of which 10 are incorrect.**

What is the classifier’s Precision on the test sample?

* 1. 75%

1. **You are evaluating a binary classifier. There are 50 positive outcomes in the test data, and 100 observations. Using a 50% threshold, the classifier predicts 40 positive outcomes, of which 10 are incorrect.**

What is the classifier’s Recall on the test sample?

* 1. 60%

1. **You are evaluating a binary classifier. There are 50 positive outcomes in the test data, and 100 observations. Using a 50% threshold, the classifier predicts 40 positive outcomes, of which 10 are incorrect.**

What is the classifier’s F1 score on the test sample?

* 1. 66.7%

1. **You are evaluating a binary classifier. There are 50 positive outcomes in the test data, and 100 observations. Using a 50% threshold, the classifier predicts 40 positive outcomes, of which 10 are incorrect.**

Increasing the threshold to 60% results in 5 additional positive predictions, all of which are correct. Which of the following statements about this new model (compared with the original model that had a 50% threshold) is TRUE?

* 1. The area under the ROC curve would remain the same.

1. **You are evaluating a binary classifier. There are 50 positive outcomes in the test data, and 100 observations. Using a 50% threshold, the classifier predicts 40 positive outcomes, of which 10 are incorrect.**

The threshold is now increased further, to 70%. Which of the following statements is TRUE?

* 1. The Recall of the classifier would increase or remain the same.

1. (True/False) A simplified way to interpret K Nearest Neighbors is by thinking of the output of this method as a decision boundary which is then used to classify new points.
   1. True
2. These are all characteristics of the k nearest neighbors algorithm EXCEPT:
   1. It determines the value for *k*
3. (True/False) An advantage of k nearest neighbor methods is that they can leverage categorical data without encoding.
   1. False
4. Usually the first step to fit a k nearest neighbor classifier using scikit learn is to:
   1. import KneighborsClassifier from the sklearn.neighbors module

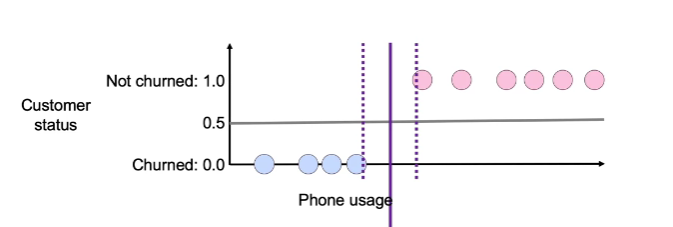
e.g. from sklearn.neighbors import KNeighborsClassifier

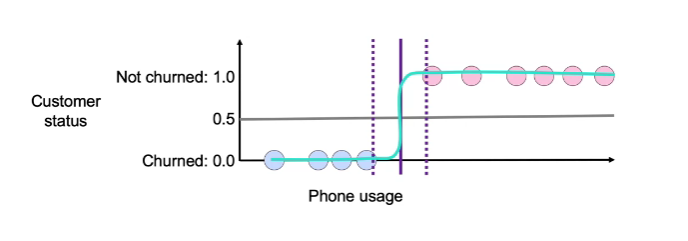
1. Which one of the following statements is true regarding K Nearest Neighbors?
   1. K Nearest Neighbors (KNN) assumes that points which are close together are similar.
2. Which one of the following statements is most accurate?
   1. K nearest neighbors (KNN) needs to remember the entire training dataset in order to classify a new data sample.
3. Which one of the following statements is most accurate about K Nearest Neighbors (KNN)?
   1. KNN can be used for both classification and regression.
4. (True/False) K Nearest Neighbors with large *k* tend to be the best classifiers.
   1. False
5. When building a KNN classifier for a variable with 2 classes, it is advantageous to set the neighbor count k to an odd number.
   1. True
6. The Euclidean distance between two points will always be shorter than the Manhattan distance:
   1. True
7. The main purpose of scaling features before fitting a k nearest neighbor model is to:
   1. Ensure that features have similar influence on the distance calculation
8. These are all pros of the k nearest neighbor algorithm EXCEPT:
   1. It is sensitive to the curse of dimensionality

