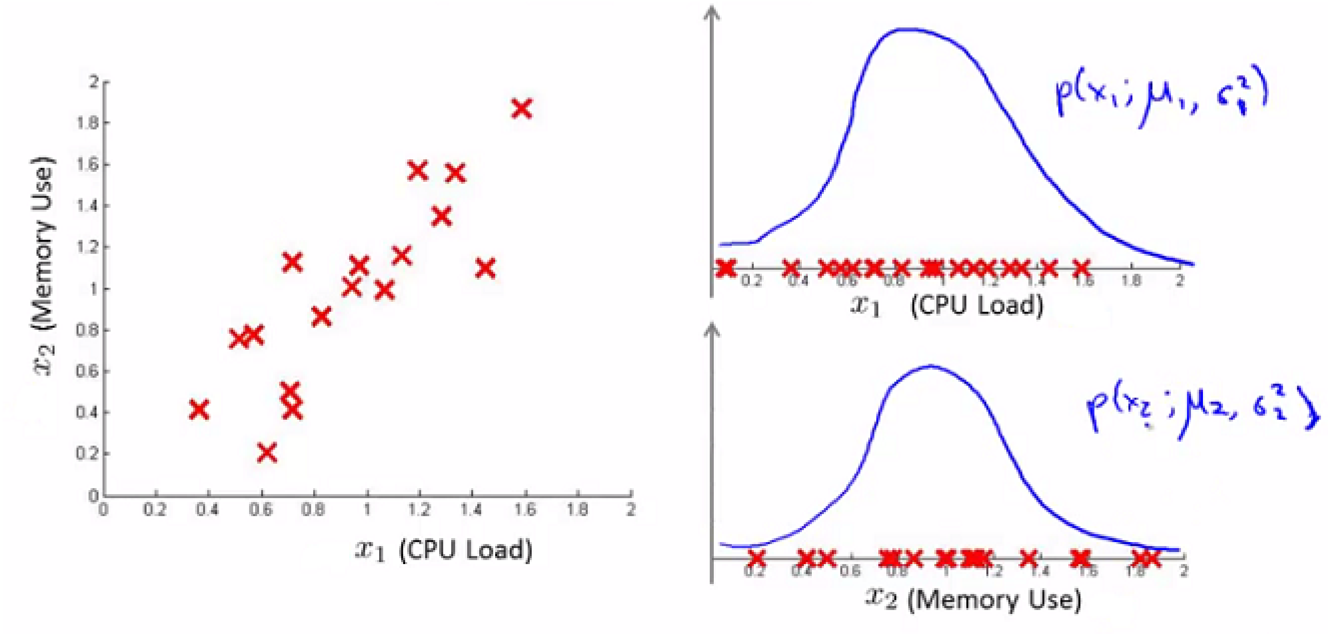
Multivariate Gaussian Distribution

The multivariate gaussian distribution is an extension of anomaly detection and may (or may not) catch more anomalies.

