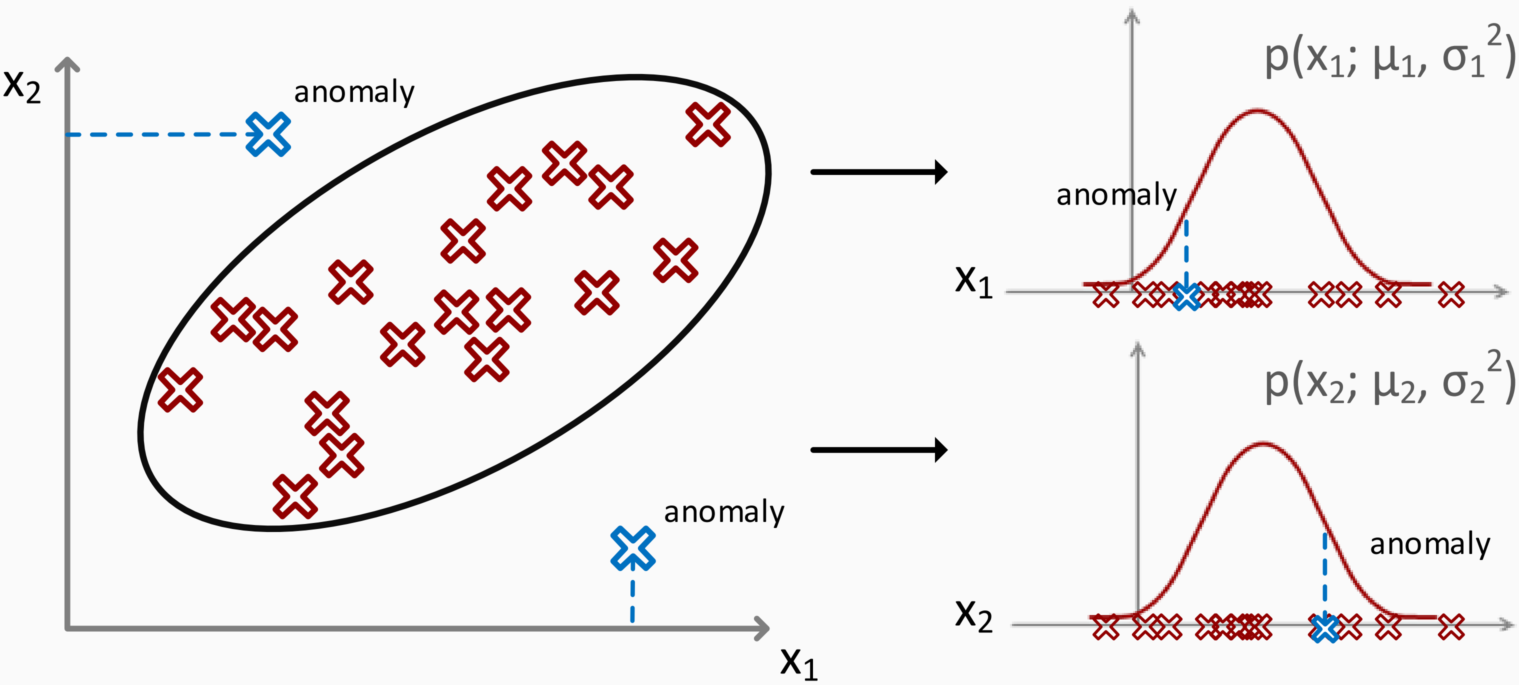
multivariate gaussian distribution

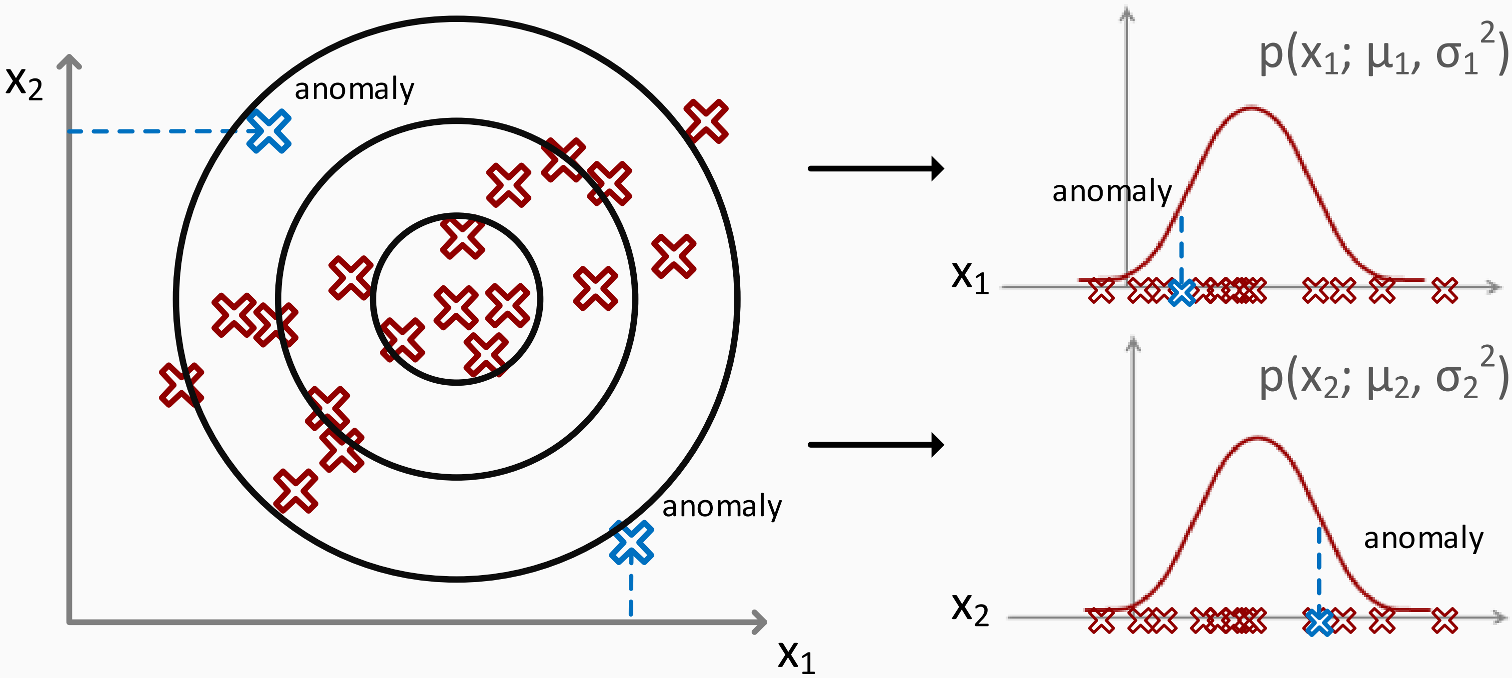
certain instances of single variable gaussian distribution might register examples amongst separtaley evaluated features as normal or non-anomalous. There are instances in which a multivariate gausian distribution would otherwise detect anomaly within a dataset that otherwise would have been misclassified elsewhere

multivariate gaussian model classification (left):



both variables plotted with a single variable gausian distribution algorithm would identify the anomolous probability regions not as the ellispe is shown in the above left illustration but as a serious of circles representative of the distributions' σ

single variable gaussian distribution (left):



