Anomaly Detection

Cuestionario, 5 preguntas

1 point

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

- [ ] Given data from credit card transactions, classify each transaction according to type of purchase (for example: food, transportation, clothing).
- [x] From a large set of primary care patient records, identify individuals who might have unusual health conditions.
- [ ] From a large set of hospital patient records, predict which patients have a particular disease (say, the flu).
- [x] In a computer chip fabrication plant, identify microchips that might be defective.

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 mis-flagging far too many good transactions as fradulent. What should you do?

- [ ] Increase ε
- [x] Decrease ε

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:

- [ ] $x_3 = x_1^2 \times x_2$
- [ ] $x_3 = x_1 + x_2$
- [ ] $x_3 = x_1 \times x_2$
- [x] $x_3 = \frac{x_1}{x_2}$

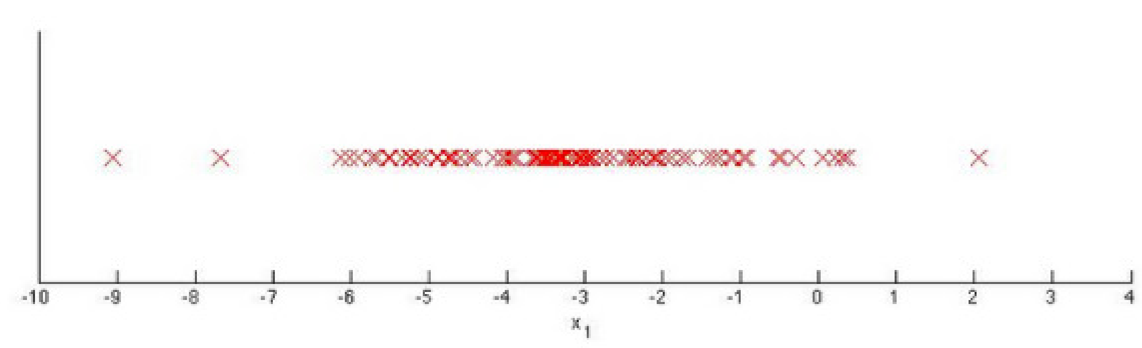
1 point

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

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

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 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?

- [x] $\mu_1 = -3, \sigma_1^2 = 4$
- [ ] $\mu_1 = -6, \sigma_1^2 = 4$
- [ ] $\mu_1 = -3, \sigma_1^2 = 2$
- [ ] $\mu_1 = -6, \sigma_1^2 = 2$

- [ ] Entiendo que enviar trabajo que no es mío resultará en la desaprobación permanente de este curso y la desactivación de mi cuenta de Coursera.

Obtén más información sobre el Código de Honor de Coursera

Ingresa tu nombre legal

Submit Quiz