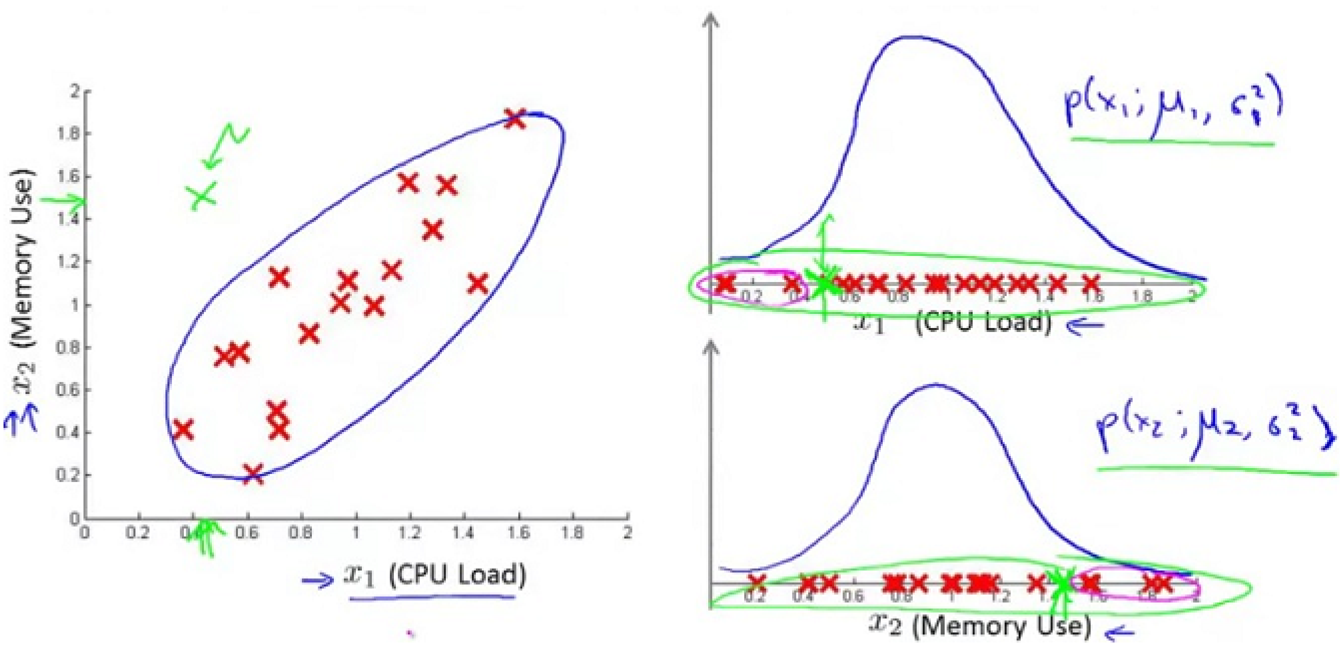
Motivating example: Monitoring machines in a data center

Let's say that we've a unlabeled data looks like this:



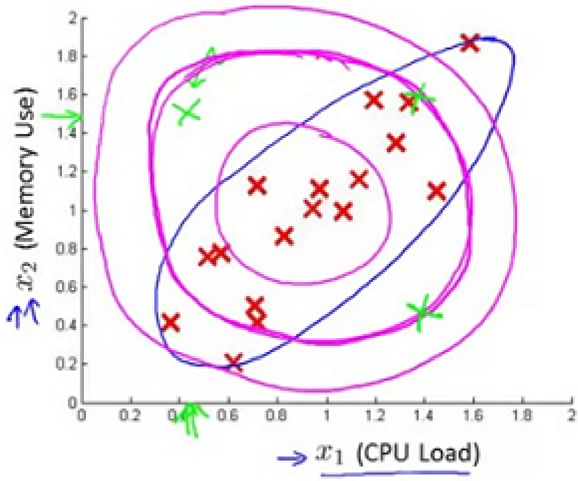
Say we can fit a Gaussian distribution to CPU load and memory use.

- Lets say in the test set we have an example which looks like an anomaly (e.g. x_1 = 0.45, x_2 = 1.5)
  - Looks like most of data lies in a region far away from this example
    - Here memory use is high and CPU load is low (if we plot x_1 vs. x_2 our green example looks miles away from the others)
  - It looks like, the CPU load, and the memory use, they sort of grow linearly with each other.



- Problem is, if we look at each feature individually they may fall within acceptable limits - the issue is we know we shouldn't don't get those kinds of values together
  - But individually, they're both acceptable

This is because our function makes probability prediction in concentric circles around the the means of both.



To fix this problem, we're going to develop a **modified version** of the anomaly detection algorithm, using something called the multivariate Gaussian distribution also called the **multivariate normal distribution**.

Multivariate Gaussian (Normal) Distribution

$x \in \mathbb{R}^n$. Instead of modeling $p(x_1), p(x_2), \ldots$ separately, we will model $p(x)$ all in one go. Our parameters will be: $\mu \in \mathbb{R}^n$ and $\Sigma \in \mathbb{R}^{n \times n}$ (covariance matrix).