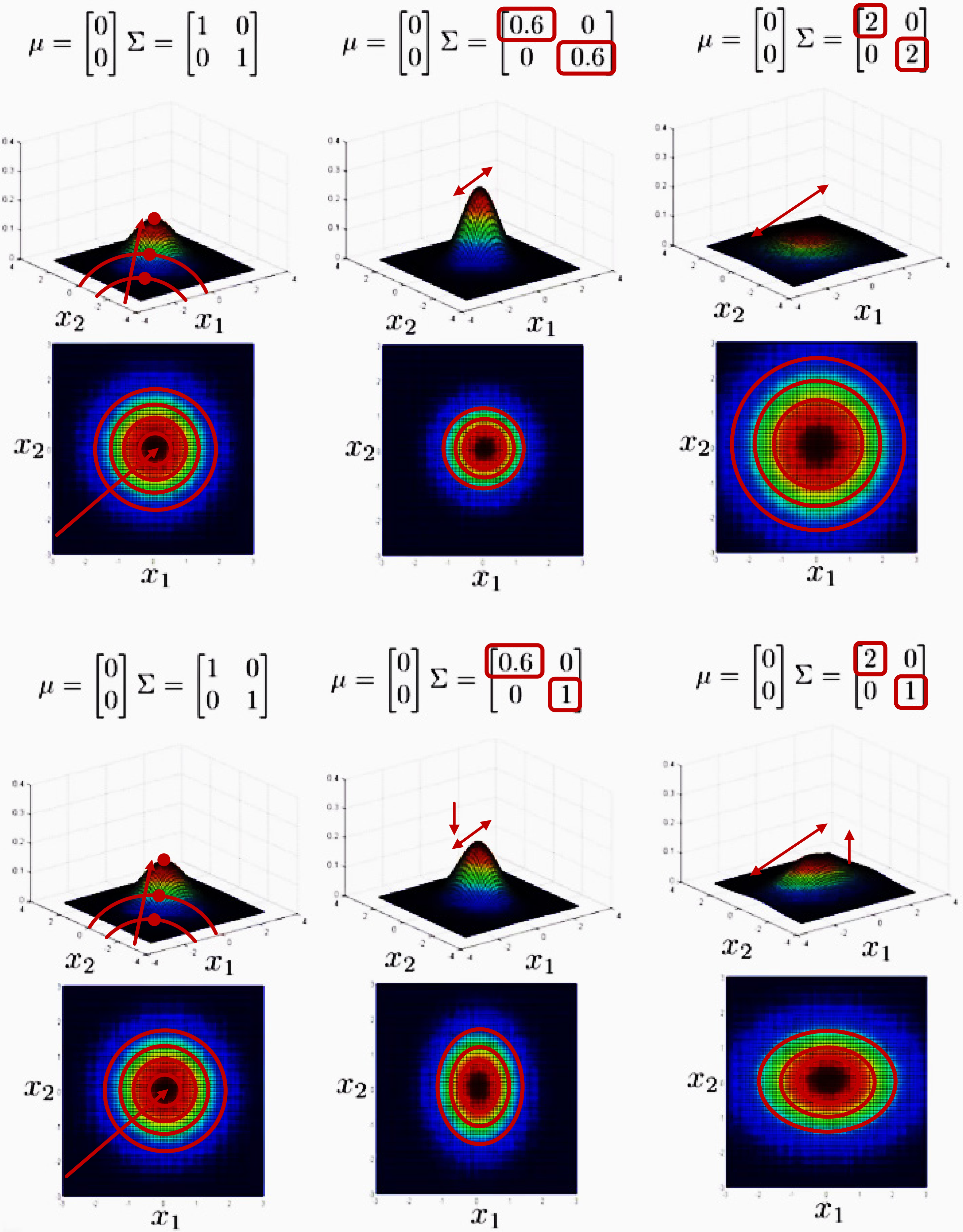
$x \in \mathbb{R}^n$, do not model multiple variables $p(x_1, x_2, \ldots, x_n)$ separately.

