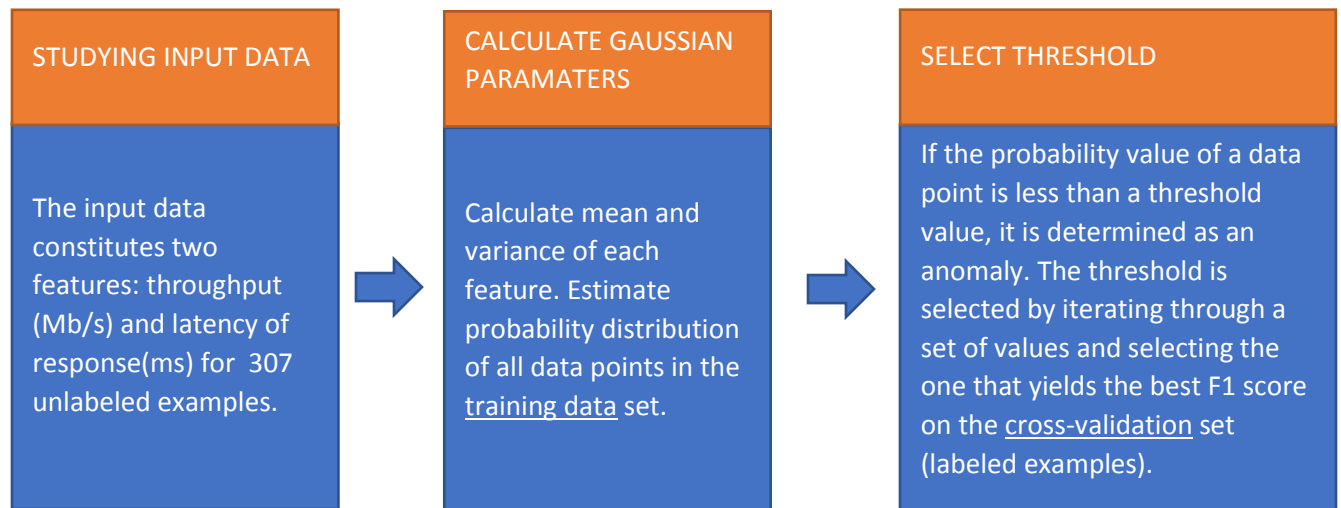


## Anomaly detection algorithm

Objective: To build an algorithm to detect failing servers (anomalous behaviour) in a network of computers

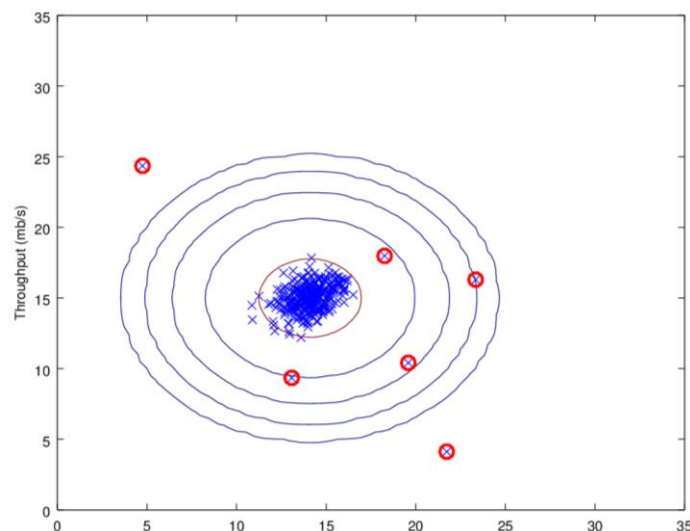
Approach: Fit a Gaussian model to data, and detect data points with low probabilities as anomalies.

Methodology:



Result:

A model that can detect anomalous behavior in a network of computers has been built. This model is not constrained by the dimension of the data. Hence the same model can be trained on datasets with additional features and can be used to detect anomalies in high dimensional dataset. The detected anomalies are marked in red as shown below.



### Appendix:

This model was built in Octave 4.0.3. It can be run in MATLAB, as well.

Note: 30 percent of attached program files were provided by “Coursera”

### Attached files:

ex8 – Anomaly detection model

estimateGaussian-Function to calculate mean, variance, and estimate Gaussian distribution

SelectThreshold-Function to select threshold value

### Formulas:

$$\text{Mean}, \mu_i = \frac{1}{m} \sum_{j=1}^m x_i^j$$

$$\text{Variance}, \sigma_i^2 = \frac{1}{m} \sum_{j=1}^m (x_i^j - \mu_i)^2$$

$$\text{Probability distribution}, \quad p(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$F1 \text{ score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}}$$

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}$$