1. For which of the following problems would anomaly detection be a suitable algorithm?

- [ ] In a computer chip fabrication plant, identify microchips that might be defective.

This should be selected

- [x] Given data from credit card transactions, classify each transaction according to type of purchase (for example: food, transportation, clothing).

This should not be selected

Anomaly detection is not appropriate for a traditional classification problem.

- [ ] From a large set of primary care patient records, identify individuals who might have unusual health conditions.

This should be selected

- [ ] From a large set of hospital patient records, predict which patients have a particular disease (say, the flu).

Un-selected is correct

2. Suppose you have trained an anomaly detection system for fraud detection, and your system that flags anomalies when $p(x)$ is less than ε, and you find on the cross-validation set that it is missing many fradulent transactions (i.e., failing to flag them as anomalies). What should you do?

- ( ) Decrease ε
- (•) Increase ε

Correct

By increasing ε, you will flag more anomalies, as desired.

3. Suppose you are developing an anomaly detection system to catch manufacturing defects in airplane engines. You model uses

$p(x) = \prod_{j=1}^{n} p(x_j; \mu_j, \sigma_j^2)$.

You have two features x_1 = vibration intensity, and x_2 = heat generated. Both x_1 and x_2 take on values between 0 and 1 (and are strictly greater than 0), and for most "normal" engines you expect that $x_1 \approx x_2$. One of the suspected anomalies is that a flawed engine may vibrate very intensely even without generating much heat (large x_1, small x_2), even though the particular values of x_1 and x_2 may not fall outside their typical ranges of values. What additional feature x_3 should you create to capture these types of anomalies:

- (selected) $x_3 = \frac{x_1}{x_2}$

Correct
This is correct, as it will take on large values for anomalous examples and smaller values for normal examples.

- $x_3 = \frac{1}{x_2}$
- $x_3 = \frac{1}{x_1}$
- $x_3 = x_1 + x_2$

4. Which of the following are true? Check all that apply.

- [ ] If you are developing an anomaly detection system, there is no way to make use of labeled data to improve your system.

Un-selected is correct

- [ ] When choosing features for an anomaly detection system, it is a good idea to look for features that take on unusually large or small values for (mainly the) anomalous examples.

This should be selected

If you have a large labeled training set with many positive examples and many negative examples, the anomaly detection algorithm will likely perform just as well as a supervised learning algorithm such as an SVM.

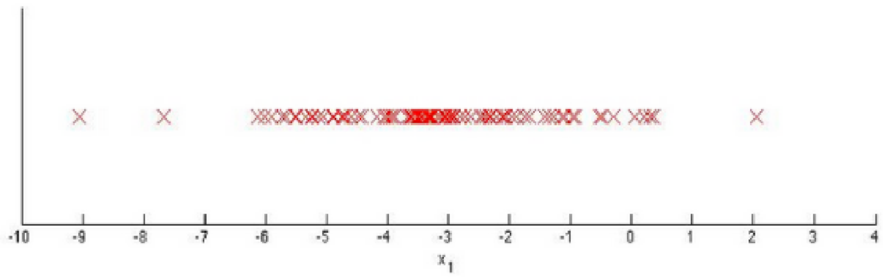
This should not be selected

Anomaly detection only models the negative examples, whereas an SVM learns to discriminate between positive and negative examples, so the SVM will perform better when you have many positive and negative examples.

If you do not have any labeled data (or if all your data has label $y = 0$), then is is still possible to learn $p(x)$, but it may be harder to evaluate the system or choose a good value of ϵ.

This should be selected

5. You have a 1-D dataset $\{x^{(1)}, \ldots, x^{(m)}\}$ and you want to detect outliers in the dataset. You first plot the dataset and it looks like this:



Suppose you fit the gaussian distribution parameters μ_1 and σ_1^2 to this dataset. Which of the following values for μ_1 and σ_1^2 might you get?

$\mu_1 = -3, \sigma_1^2 = 4$

Correct

Correct

This is correct, as the data are centered around -3 and tail most of the points lie in [-5, -1].

- ◯ $\mu_1 = -6, \sigma_1^2 = 4$
- ◯ $\mu_1 = -3, \sigma_1^2 = 2$
- ◯ $\mu_1 = -6, \sigma_1^2 = 2$

1. For which of the following problems would anomaly detection be a suitable algorithm?

- [x] Given a dataset of credit card transactions, identify unusual transactions to flag them as possibly fraudulent.

Correct

By modeling "normal" credit card transactions, you can then use anomaly detection to flag the unusuals ones which might be fraudulent.

- [ ] Given data from credit card transactions, classify each transaction according to type of purchase (for example: food, transportation, clothing).

Un-selected is correct

- [x] From a large set of primary care patient records, identify individuals who might have unusual health conditions.

Correct

Since you are just looking for unusual conditions instead of a particular disease, this is a good application of anomaly detection.

- [ ] Given an image of a face, determine whether or not it is the face of a particular famous individual.

Un-selected is correct

2. Suppose you have trained an anomaly detection system for fraud detection, and your system that flags anomalies when $p(x)$ is less than ε, and you find on the cross-validation set that it mis-flagging far too many good transactions as fradulent. What should you do?

- ( ) Increase ε
- (•) Decrease ε

Correct
By decreasing ε, you will flag fewer anomalies, as desired.

4. Which of the following are true? Check all that apply.

- [x] When developing an anomaly detection system, it is often useful to select an appropriate numerical performance metric to evaluate the effectiveness of the learning algorithm.

Correct
You should have a good evaluation metric, so you can evaluate changes to the model such as new features.

- [ ] When evaluating an anomaly detection algorithm on the cross validation set (containing some positive and some negative examples), classification accuracy is usually a good evaluation metric to use.

Un-selected is correct

- [ ] In a typical anomaly detection setting, we have a large number of anomalous examples, and a relatively small number of normal/non-anomalous examples.

Un-selected is correct

- [x] In anomaly detection, we fit a model $p(x)$ to a set of negative ($y = 0$) examples, without using any positive examples we may have collected of previously observed anomalies.

Correct
We want to model "normal" examples, so we only use negative examples in training.