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.
   * Use OLS regression.
   * Because
     1. The dependent variable is continuous
     2. OLS is simple
     3. Not much data to use with some of the other methods (e.g. RF, SVM regressions)
2. You have more features (columns) than rows in your dataset.
   * Lasso Logistic regression if classification problem
     1. Lasso will zero-out many features
3. Identify the most important characteristic predicting likelihood of being jailed before age 20.
   * Logistic regression (look at highest R2 of a single factor) or RFC (look at feature importance)
4. Implement a filter to “highlight” emails that might be important to the recipient
   * If importance is about the existence of certain words, use Naïve Bayes.
   * If importance is about a certain topic use KNN
   * If importance is about some subtle combination of factors use Random Forests (can you apply SVM to NL questions???)
5. You have 1000+ features.
   * Use regression bases framework because it is fast
   * Or use PCA (or correlation analysis) to do some feature selection and then use SVM, RFC.
   * Aside from size question, another important here is whether relationship is linear or nor
6. Predict whether someone who adds items to their cart on a website will purchase the items.
   * Try Logistic Regression, SVM and RFC and see which one gives better results.
7. Your dataset dimensions are 982400 x 500
   * Try logistic Regression or SVM (due to size RFC is probably too much).
8. Identify faces in an image.
   * KNN, because faces do not depend on a particular pixel
   * Maybe boosting with weak learner KNN
9. Predict which of three flavors of ice cream will be most popular with boys vs girls.
   * Target variable seems to be a 9 (3x3) class pair (b,g) combination.
   * Decision Tree?