You now have a fairly substantial starting toolbox of supervised learning methods that you can use to tackle a host of exciting problems. To make sure all of these ideas are organized in your mind, please go through the list of problems below. 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.
2. You have more features (columns) than rows in your dataset.
3. Identify the most important characteristic predicting likelihood of being jailed before age 20.
4. Implement a filter to “highlight” emails that might be important to the recipient
5. You have 1000+ features.
6. Predict whether someone who adds items to their cart on a website will purchase the items.
7. Your dataset dimensions are 982400 x 500
8. Identify faces in an image.
9. Predict which of three flavors of ice cream will be most popular with boys vs girls.
10. Predict the running times of prospective Olympic sprinters using data from the last 20 Olympics.

Run the EDA and workflow pipeline for a sample of the data (metadata), and then use regression techniques based on the accuracy that we are getting. Few models include:

Linear, Lasso, Ridge, Elastic Net CV.

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

We could implement regularization techniques or models such as ridge regression, subset selection, PCA.

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

This is a classification model. We would need a feature that explains or is important enough to explain why someone is jailed. We could use Logisitc regression, Gradient bosstic, etc. to see which model has a better precision recall and accuracy.

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

It is a classification problem, we would need a model that is best in classifying textual data.

1. You have 1000+ features.

Find the most important features, perform dimensional reduction like PCA to preserve atleast 95-97% of the data.

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

This is again a classification problem, this is more of a binary problem and logistic regression would fit better.

1. Your dataset dimensions are 982400 x 500

Linear models are preferred as they can handle datasets with huge sizes.

1. Identify faces in an image.

Convolutional Neural Network would be an apt model for this problem. The reason is that CNN can capture complex images using learnable weights.

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

Okay I believe there are two columns, ice cream flavor and gender. If we have features like toppings, shapes, etc. we would have a better understanding. This is anyway a binary classification model.