CS-541
Quiz 1
Wyatt Blair

1.
    a. True
    b. False
    c. False
    d. True
    e. False
2.
    a. Stochastic gradient descent takes a single data point or a subset of data points and updates w using only those points. Gradient descent takes the entire set of data and updates w using all of those points each time. Because of this, gradient descent can be more prone to overfitting and will sometimes fall into local minima more easily than stochastic gradient descent.
    b. The biggest difference between the two regimes is that Supervised Learning has labeled data (and therefore requires a more complete knowledge in the dataset) and Unsupervised Learning does not have labeled data.
    c. Cross-Validation and feature selection. Cross-validation takes different train and test subsets and then trains different instances of the model using these subsets. From there a more accurate estimation of the error is obtained. Feature selection removes any redundant features.
    d. Since we are using hidden layers, we are not necessarily sure what features are triggering which neurons, meaning the design of good features ends up not being as important. Collecting good data is *ALWAYS* important of course.