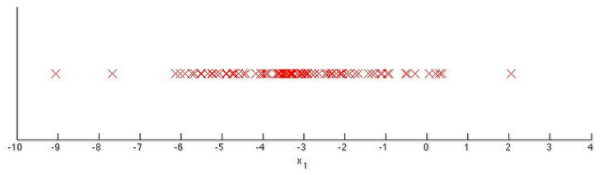
* For which of the following problems would anomaly detection be a suitable algorithm?
* **From a large set of primary care patient records, identify individuals who might have unusual health conditions.**
* **In a computer chip fabrication plant, identify microchips that might be defective.**
* Suppose you have trained an anomaly detection system for fraud detection + your system that flags anomalies when p(x) is less than ε, and you find on the CV set that it is missing many fraudulent transactions (failing to flag them as anomalies). What should you do?
* **Increase ε**
* Suppose you are developing an anomaly detection system to catch manufacturing defects in airplane engines. You model uses



You have 2 features x1 = vibration intensity, + x2 = heat generated. Both x1 + x2 take on values between 0-1 (+ are strictly > 0), + for most "normal" engines you expect x1 ≈ x2. 1 of the suspected anomalies is a flawed engine may vibrate very intensely even w/out generating much heat (large x1, small x2), even though the particular values of x1 + x2 may not fall outside their typical ranges of values. What additional feature x3 should you create to capture these types of anomalies:

* **x3 = x1 / x2**
* Which of the following are true? Check all that apply.
* **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.**
* **In a typical anomaly detection setting, we have a large number of normal/non-anomalous examples, and a relatively small number of anomalous examples.**
* **When evaluating an anomaly detection algorithm on CV set (containing some positive + some negative examples), classification accuracy is usually NOT a good evaluation metric to use**
* **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.**
* You have a 1-D dataset {x(1),…,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 μ1 and σ21 to this dataset. Which of the following values for μ1 and σ21 might you get?

* **μ1 = −3, σ^2(1) = 4**