1. Predict the running times of prospective Olympic sprinters using data from the last 20 Olympics.

Linear regression. Target variable is continuous and feature variables should be correlated with target.

1. You have more features (columns) than rows in your dataset.

Reduce the number of features in the dataset using PCA and then pick an supervised learning model based on the type of data.

1. Identify the most important characteristic predicting likelihood of being jailed before age 20.

Random forest has a feature importance attribute that tells which features are most important to the data set.

1. Implement a filter to “highlight” emails that might be important to the recipient

Naïve Bayes. This is useful when target variable is categorical. In this case the email will either be important (1) or not important (0).

1. You have 1000+ features.

Lasso regression is useful for data with thousands of features. This works as feature selection by return 0 for small parameter estimates. The variables can be removed, and the model can be run again.

1. Predict whether someone who adds items to their cart on a website will purchase the items.

KNN classifier could work. The target is binary. The user buys (1) or doesn’t buy (0). I guess it depends on what the other variables in the data set are.

1. Your dataset dimensions are 982400 x 500

Depending on the type of variables, using PCA could be a good idea to reduce the number of features.

1. Identify faces in an image.

Random Forest classifier. The image will either identify the face or it will not identify the face. Logistic regression may also work.

1. Predict which of three flavors of ice cream will be most popular with boys vs girls.

Random forest or decision tree if data is balanced.