**Intuition behind Anomaly Detection**

|  |
| --- |
|  |
| Used for fraud detection on website detection, compare special case to average uses of webpage |

**Gaussian distribution**

|  |
| --- |
| Formula |
| Image result for formula of bell shaped curve |
| Effects of parameters |
|  |
| Implementation |
| Mu = (1/m) .\* mean(X, 1); (vector of (1\*n))    X = X(actual)  For i from 1 to m:  X(i, :) = ((X(i, :) – mu) .^ 2)  Sigma = (1/m) .\* sum(X, 1) size(1\*n)    (1/(sqrt(2\*pi) .\* sigma)) \* exp(-1 .\* (for i-m: X(i, :) – mu) .^ 2 ./ (2 .\* sigma .^ 2)) |

**Application: Develop/Evaluating Anomaly Detection & choosing features to use**

|  |
| --- |
| **Assume**:   1. Data are labeled but single variate || normal = 0. Anomalous = 1 2. Training set are normal examples and not anomalous 3. Cross validation set and test set can contain anomalous example   **Note**:   1. Typical range of #anomalous are 2-50   2. Training set are fully unlabeled, but as predicting test set, the normal will be labeled as 0 and anomaly will be labeled as 1  3. Evaluate the anomaly detection algorithm by using evaluation metrics since y = 1 are skewed classes. (ex. to choose threshold epsilon that maximize F1 score) |
| **Question:**  Sometimes supervised learning can also classify anomaly, so when should anomaly detection be used and when should supervised learning be used? |
| **Note for deciding:**  Before choosing which algorithm to use, plot the histogram of the data to make sure it is a bell shaped curve **hist() command in Octave**  In this case, it is a better to use a anomaly detection algo.  Else use a supervised learning may give a more accurate prediction |
|  |
| **Manipulating features for anomaly detection** |
| To make the hist() curve more gaussian, play around with the features.  **Example:** |
|  |
| **Come up with features for anomaly detection** |
| 1. Look at missed classification of anomalous example, and compare it to the normal data to come up with a new feature that distinguish the data from the normal examples around it 2. Chose features that has extreme values in the event of an anomaly |
|  |

**Multivariate Gaussian Distribution**

|  |
| --- |
| MGD allows  to be  **Constrains when using MGD:**   1. Computationally expensive 2. Must have m > n, else non-invertable |
| Numerator = matrix of (n\*m)(m\*n) = (n\*n) |
|  |