✓ **Congratulations! You passed!**

Next Item

✓ 1 / 1
point

1.

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

☑ 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.

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

Un-selected is correct

☑ In a computer chip fabrication plant, identify microchips that might be defective.

Correct
The defective chips are the anomalies you are looking for by modeling the properties of non-defective chips.

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

Un-selected is correct

✓ 1 / 1
point

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 fraudulent transactions (i.e., failing to flag them as anomalies). What should you do?

○ Increase ε

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

○ Decrease ε

✔ 1 / 1 point

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:

○ $x_3 = x_1 \times x_2^2$

○ $x_3 = \frac{x_1}{x_2}$

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

○ $x_3 = (x_1 + x_2)^2$

○ $x_3 = x_1^2 \times x_2^2$

✔ 1 / 1 point

4.

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

☐

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.

Un-selected is correct

☐ 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.

Correct

Only negative examples are used in training, but it is good to have some labeled data of both types for cross-validation.

☐ 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.

Correct

These are good features, as they will lie outside the learned model, so you will have small values for $p(x)$ with these examples.
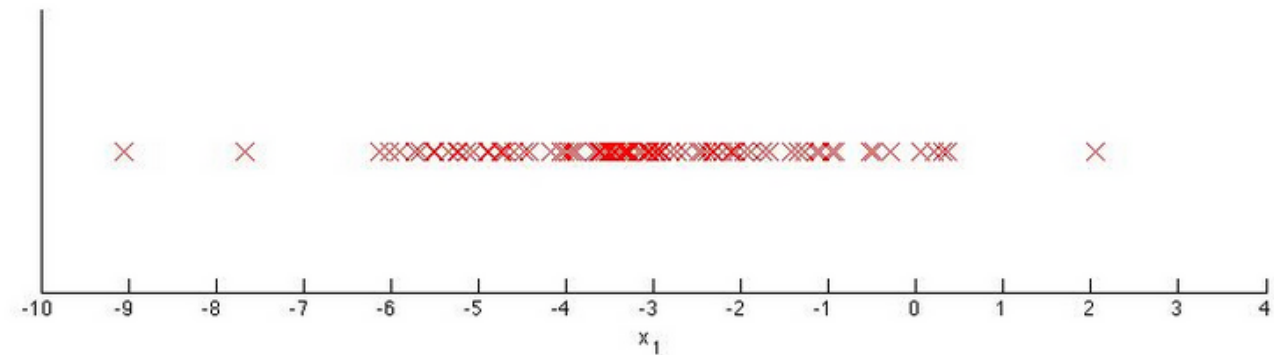

✖ 0 / 1
point

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

Anomaly Detection
dataset and it looks like this:

4/5 points (80.00%)

Quiz, 5 questions

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?

○ $\mu_1 = -3, \sigma_1^2 = 4$

○ $\mu_1 = -6, \sigma_1^2 = 4$

⦿ $\mu_1 = -3, \sigma_1^2 = 2$

▲

This should not be selected
This is the correct value for μ_1, but most of the data are in [-5, -1], so σ_1^2 is 4, not 2.

○ $\mu_1 = -6, \sigma_1^2 = 2$