Problem Motivation

Anomaly Detection ex. 이상치 탐지

QA (Quality Assurance testing)

Aircraft engine features:

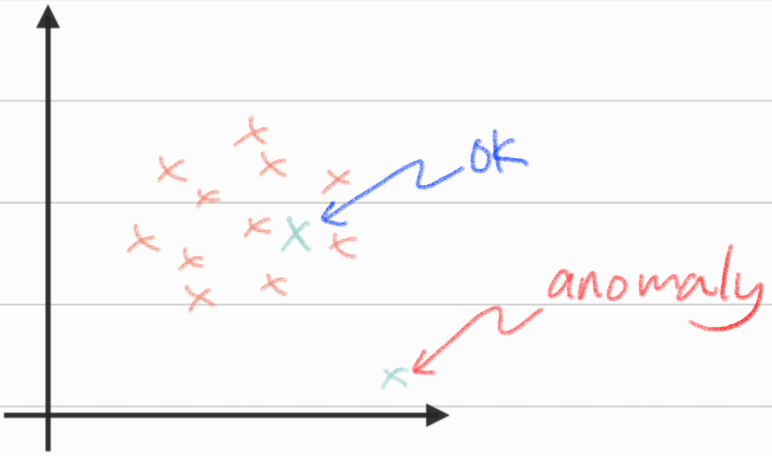
x_1 = heat generated

x_2 = vibration intensity

⋮

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

New engine: x_{test}



Density estimation

$p(x)$: model of the probability of x

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

$p(x_{test}) \geq \varepsilon \rightarrow$ okay ex. ① Fraud detection:

$x^{(i)}$ = features of user i's activities

Model $p(x)$ from data

Identify unusual users by checking which have $p(x) < \varepsilon$

② Manufacturing

③ Monitoring computers in a data center

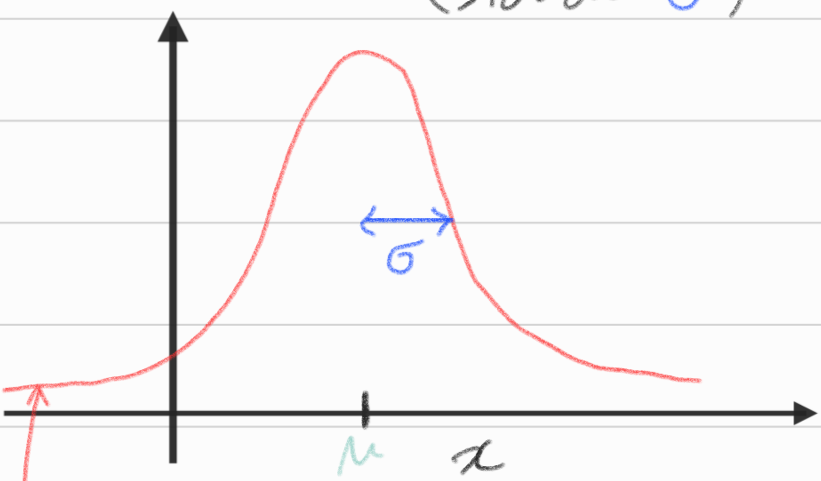
$x^{(i)}$ = features of machine i

Gaussian Distribution (Normal Distribution)

Say $x \in \mathbb{R}$,

If x is a distributed Gaussian with mean μ, variance σ^2 $\qquad x \sim \mathcal{N}(\mu, \sigma^2)$

(std. dev. σ)



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

Parameter estimation

Dataset: $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$ $\quad x^{(i)} \in \mathbb{R}$

$$x^{(i)} \sim N(\mu, \sigma^2)$$

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)} \qquad \sigma^2 = \frac{1}{m}\sum_{i=1}^{m}(x^{(i)} - \mu)^2$$

Algorithm

Training set: $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$ $\quad x^{(i)} \in \mathbb{R}$

$p(x)$

$$= p(x_1; \mu_1, \sigma_1^2)\, p(x_2; \mu_2, \sigma_2^2)\, p(x_3; \mu_3, \sigma_3^2) \cdots p(x_n; \mu_n, \sigma_n^2)$$

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

곱셈

Anomaly detection algorithm

1. Choose features x_i that you think might be indicative of anomalous examples.
2. Fit parameters $\mu_1, \dots, \mu_n, \sigma_1^2, \dots, \sigma_n^2$

$$\mu_j = \frac{1}{m}\sum_{i=1}^{m} x_j^{(i)}$$ (average value of feature j)

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

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

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

곱셈

모든 feature (n)에 대한 p(x) 곱!

Anomaly if $p(x) < \varepsilon$