Plan:

1. Classification. The classification problem is just like the regression problem, except that the values we now want to predict take on only a small number of discrete values, so we are going to use a method called Logistic regression.
    a. Hypothesis function
    b. Decision boundary
2. Cost function and logistic function. We cannot use the same cost function that we use for linear regression because of for our logistic function output will be a non-convex function.
    a. Simplified cost-function
    b. Gradient descent
    c. Advanced optimization
3. Multiclass classification. Understanding of classification with more than 2 classes.
4. Underfitting and Overfitting. The number of features in a model can lead to two different problems - if it's too small - underfitting, if it's too big - overfitting.
    a. Strategies to avoid these problems
    b. Regularization

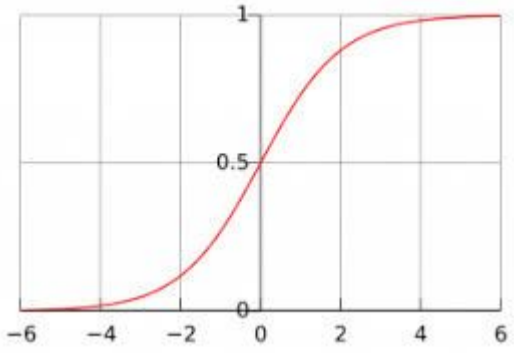
Questions:

1. How many classification models we need to classify data as one of two classes?
2. How many classification models we need to classify data as one of five classes?
3. What's gonna happen with the model's accuracy on training data after including regularization?

Glossary:

Overfitting - model's inability to reliably predict future values caused by corresponding too closely on a training dataset - *We can solve overfitting problem by using regularization.*

BFGS - Broyden–Fletcher–Goldfarb–Shanno algorithm - quasi-Newton iterative optimization algorithm - *BFGS is extremely effective but very hard to understand.*



One-vs-all (one-vs-rest):