- μ - Is an n-dimensional vector (where n is number of features)
- Σ - Is an $[n\ x\ n]$ matrix - the covariance matrix

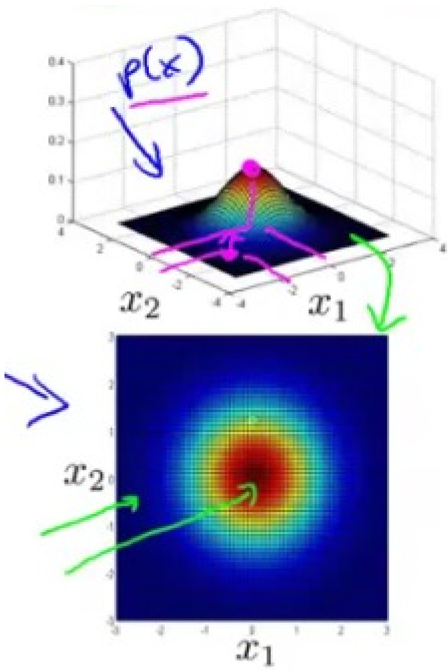
$$p(x; \mu, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} exp(-1/2(x-\mu)^T\Sigma^{-1}(x-\mu))$$

What does this $p(x)$ look like?: Examples

Let's take a two dimensional example (n = 2), with the following values of the parameters μ and Σ:

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$p(x)$ looks like this:



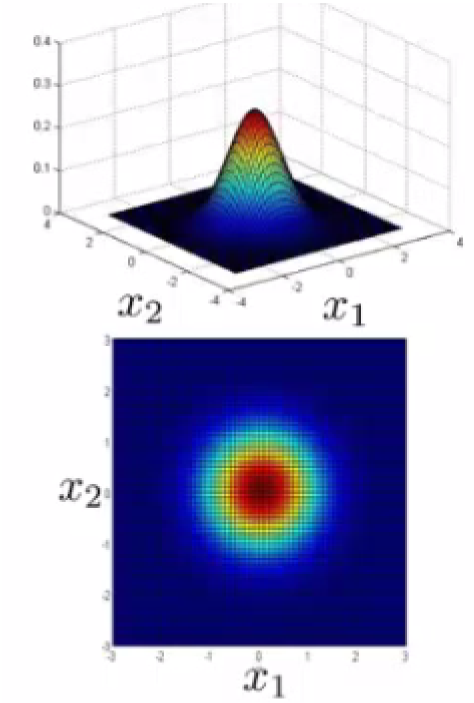
For inputs of x_1 and x_2 the height of the surface gives the value of $p(x)$

- With this distribution, we see that it faces most of the probability near (0,0) and then as we go out from (0,0) the probability of x_1 and x_2 goes down.

What happens if we change Sigma (Σ)?

