Problem 1
(10 points) Explain how changing $\Sigma$ in the multivariate Gaussian distribution changes the contours where $\Sigma$ is the covariance matrix. *Hint:* you can consider that $\Sigma$ is a $2 \times 2$ matrix. It is the same for higher dimensions.

Problem 2
(10 points) How would you check that your features are Gaussian? Give two possible transformations that could make them Gaussian?

Problem 3
(10 points) When would you use anomaly detection versus when would you use supervised learning?

Problem 4
(10 points) What are the steps required to implement anomaly detection? Describe the formula you are implementing and explain what each variable represents. Finally, describe how to compute these variables and how to decide if a training example is an anomaly?

Problem 5
(10 points) What is the difference between mini-batch gradient descent and stochastic gradient descent?

Problem 6
(10 points) How would you update your theta in a map-reducible learning algorithm (i.e. linear regression)?

Problem 7
Choose True/False. No justification needed.

1. (10 points) In general, the smaller learning rate, the better the solution with stochastic gradient descent. One explanation is that stochastic gradient descent will make smaller steps around the global minimum hence getting closer than with a bigger learning rate.
2. (10 points) Your learning algorithm is map-reducible if it can be expressed as computing sums of functions over the data set.

3. (10 points) One strategy to debug stochastic gradient descent is to plot the average of the cost function on a batch of data points against the number of iterations.

4. (10 points) In order to converge to the global minimum with the stochastic gradient method, you need to decrease the learning rate $\alpha$ over time.