**Algorithm Understanding**

**Is SVM (Support Vector Machine) a supervised or unsupervised learning algorithm?  
Why is SVM such a powerful classification method?  
What are 3 disadvantages of SVMs?**

* SVM (Support Vector Machine) is a supervised learning algorithm.
* Because it can discover the best boundary between classes, known as the maximum margin hyperplane, SVM is a powerful classification approach. This boundary is designed to maximize the distance between itself and the nearest data points from each class, which are known as the support vectors. This produces a border that not only effectively separates the classes but also has the best generalization performance, which means it is less likely to misclassify fresh unseen data. Furthermore, SVM can be expanded to accommodate non-linearly separable data by employing the kernel approach, which projects the data into a higher dimensional space where it can be separated linearly.
  1. Training an SVM can be computationally costly, especially when working with large datasets or when utilizing more sophisticated kernel functions. This can make the model challenging to train and scale to massive data situations.
  2. Sensitive to kernel function: The performance of an SVM might be very dependent on the kernel function utilized. Choosing the incorrect kernel might lead to poor performance or failure to converge.
  3. Interpretability is limited: Because Support Vector Machines do not provide explicit explanations of how algorithms arrive at their predictions, it can be difficult to grasp why a particular conclusion was reached. This lack of interpretability might be an issue in some applications, such as healthcare or finance, where understanding the reasons behind a choice is critical.

**Interview Readiness**

**What is the time complexity of SVM?   
What is it for Logistic Regression?**

* The optimization algorithm determines SVM training time. The Sequential Minimal Optimization (SMO) approach, which breaks the problem into smaller subproblems that may be solved analytically, is the most popular. The SMO algorithm takes O(n^2\*k) iterations to converge, where n is the number of training samples.
* The optimization algorithm determines logistic regression model training time. Batch gradient descent and Newton-Raphson are popular algorithms.

The batch gradient descent approach takes O(nkd) time to converge, where n is the number of training instances, k is the number of iterations, and d is the number of features.

**Interview Readiness**

**Explain feature importance for the Random Forest algorithm?  
When examining feature importance, what is Gini impurity or information gain?**

* Feature relevance in a random forest is the relative importance of each dataset feature on the prediction outcome. The most typical random forest feature significance calculations are:
  1. Gini importance: When a feature is separated, Gini impurity decreases. Misclassifying a random sample in data is measured by the Gini impurity. Gini significance is the average decrease in Gini impurity across all trees in the forest, weighted by the number of samples passing through the node where the feature is employed in a split.
  2. Permutation importance: This method includes shuffling feature values for a small portion of data and assessing the prediction accuracy loss. A feature's permutation relevance is its average accuracy drop over random permutations.
  3. Mean decrease impurity (MDI): This alternative to Gini importance measures feature importance by reducing impurity (measured by Gini or Entropy).
  4. Mean decrease accuracy (MDA): This alternative to permutation importance measures feature importance by reducing accuracy.
* Decision tree-based algorithms employ Gini impurity and information gain to choose the optimum feature to split on at each node.

Gini impurity measures how often a randomly chosen element from the collection would be erroneously identified if randomly labeled according to the set's label distribution. The Gini impurity is the sum of the square of the probability of each label in the set and varies from 0 to 1, with 0 indicating that all items in the set have the same label and 1 indicating that all elements have an equal likelihood of being any label.

**Interview Readiness**

**SHAP (SHapley Additive exPlanations) is a game theoretic approach to explain the output of any machine learning model, what is it and how does it work?**

* SHAP (SHapley Additive exPlanations) assigns feature relevance scores to input dataset features to explain machine learning model output. Shapley values from cooperative game theory are used to fairly allocate a value among a group of people based on their efforts. SHAP values for a prediction calculate the average marginal contribution of each feature, taking into consideration all potential feature combinations. Local and global explanations are used to do this (i.e., the contribution of each feature to the overall model behavior).