

1. Predict the running times of prospective Olympic sprinters using data from the last 20 Olympics
 - a. This would be a ordinal regression problem because we are predicting a continuous value and place matters in the prediction of running times.
2. You have more features (columns) than rows in your dataset.
 - a. Use PCA to help reduce the number of columns then choose an accurate classifier or regression model depending on the question.
3. Identify the most important characteristic predicting likelihood of being jailed before age 20.
 - a. We are looking for the most prominent variable, a random forest or could help reveal this and show more accuracy than a decision tree.
4. Implement a filter to “highlight” emails that might be important to the recipient
 - a. A random forest could be used to classify the right emails based on text
5. You have 1000+ features
 - a. Lasso regression will reduce the number of features
6. Predict whether someone who adds items to their cart on a website will purchase the items
 - a. Logistic regression because it will highlight the probability of an outcome based
7. Your dataset dimensions are 982400 x 500
 - a. Lasso logistic regression
8. Identify faces in an image.
 - a. SVM Classification using a hard boundary or logistic regression - the probability of the image being that person

9. Predict which of three flavors of ice cream will be most popular with boys vs girls.
 - a. To accurately predict the flavor based on gender, a random forest would be highly accurate