Multivariate Gaussian Distribution:

$\underbrace{\mu}_{\text{mean}} \in R^{n\times 1}$, $\underbrace{\Sigma}_{\text{variance}} \in R^{n\times n} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \sigma_n^2 \end{bmatrix}$

show the correlation of the variables,

Density function:

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

It needs $m \geq 10n$ (or at least $m > n$) and no redundant features for Σ is invertible.

If the variables are independent with each other, they are all 0.

If they are negative correlated, then the values are negative.

How to get μ and Σ:

$$\mu = \frac{1}{m} \cdot \sum_{i=1}^{m} x^{(i)}$$

$$\Sigma = \frac{1}{m} \cdot \sum_{i=1}^{m} \underbrace{(x^{(i)}-\mu)}_{n\times 1}\underbrace{(x^{(i)}-\mu)^T}_{1\times n}$$

$\Downarrow$

$n \times n$

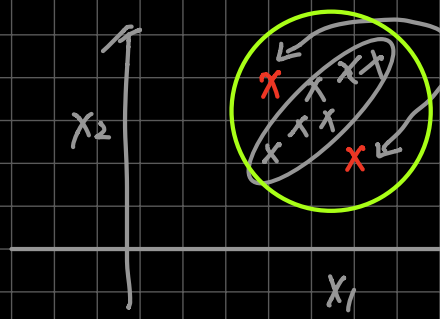
Rotation Direction: positive

How to apply in anomaly detection:

$$p(x^{(i)}) < \varepsilon \Rightarrow x^{(i)} \text{ is an anomaly data.}$$

What's the advantage of Multivariate Gaussian Distribution:

It consider the relationship between features, while the $\prod_{i=1}^{n} p(x_i^{(j)}) < \varepsilon \Rightarrow x^{(j)}$ is anormaly doesn't consider this. So it will be a nightmare when the data's distribution like this:



this two anomaly point will be consider normal because in both direction $p(x_i^{(j)})$ is not a small value.

Disadvantage:

① Calculation is expensive.

② It needs $m >> n$ (for Σ^{-1}) } If you don't want to run into these problems, just use the original model.

Comparison of Anomaly Detection and Supervised Learning:

	Anomaly Detection	Supervised Learning
Data features	Unbalanced ⇒ very small ① of positive examples	Balanced (little unbalanced)
Learning features	② Many different types of anomalies, hard for any algorithm to learn from positive examples. ③ future anomalies may look nothing like any of the	① Enough positive examples ② Simple type ③ future positive examples are similar to ones in training set.

	anomalies we've seen before	
Real Examples	Fraud Detection Aircraft engine Manufacturing Monitoring machines.	Other classification APPs, regression APPs.

△ How to choose features:

Choose the features that might take on unusually larger or small values in the event of an anomaly.

Note that the features can be original features or synthetic features. e.g $X_3 = \frac{X_1}{X_2^2}$