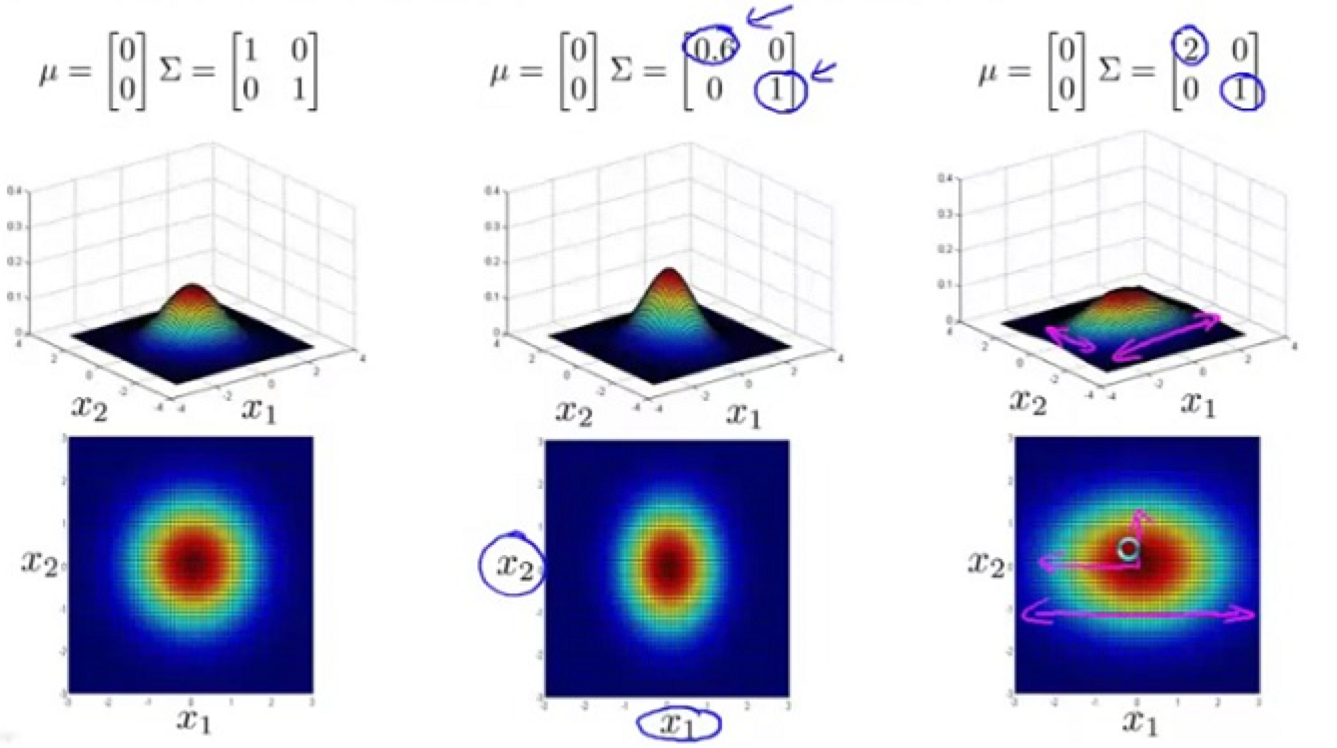
$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \ \Sigma = \begin{bmatrix} 0.6 & 0 \\ 0 & 0.6 \end{bmatrix}$$

and with these values in the parameters we obtain the following plot:



- Now the width of the bump decreases and the height increases

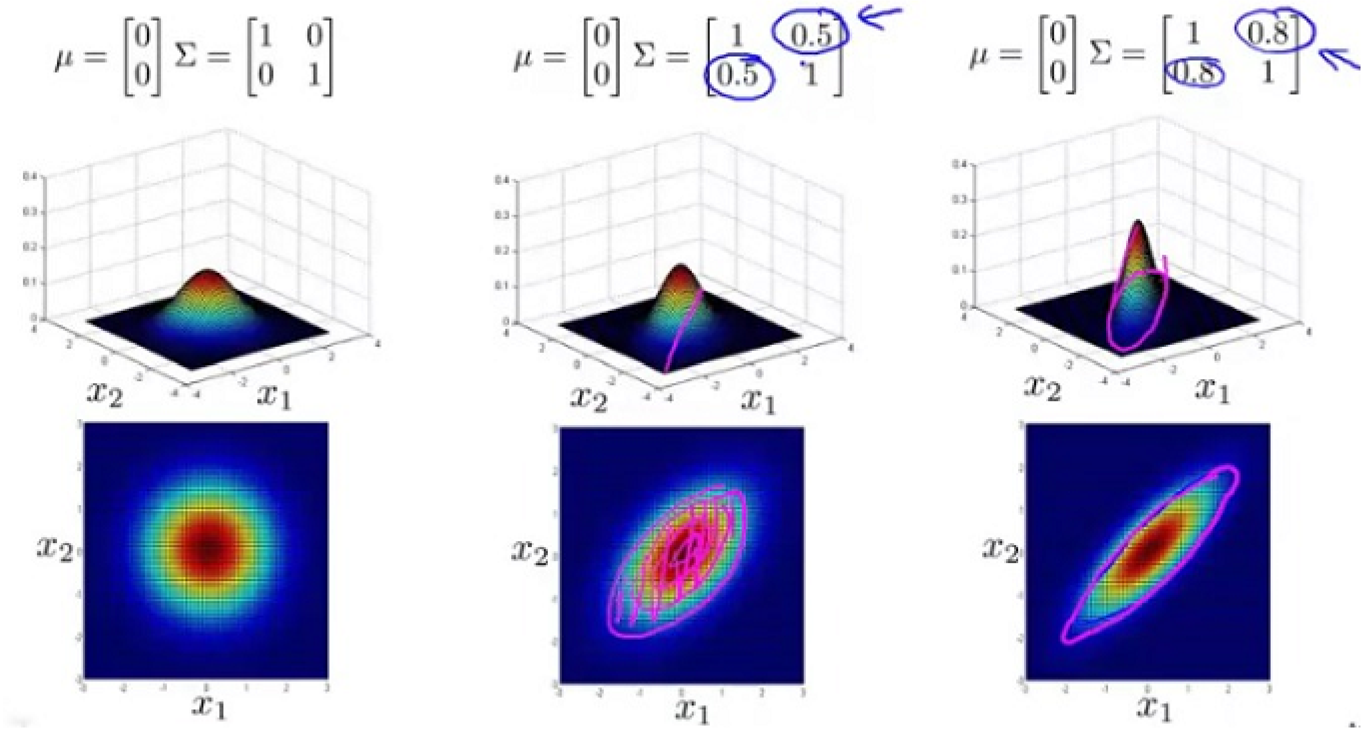
If we set sigma to be different values this changes the covariance matrix and we change the shape of our graph



