For each, identify which supervised learning method(s) would be best for addressing that particular problem. Explain your reasoning and discuss your answers with your mentor.

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

Since this is a continuous variable I would choose a regression and a Random forest Regression since it is very efficient model.

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

I would use Random Forest or Lasso Regression model since both have feature selection embedded in its algorithm

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

I would use Recursive Feature elimination and find the most important feature/characteristic of being jailed. I would then use Logistic regression or any other classifiers.

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

KNN classifier is good in finding similarities

1. You have 1000+ features.

Lasso Regression has embedded feature selection algorithm of giving penalties or can do PCA and then model with Lasso or Random forest or Gradientboost

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

SVC or KNN classifier

1. Your dataset dimensions are 982400 x 500

SVM and use sub-sampling.

1. Identify faces in an image.

We can use PCA and SVM (Support vector machine)

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

KNN classifier