Feedback — XV. Anomaly Detection

Help Center

You submitted this quiz on **Fri 10 Apr 2015 1:17 PM CEST**. You got a score of **5.00** out of **5.00**.

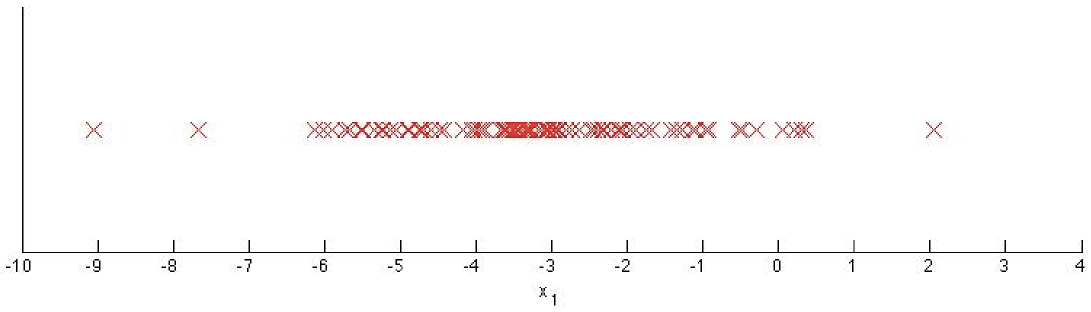
Question 1

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

Your Answer		Score	Explanation
☐ 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.
☑ 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.
☐ 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.
☑ 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 appliation of anomaly detection.
Total		1.00 / 1.00	

Question 2

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?

Your Answer		Score	Explanation
○ $\mu_1 = -3, \sigma_1^2 = 2$			
◉ $\mu_1 = -3, \sigma_1^2 = 4$	✔	1.00	This is correct, as the data are centered around -3 and tail most of the points lie in [-5, -1].
○ $\mu_1 = -6, \sigma_1^2 = 2$			
○ $\mu_1 = -6, \sigma_1^2 = 4$			
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 mis-flagging far too many good transactions as fradulent. What should you do?

Your Answer		Score	Explanation
○ Increase ε			
◉ Decrease ε	✔	1.00	By decreasing ε, you will flag fewer 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. 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:

Your Answer		Score	Explanation
$x_3 = x_1^2 \times x_2$			
$x_3 = x_1 \times x_2$			
$x_3 = x_1 + x_2$			
$x_3 = \frac{x_1}{x_2}$	✔	1.00	This is correct, as it will take on large values for anomalous examples and smaller values for normal examples.
Total		1.00 / 1.00	

Question 5

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

Your Answer		Score	Explanation
☑ 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.	✔	0.25	We want to model "normal" examples, so we only use negative examples in training.
☐ In a typical anomaly	✔	0.25	It is the reverse: we have many normal

detection setting, we have a large number of anomalous examples, and a relatively small number of normal/non-anomalous examples.			examples and few anomalous examples.
☑ 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.
☐ 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).
Total		1.00 / 1.00	