# Support Vector Machines

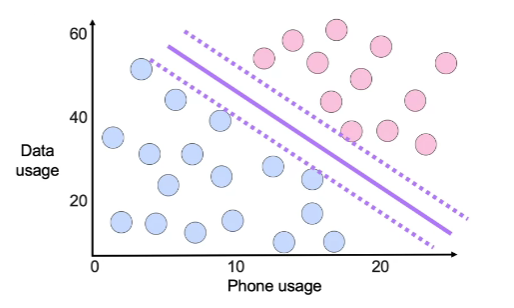
1. All of these are characteristics of SVMs, EXCEPT:
   1. Support Vector Machine models are non-linear.
2. (True/False) SVMs calculate predicted probabilities in the range between 0 and 1.
   1. False
3. (True/False) Any linear model can be turned into a non-linear model by applying a kernel to the model
   1. True
4. (True/False) SVMs with kernels are recommended for large data sets with many features
   1. False
5. Usually the first step to fit a support vector machine classifer model using scikit learn is to:
   1. import SVC from the sklearn.svm module

e.g. from sklearn.svm import SVC

1. Select the TRUE statement regarding the cost function for SVMs:
   1. SVMs use the Hinge Loss function as a cost function
2. Which statement about Support Vector Machines is TRUE?
   1. Support Vector Machine models rarely overfit on training data.
3. (True/False) A large *c* term will penalize the SVM coefficients more heavily.
   1. False
4. Regularization in the context of support vector machine (SVM) learning is meant to \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.
   1. lessen the impact that some minor misclassifications have on the cost function
5. Support vector machines can be extended to work with nonlinear classification boundaries by \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.
   1. using the kernel trick
6. Select the image that displays the line at the optimal point in the phone usage that the data can be split to create a decision boundary.
   1. 
7. The below image shows the decision boundary with a clear margin, such decision boundary belongs to what type machine learning model?



* 1. Support Vector Machine

1. SVM with kernals can be very slow on large datasets. To speed up SVM training, which methods may you perform to map low dimensional data into high dimensional beforehand?
   1. Nystroem
   2. RBF Sampler
2. Concerning the Machine Learning workflow what model choice would you pick if you have "Few" features and a "Medium" amount of data?
   1. SVC with RBF
3. Select the image that best displays the line that separates the classes.
   1. 
4. Which of the following statements about Decision Tree models is TRUE?
   1. Decision Tree models are non-linear.
5. (True/False) Decision Trees are considered a greedy algorithm.
   1. True
6. Usually the first step to fit a decision tree classifier using scikit learn is to:
   1. import decision tree classifier from the sklearn.tree module

e.g. from sklearn.tree import DecisionTreeClassifier

1. These are all characteristics of decision trees, EXCEPT:
   1. They have well rounded decision boundaries
2. Decision trees used as classifiers compute the value assigned to a leaf by calculating the ratio: number of observations of one class divided by the number of observations in that leaf E.g. number of customers that are younger than 50 years old divided by the total number of customers.

How are leaf values calculated for regression decision trees?

* 1. average value of the predicted variable

1. These are two main advantages of decision trees:
   1. They are very visual and easy to interpret
2. How can you determine the split for each node of a decision tree?
   1. Find the split that minimizes the gini impurity.
3. Which of the following describes a way to regularize a decision tree to address overfitting?
   1. Decrease the max depth.
4. What is a disadvantage of decision trees?
   1. They tend to overfit.
5. What method can you use to minimize overfitting of a machine learning model?
   1. Tune the hyperparameters of your model using cross-validation.
6. Concerning Classification algorithms, what is a characteristic of K-Nearest Neighbors?
   1. Fitting can be slow
7. Concerning Classification algorithms, what are the characteristics of Logistic Regression?
8. When evaluating all possible splits of a decision tree what can be used to find the best split regardless of what happened in prior or future steps?
   1. Greedy Search
9. (True/False) A model that averages the predictions of multiple models reduces the variance of a single model and has high chances to generalize well when scoring new data.
   1. True
10. (True/False) Bagging is a tree ensemble that combines the prediction of several trees that were trained on bootstrap samples of the data.
    1. True
11. (True/False) In general, a random forest can be considered a special case of bagging and it tends to have better out of sample accuracy
    1. True
12. Usually the first step to fit a random forest classifier model using scikit learn is to:
    1. import random forest classifer from the sklearn.ensemble module

