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

**Candidates**

OLS, KNN Regression, DecisionTreeRegression, SupportVectorRegression

+ RandomForest, GradientBoosting

I think OLS and multivariate regression are not the best for this kind of task. Because they assume predictors are independent but multicollinearity is very likely. KNN may work well. Another Non-parametric way is Decision tree. DT overfit easily and don’t generalize very well. Sprinters running times are numbers close to each other so DT can generate a completely different tree for a small. Random Forest and Boosting may circumvent DT problems.

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

PCA can reduce the dimensions. I can also select good features using RandomForest or Support Vector Machine. Adding L1 or L2 penalty to the loss function will act as embedded feature selection.

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

Background of parents, past crime history of parents, neighborhood, education performance, mental status, bad habits like smoking, drinking, crime rate in the location, unemployment rate.

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

It is a binary classification problem. The features are important keywords. SVC perform very well for this kind of tasks.

1. You have 1000+ features.

L1 regularized regression will be good. PCA, Radom Forest or SVC can show the most important features.

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

It is simple problem. It can be done using Naïve Bayes

1. Your dataset dimensions are 982400 x 500

A lot of data. Models can generalize better. KNN/OLS/DT or any other model can be used. There is a lot of data so DT may do better.

1. Identify faces in an image.

It is a classification problem. The pixels are taken as features. Multinomial Logistic regression works well because the pixels are not likely to be correlated with each other. Convolutional neural networks are the best for this kind of task.

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

Gender is the predictor. The outcome is the selected ice cream flavor. A supervised learning model can be trained to predict the best ice cream flavor for boys and girls.

Any classifier can do the task.

https://hackernoon.com/choosing-the-right-machine-learning-algorithm-68126944ce1f