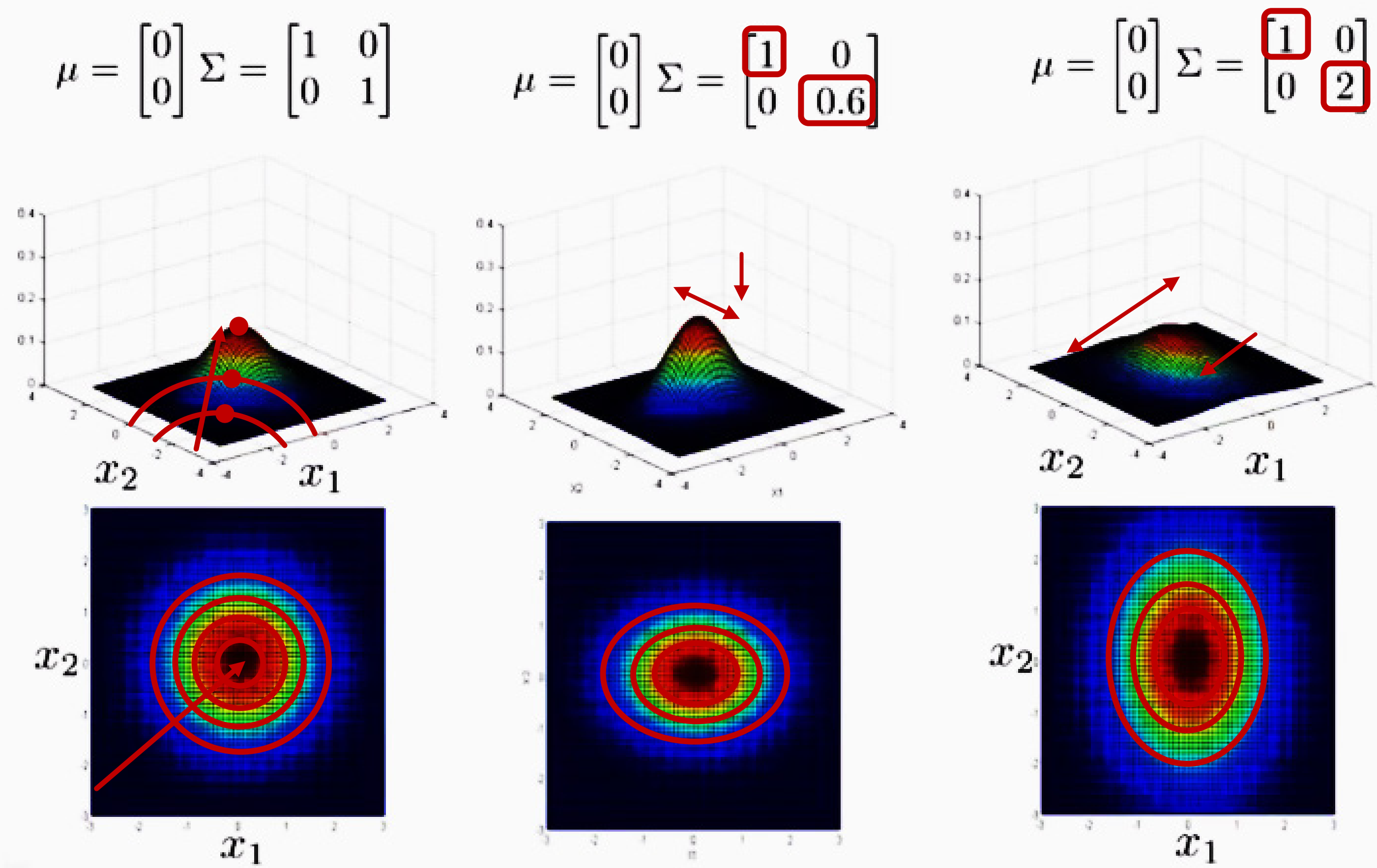
model $p(x)$ together at once.

parameters: $\mu \in \mathbb{R}^n, \Sigma \in \mathbb{R}^{n \times n}$ covariance matrix

