

Feedback — XV. Anomaly Detection

[Help](#)

You submitted this quiz on **Sun 18 May 2014 11:47 PM IST**. You got a score of **5.00** out of **5.00**.

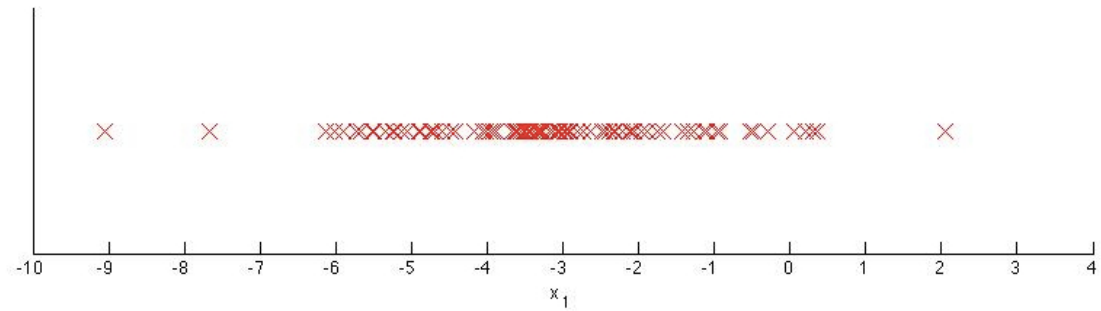
Question 1

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

Your Answer	Score	Explanation
<input type="checkbox"/> Given data from credit card transactions, classify each transaction according to type of purchase (for example: food, transportation, clothing).	✓ 0.25	Anomaly detection is not appropriate for a traditional classification problem.
<input checked="" type="checkbox"/> Given a dataset of credit card transactions, identify unusual transactions to flag them as possibly fraudulent.	✓ 0.25	By modeling "normal" credit card transactions, you can then use anomaly detection to flag the unusual ones which might be fraudulent.
<input checked="" type="checkbox"/> In a computer chip fabrication plant, identify microchips that might be defective.	✓ 0.25	The defective chips are the anomalies you are looking for by modeling the properties of non-defective chips.
<input type="checkbox"/> Given an image of a face, determine whether or not it is the face of a particular famous individual.	✓ 0.25	This problem is more suited to traditional supervised learning, as you want both famous and non-famous images in the training set.
Total	1.00 / 1.00	

Question 2

You have a 1-D dataset $\{x^{(1)}, \dots, x^{(m)}\}$ 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?

Your Answer	Score	Explanation
<input type="radio"/> $\mu_1 = -6, \sigma_1^2 = 2$		
<input checked="" type="radio"/> $\mu_1 = -3, \sigma_1^2 = 4$	1.00	This is correct, as the data are centered around -3 and tail most of the points lie in [-5, -1].
<input type="radio"/> $\mu_1 = -3, \sigma_1^2 = 2$		
<input type="radio"/> $\mu_1 = -6, \sigma_1^2 = 4$		
Total	1.00 / 1.00	

Question 3

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?

Your Answer	Score	Explanation
<input type="radio"/> Decrease ε		
<input checked="" type="radio"/> Increase ε	1.00	By increasing ε , you will flag more anomalies, as desired.
Total	1.00 / 1.00	

Question 4

Suppose you are developing an anomaly detection system to catch manufacturing defects in airplane engines. Your 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:

Your Answer	Score	Explanation
<input type="radio"/> $x_3 = x_1^2 \times x_2^2$		
<input type="radio"/> $x_3 = x_1^2 \times x_2$		
<input checked="" type="radio"/> $x_3 = \frac{x_1}{x_2}$	✓ 1.00	This is correct, as it will take on large values for anomalous examples and smaller values for normal examples.
<input type="radio"/> $x_3 = \frac{1}{x_1}$		
Total	1.00 / 1.00	

Question 5

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

Your Answer	Score	Explanation
<input checked="" type="checkbox"/> 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.	✓ 0.25	You should have a good evaluation metric, so you can evaluate changes to the model such as new features.
<input type="checkbox"/> When evaluating an anomaly detection algorithm	✓ 0.25	Classification accuracy is a poor metric because of the skewed classes in the cross-

on the cross validation set (containing some positive and some negative examples), classification accuracy is usually a good evaluation metric to use.

validation set (almost all examples are negative).

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

✓ 0.25

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

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

✓ 0.25

It is the reverse: we have many normal examples and few anomalous examples.

Total

1.00 /
1.00