Anomaly detection

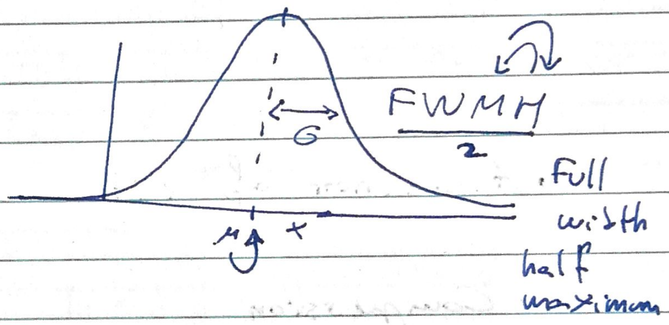
- Gaussian distribution

Say $x \in \mathbb{R}$. $x \sim N(\mu, \sigma^2)$

variance

distributed as   Normal distribution   mean

$$\Rightarrow \mathbb{P}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$



$$\mu \equiv \text{mean} = \frac{1}{m} \sum_{i=1}^{m} x^i$$

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

- Given a training set $X = \{x^1, \dots x^m\}$

$$\Rightarrow \mathbb{P}(X) = \prod_{j=1}^{n} \mathbb{P}(x_j; \mu_j, \sigma_j^2) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma_j} e^{-\frac{(x_j-\mu_j)^2}{2\sigma_j^2}}$$

③ Example of distributing datasets

Lets say we have 10K items flawed as good, and 20 flawed as anomalous ($y = 1$)

⇒ Training set 6K ($y=0$)
Cross validation set 2K ($y=0$), 10 anomalous ($y=1$)
Test set 2K ($y=0$), 10 anomalous ($y=1$)

- Fit model $p(x)$ on training set

- On CV / Test, predict $y = \begin{cases} 1 & \text{if } p(x) < \varepsilon \text{ anomaly} \\ 0 & \text{if } p(x) \geq \varepsilon \text{ normal} \end{cases}$

ε = benchmark

- Possible evaluation metrics
  - TP, TN, FP, FN
  - Precission / recall
  - F_1-score

We can use CV ~~test~~ set to find the proper ε

Anomaly detection VS Supervised learning

- The distribution is highly skewed. Maybe positive are large and negatives just a few
- Many "different types" of anomalies. The algorithm "won't learn" what the anomalies look like for future predictions

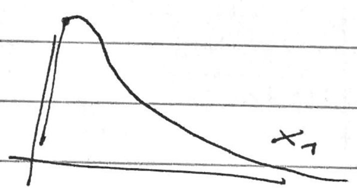
- Positive and negative cases are more or less equally distributed
- Enoug (+) cases for the algorithm to get a sense of what (+) cases are like and future ones

- Fraud detection $\rightarrow y = 1$
- Manufacturing
- Monitoring machines in data center

- Email spam classification
- Weather predictions
- Cancer classification...

- Transformations:

If x_1 like for instance



we can modify it like

$$x_1 \rightarrow \log(x_1)$$
$$\searrow \log(x_1+1)$$
$$\searrow (x_1)^{1/3}$$

whatever that gives us the Gaussian shape

• Multivariate Gaussian distribution

$$\mathbb{P}(x; \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

; $\Sigma \in \mathbb{R}^{n \times n}$

• Examples $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ ← centred in 0 circle radious 1

$\Sigma = \begin{pmatrix} a & c \\ c & b \end{pmatrix}$

• If $a, c > 0 \Rightarrow$ elipse falling to the right

• If $a, c < 0 \Rightarrow$ elipse falling to the left

Al final son autovectores multiplicando en la ecuación general de una elipse

$$\mu = \frac{1}{m} \sum_{i=1}^{m} x^i \quad ; \quad \Sigma = \frac{1}{m} \sum_{i=1}^{m} (x^i - \mu)(x^i - \mu)^T$$

It looks like in Multivariate Gaussian Σ is a diagonal matrix

@

Original model	Multivariate Gaussian
$\mathbb{P}(x;\mu,\sigma^2) = \prod \mathbb{P}(x_i;\mu_i,\sigma_i^2)$	$\mathbb{P}(x;\mu,\Sigma) = \frac{e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}}{(2\pi)^{n/2}\lvert\Sigma\rvert^{1/2}}$
• Manually create features to capture anomalies	• Automatically captures correlations between features
• Computationally cheaper	• Computationally expensive
• OK if training set is small	• Must have $m \geq n$ or else Σ is not invertible