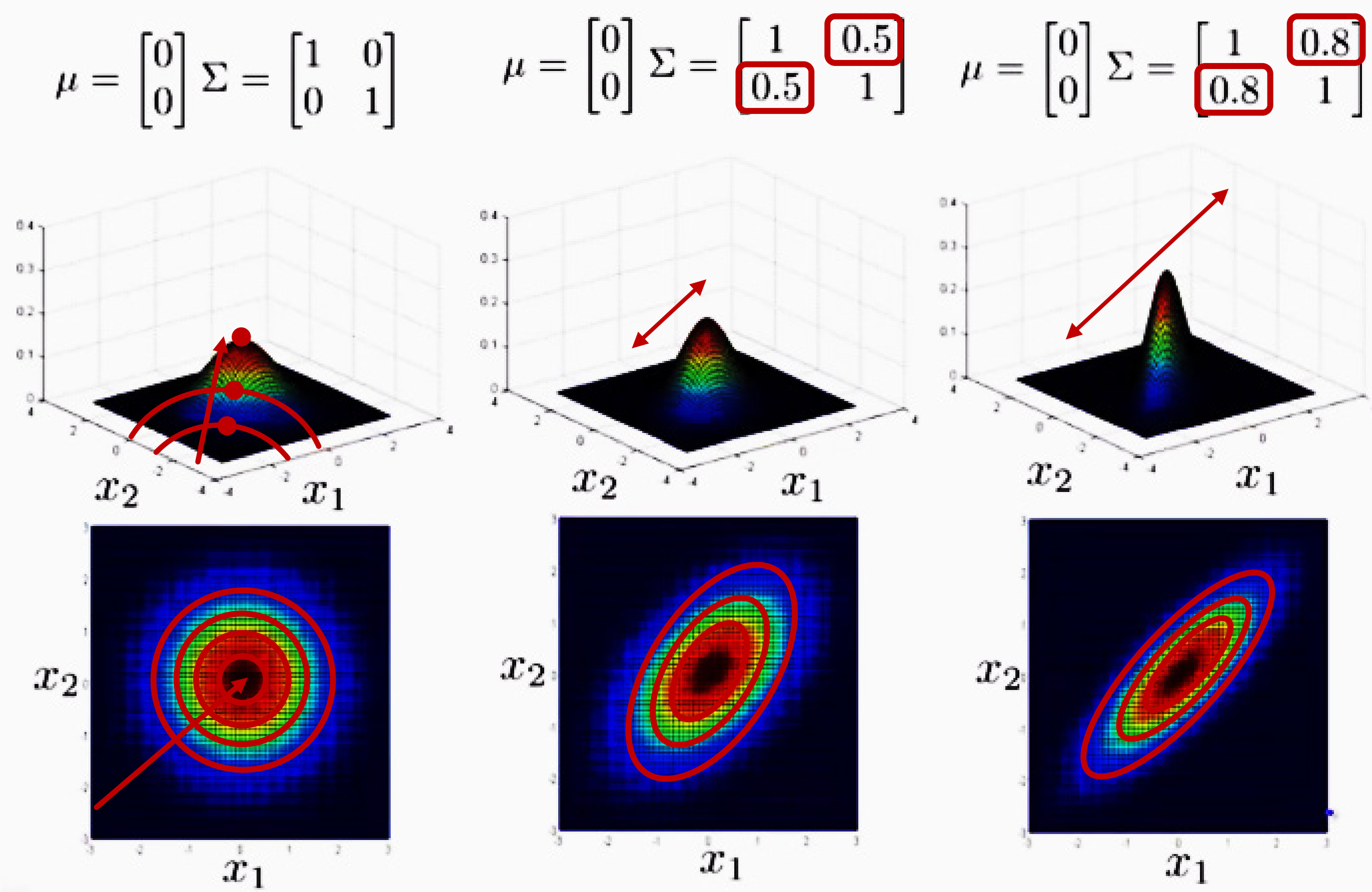
$$p(x; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} \, |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T \, \Sigma^{-1}(x-\mu)\right)$$

$|\Sigma| =$ determinant of Σ | det(Sigma) effects of transforming parameters in $p(x; \mu, \Sigma)$



