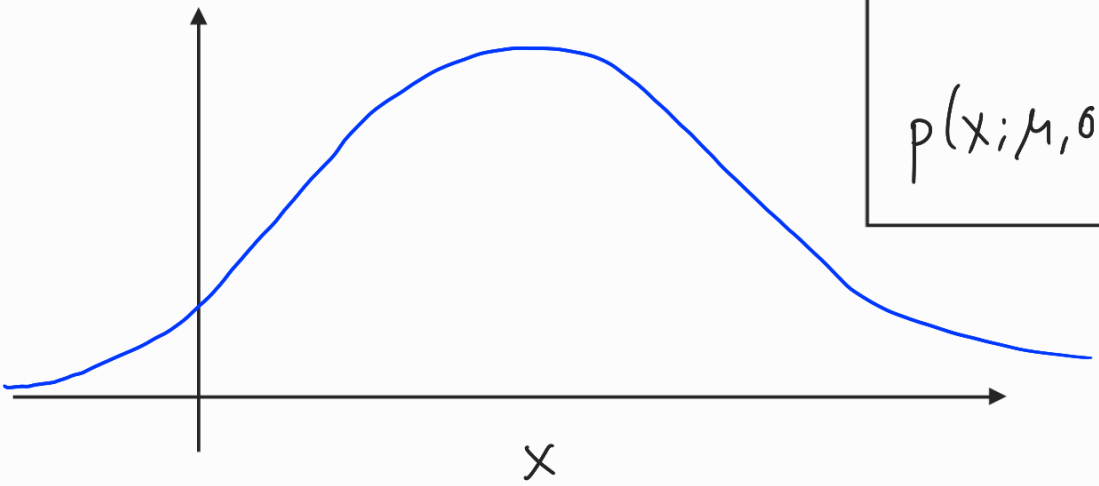
# Gaussian distribution:

$X \in \mathbb{R}$ with mean $= \mu$ and variance $= \sigma^2$;

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

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




$\mu \Rightarrow$ sets the center of the distribution

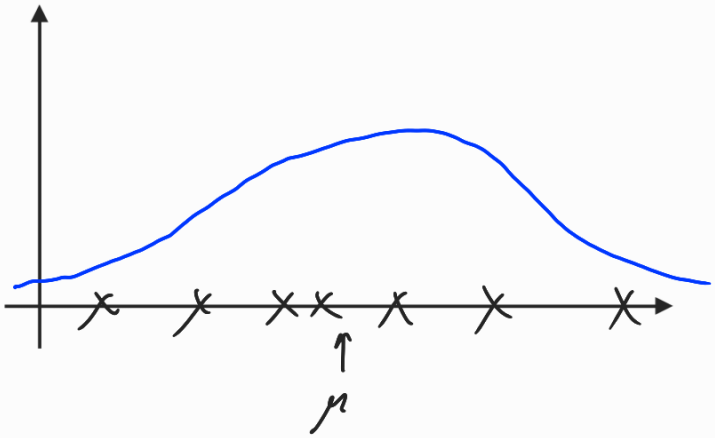
$\sigma^2 \Rightarrow$ height and width $\nearrow$ smaller : $\uparrow$ $\downarrow$

$\searrow$ bigger : $\downarrow$ $\uparrow$

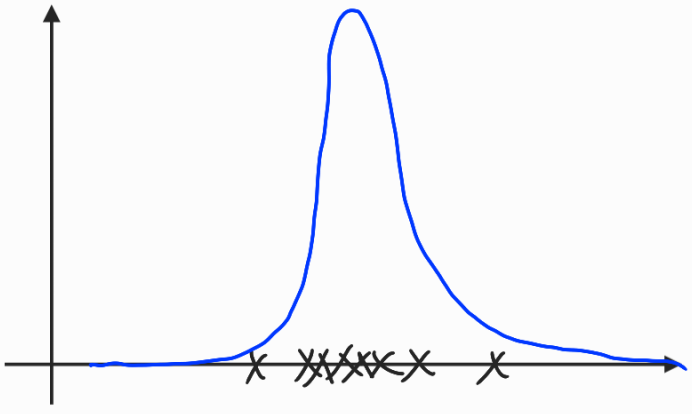
height | width

That's like that, because the AREA is ALWAYS the same.

As the $\sigma^2$ grows:




As it dicreases:



Real values are more far from $\mu$.

## Density estimation:

Training set: $\{x^{(1)}, \ldots, x^{(m)}\}$ $\forall\ x \in \mathbb{R}^n$

With its corresponding $\{\{\mu_1, \sigma_1^2\}, \ldots, \{\mu_m, \sigma_m^2\}\}$

$$p(x) = \prod_{i=1}^{m} p(x_i; \mu_i, \sigma_i^2) = \prod_{i=1}^{m} \frac{e^{\left(-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right)}}{\sqrt{2\pi}\,\sigma_i}$$

## Anomaly detection algorithm:

1. Choose features that may be anomalous
2. Fit parameters $\mu$ and $\sigma$
3. Given new $x$, compute $p(x) < \varepsilon$

Example:

- 1000 good
- 20 anomalous

| | Good | Anomalous |
|---|---|---|
| Training set: | 6000 | 0 |
| Cross validation: | 2000 | 10 |
| Test set: | 2000 | 10 |

How to evaluate model?

- Accuracy: NO, very skewed
- F1 score: YES

Choose $\varepsilon$: with cross validation, iteratively maximize F1.

| | Normal | Anomalous | |
|---|---|---|---|
| Anomaly detection | ↑ ↑ ↑ | ↑ | many different types of anomalies |
| Supervised learning | ↓ ↓ ↓ | ↓ | |

Multivariate Gaussian distribution: used when we need more complex functions than just  . Being able then to create ellipses with $\Sigma^{-1}$ and moving them with $\mu$: $\forall \mu \in \mathbb{R}^n$, $\Sigma^{-1} \in \mathbb{R}^{n \times n}$

1. Fit model $p(x)$

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

2. Given new $x$, compute:

$$p(x) = \frac{e^{\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)}}{(2\pi)^{n/2} |\Sigma|^{1/2}}$$

$p(x) < \varepsilon$ : Anomaly

| Original model | Multivariate model |
|---|---|
| - Manually create features | - Automatically captures corrs. |
| - Cheaper | - Expensive |
| - OK if training size small (m small) | - Number samples > number features (if not, $\nexists \Sigma^{-1}$) |