Using these values we can, therefore, define the shape of this to better fit the data, rather than assuming symmetry in every dimension

- With this property we can use it to model correlation between data

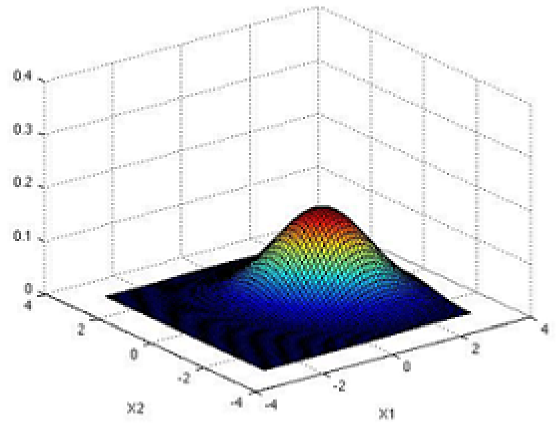
If we start to change the off-diagonal values in the covariance matrix we can control how well the various dimensions correlation

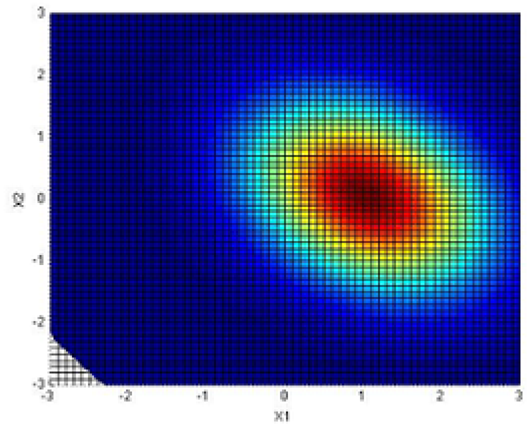


The important effect is that we can model oblong gaussian contours, allowing us to better fit data that might not fit into the normal circular contours.

- We can also make the off-diagonal values negative to show a negative correlation
- Varying Σ changes the shape, width, and orientation of the contours. Changing μ will move the center of the distribution.

Video Question: Consider the following multivariate Gaussian:





Which of the following are the μ and Σ for this distribution?

- $\mu = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}$
- $\mu = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & 0.3 \\ 0.3 & 1 \end{bmatrix}$

$$\mu = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & -0.3 \\ -0.3 & 1 \end{bmatrix}$$

- $\mu = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & -0.3 \\ -0.3 & 1 \end{bmatrix}$