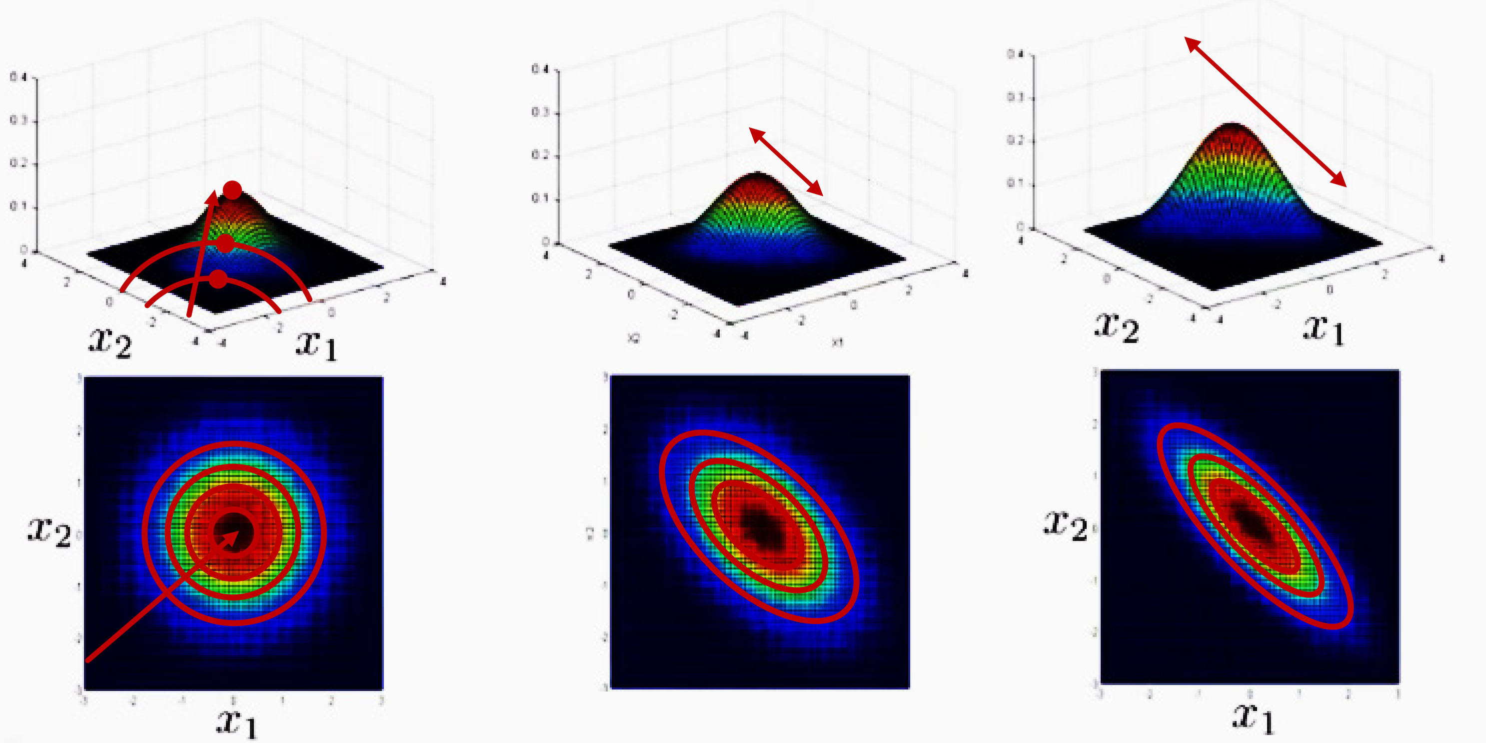


multivariate gaussian distribution can measure correlation amongst features:



