1. The Mean Squared Error (MSE) is a measure of how close a fitted line is to data points and the Root Mean Squared Error (RMSE) is just the square root of the mean square error. Many best-fit algorithms use the MSE/RMSE error methods to find a regression line. RMSE better reflects performance when dealing with large error values because it penalizes large error values.

2. a) When training a model, as you add more features to the dataset, you often need to increase the dataset's size to ensure the model learns reliably. This is because excessively adding features can often lead to overfitting if the dataset size is not increased.

3. (c) Insufficient training data size. Overfitting is the cause of the model beginning to memorize the training set rather than learn the patterns of it. This can be fixed by either adding more data to the training set or decreasing the data in the training set to force the model to focus more on the patterns.

4. Perceptron
   (a) is a linear classifier. A. True
   (b) and cannot be trained on linearly inseparable data B. False. If a dataset is not linearly separable, learning will just never reach a point where all vectors are classified properly.

5. Logistic Regression
   a. is a linear classifier. A. True
   b. and always has a unique solution B. False

6. (e) For convex loss functions (i.e. with a bowl shape), both stochastic gradient descent and batch gradient descent will eventually converge to the global optimum.

7. $\begin{matrix} -2 \\ -1 \\ 3 \end{matrix} = \beta \begin{matrix} -1 \\ 1 \\ 3 \end{matrix} + \epsilon i$

8.

9. (Extra Credit) It is ok to initialize parameters with zeros when training a logistic regression model because it has a convex cost function. This means there will only be a single minimum point which contrasts with the many minima of neural networks where it is a bad idea to initialize parameters with zeroes.