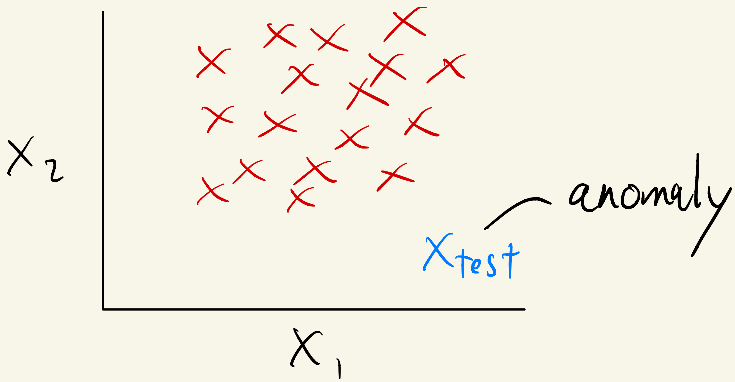
Anomaly Detection Example:

May 30, 2021

e.g. x_1 = heat generated
x_2 = vibration intensity

Dataset: $\{x^{(1)}, x^{(2)} \ldots x^{(m)}\}$

New Engine: x_{test}

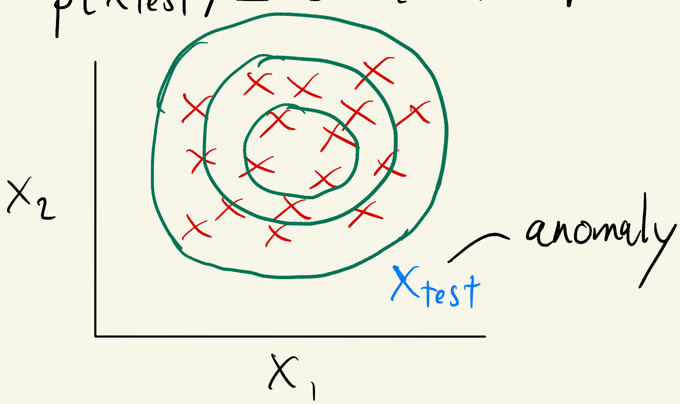


So how do we determine if x_{test} is anomalous?

Create probability function $p(x)$.

↳ $p(x_{test}) < \varepsilon \rightarrow$ flag anomaly

$p(x_{test}) \geq \varepsilon \rightarrow$ normal



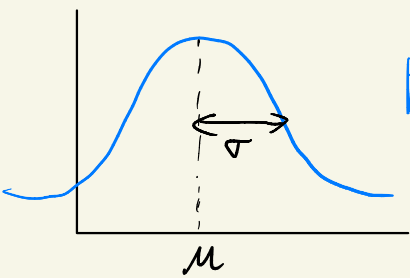
✱ What is the probability that x is normal?

Gaussian (Normal) Distribution:

Let $x \in \mathbb{R}$. If x is a distributed Gaussian w/ mean μ and variance σ^2:

$$x \sim \mathcal{N}(\mu, \sigma^2) \qquad \mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)} \qquad \sigma^2 = \frac{1}{m}\sum_{i=1}^{m}\left(x^{(i)} - \mu\right)^2$$

"distributed as"



$$p(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

* Gaussian function

Density Estimation:

$$x_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$$
$$x_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$$
$$x_3 \sim \mathcal{N}(\mu_3, \sigma_3^2)$$

* $\prod_{i=1}^{n} i = 1 \times 2 \times \ldots n$

$$p(x) = p(x_1; \mu_1, \sigma_1^2)\, p(x_2; \mu_2, \sigma_2^2) \ldots p(x_n; \mu_n, \sigma_n^2)$$
$$= \prod_{j=1}^{n} p(x_j; \mu_j, \sigma_j^2)$$

Anomaly Detection Algorithm:

1. Choose features x_i that you think are indicative of anomalous examples.

2. Fit parameters $\mu_1, \ldots, \mu_n, \sigma_1^2, \ldots, \sigma_n^2$

$$\rightarrow \mu_j = \frac{1}{m} \sum_{i=1}^{m} x_j^{(i)} \longrightarrow \mu \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} = \frac{1}{m} \sum_{i=1}^{m} x^{(i)}$$

$$\rightarrow \sigma_j^2 = \frac{1}{m} \sum_{i=1}^{m} \left(x_j^{(i)} - \mu_j\right)^2$$

3. Given new example x, compute $p(x)$.

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

$$= \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{(x_j - \mu_j)^2}{2\sigma_j^2}\right)$$

Anomaly if $p(x) < \varepsilon$.