

Feedback — XV. Anomaly Detection

[Help](#)

You submitted this quiz on **Mon 18 Aug 2014 12:36 PM PDT**. You got a score of **5.00** out of **5.00**.

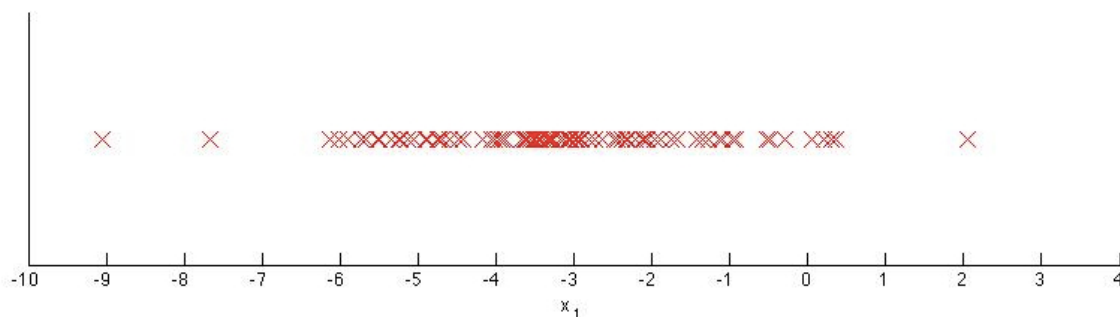
Question 1

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 unusuals ones which might be fraudulent.
<input checked="" type="checkbox"/> From a large set of primary care patient records, identify individuals who might have unusual health conditions.	✓ 0.25	Since you are just looking for unusual conditions instead of a particular disease, this is a good application of anomaly detection.
<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 checked="" type="radio"/> $\mu_1 = -3, \sigma_1^2 = 4$	<div>✓</div> 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"/> $\mu_1 = -6, \sigma_1^2 = 4$		
<input type="radio"/> $\mu_1 = -6, \sigma_1^2 = 2$		
<input type="radio"/> $\mu_1 = -3, \sigma_1^2 = 2$		
Total	1.00 / 1.00	

Question 3

Suppose you have trained an anomaly detection system that flags anomalies when $p(x)$ is less than ε , and you find on the cross-validation set that it has too many false positives (flagging too many things as anomalies). What should you do?

Your Answer	Score	Explanation
<input checked="" type="radio"/> Decrease ε	<div>✓</div> 1.00	By decreasing ε , you will flag fewer anomalies, as desired.

☐ Increase ϵ

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 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"/> $x_3 = x_1 \times x_2^2$		
<input type="radio"/> $x_3 = x_1^2 \times x_2^2$		
<input type="radio"/> $x_3 = x_1 \times x_2$		
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"/> If you are developing an anomaly detection system, there is no way to make use of labeled data to improve your system.	✓ 0.25	Labeled data are usefull in cross-validation and testing for evaluating the system and setting the parameter ϵ .
<input type="checkbox"/> 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.	✓ 0.25	Classification accuracy is a poor metric because of the skewed classes in the cross-validation set (almost all examples are negative).
<input checked="" type="checkbox"/> 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.	✓ 0.25	These are good features, as they will lie outside the learned model, so you will have small values for $p(x)$ with these examples.
Total	1.00 / 1.00	