Anomaly Detection

Total points 5

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

- [ ] Given data from credit card transactions, classify each transaction according to type of purchase (for example: food, transportation, clothing).
- [ ] Given an image of a face, determine whether or not it is the face of a particular famous individual.
- [x] From a large set of primary care patient records, identify individuals who might have unusual health conditions.
- [x] Given a dataset of credit card transactions, identify unusual transactions to flag them as possibly fraudulent.

2. 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 negatives (failing to flag a lot of anomalies). What should you do? **1 point**

- ○ Decrease ε
- ◉ Increase ε

3. Suppose you are developing an anomaly detection system to catch manufacturing defects in airplane engines. You model uses **1 point**

$$p(x) = \prod_{j=1}^{n} p(x_j; \mu_j, \sigma_j^2).$$

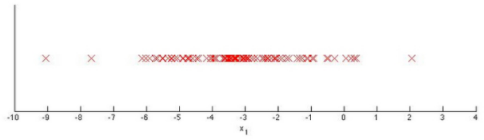
You have two features x_1 = vibration intensity, and x_2 = heat generated. Both x_1 and x_2 take on values between 0 and 1 (and are strictly greater than 0), and for most "normal" engines you expect that $x_1 \approx x_2$. One of the suspected anomalies is that a flawed engine may vibrate very intensely 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:

- ○ $x_3 = \frac{1}{x_1}$
- ○ $x_3 = x_1 + x_2$
- ○ $x_3 = \frac{1}{x_2}$
- ◉ $x_3 = \frac{x_1}{x_2}$

4. Which of the following are true? Check all that apply. **1 point**

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

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: **1 point**



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?

- ◉ $\mu_1 = -3, \sigma_1^2 = 4$
- ○ $\mu_1 = -6, \sigma_1^2 = 4$
- ○ $\mu_1 = -3, \sigma_1^2 = 2$
- ○ $\mu_1 = -6, \sigma_1^2 = 2$

Coursera Honor Code Learn more

- [x] I, **Vamshi Nenavath**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

Submit **Save draft**