e.g. from sklearn.ensemble import RandomForestClassifer

1. Usually the first step to fit a bagging classifier model using scikit learn is to:
   1. import bagging classifier from the sklearn.ensemble module

e.g. from sklearn.ensemble import BaggingClassifier

1. (True/False) Bagging tends to have less overfitting than decision trees.
   1. True
2. (True/False) Boosting tend to be well suited for data sets with outliers and rare events.
   1. True
3. All of these are characteristics of boosting algorithms, EXCEPT:
   1. They create trees independently
4. Usually the first step to fit a gradient boosting classifier model using scikit learn is to:
   1. import gradient boosting classifier from the sklearn.ensemble module

e.g. from sklearn.ensemble import GradientBoostingClassifier

1. (True/False) If you were to combine several logistic regressions using a voting ensemble, you should use a Voting Regressor.
   1. False
2. The term *Bagging* stands for bootstrap aggregating.
   1. True
3. This is the best way to choose the number of trees to build on a Bagging ensemble.
   1. Tune number of trees as a hyperparameter that needs to be optimized
4. Which type of Ensemble modeling approach is NOT a special case of model averaging?
   1. Boosting methods
5. What is an ensemble model that needs you to look at out of bag error?
   1. Random Forest
6. What is the main condition to use stacking as ensemble method?
   1. Models need to output predicted probabilities
7. This tree ensemble method only uses a subset of the features for each tree:
   1. Random Forest
8. Order these tree ensembles in order of most randomness to least randomness:
   1. Random Trees, Random Forest, Bagging
9. This is an ensemble model that does not use bootstrapped samples to fit the base trees, takes residuals into account, and fits the base trees iteratively:
   1. Boosting
10. When comparing the two ensemble methods Bagging and Boosting, what is one characteristic of Boosting?
    1. Fits entire data set
11. What is the most frequently discussed loss function in boosting algorithms?
    1. 0-1 Loss Function
12. What type of forest is a classification algorithm that potentially contains hundreds of different decision trees?
    1. Random Forest
13. When describing models what type of model will feature coefficients help to explain?
    1. Linear Model
14. What type of surrogate model tries to approximate a black-box model globally on every instance in the data set?
    1. Global Surrogate model
15. These are all methods of dealing with unbalanced classes EXCEPT:
    1. Mix of in-sample and out-of-sample.
16. (True/False) A best practice to build a model using unbalanced classes is to split the data first, then apply an upsample or undersample technique.
    1. True
17. Which of the following statements about Downsampling is TRUE?
    1. Downsampling is likely to decrease Precision.
18. Which of the following statements about Random Upsampling is TRUE?
    1. Random Upsampling preserves all original observations.
19. Which of the following statements about Synthetic Upsampling is TRUE?
    1. Synthetic Upsampling generates observations that were not part of the original data.
20. What can help humans to interpret the behaviors and methods of Machine Learning models more easily?
    1. Model Explanations
21. What type of explanation method can be used to explain different types of Machine Learning models no matter the model structures and complexity?
    1. Model Explanations
22. What reason might a Global Surrogate model fail?
    1. Large inconsistency between surrogate models and black-box models
23. When working with unbalanced sets, what should be done to the samples so the class balance remains consistent in both the train and test set?
    1. Stratify the samples
24. What approach are you using when trying to increase the size of a minority class so that it is similar to the size of the majority class?
    1. Oversampling
25. What approach are you using when you create a new sample of a minority class that does not yet exist?
    1. Synthetic Oversampling
26. What intuitive technique is used for unbalanced datasets that ensures a continuous downsample for each of the bootstrap samples?
    1. Blagging