Project Hints

Get Your Ball Rolling

- A strawman solution

```
//load data
X_train, y_train = load('train.csv')
X_test = load('test.csv')
//train classifier
classifier = train(X_train, y_train, 'your classifier', parameters)
//prediction
result = classifier.predict_proba(X_test)
```

Improvement Strategies

- Feature engineering is often the most important part
  - start with all the possible features you can come up
  - e.g., combine features to build new features
- Simple models will get your far
  - easy to train and adapt, force you to work on data
- Ensembling is a winning strategy
  - bag of models which are loosely correlated
    - loosely correlated models
    - loosely correlated data
- Keeping track of your progress
  - making full use of training data
    - cross validation
  - overfitting leader board is an issue