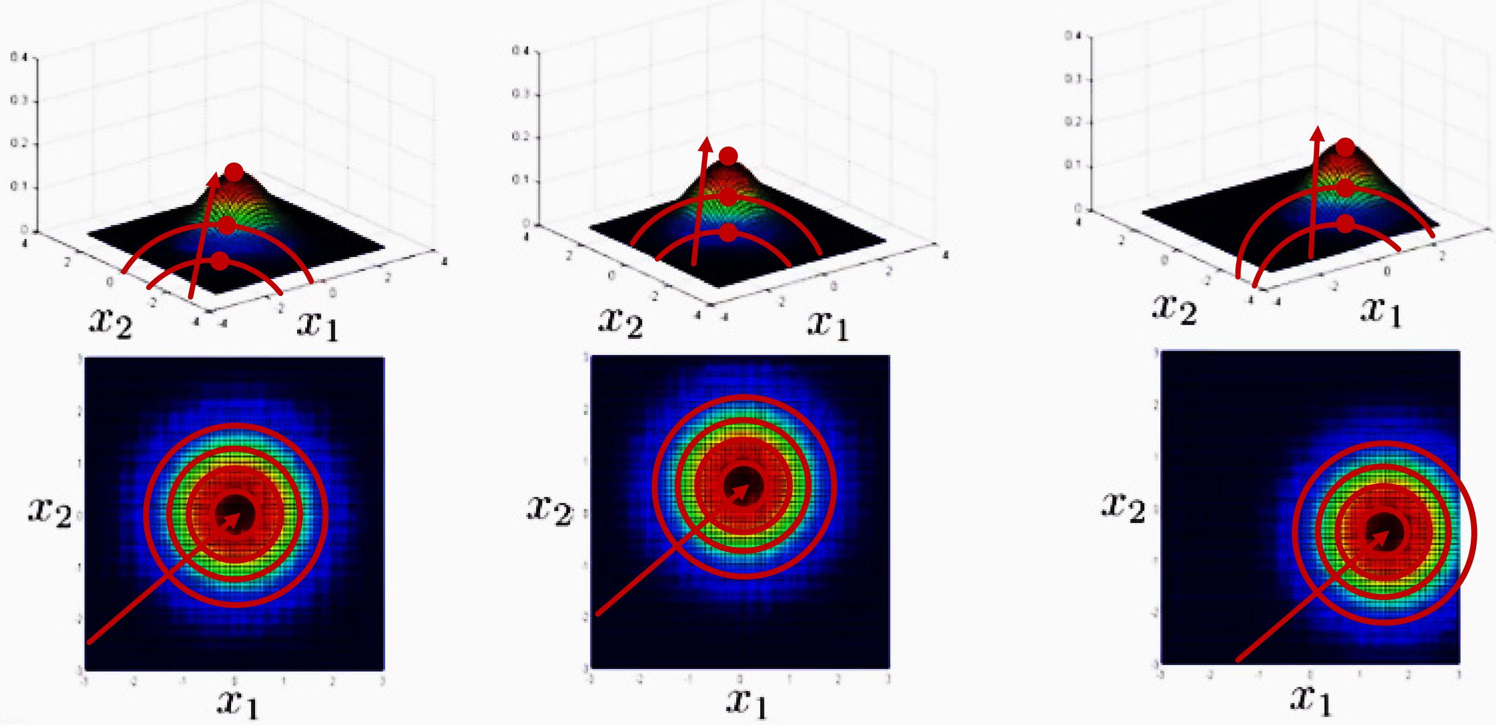
in contrast when sigma Σ is set to negative values:

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & \boxed{-0.5} \\ \boxed{-0.5} & 1 \end{bmatrix} \qquad \mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & \boxed{-0.8} \\ \boxed{-0.8} & 1 \end{bmatrix}$$

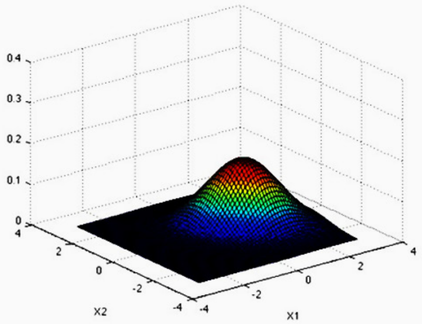


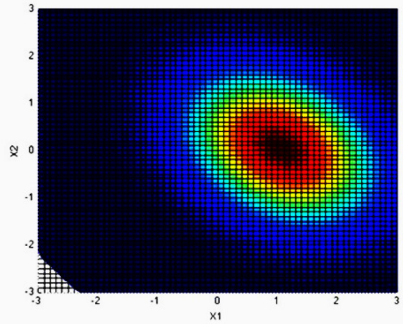
additionally, the μ parameter can also be varied:

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \mu = \begin{bmatrix} 0 \\ \boxed{0.5} \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \mu = \begin{bmatrix} \boxed{1.5} \\ \boxed{-0.5} \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



Consider the following multivariate Gaussian:





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

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

Correct Response

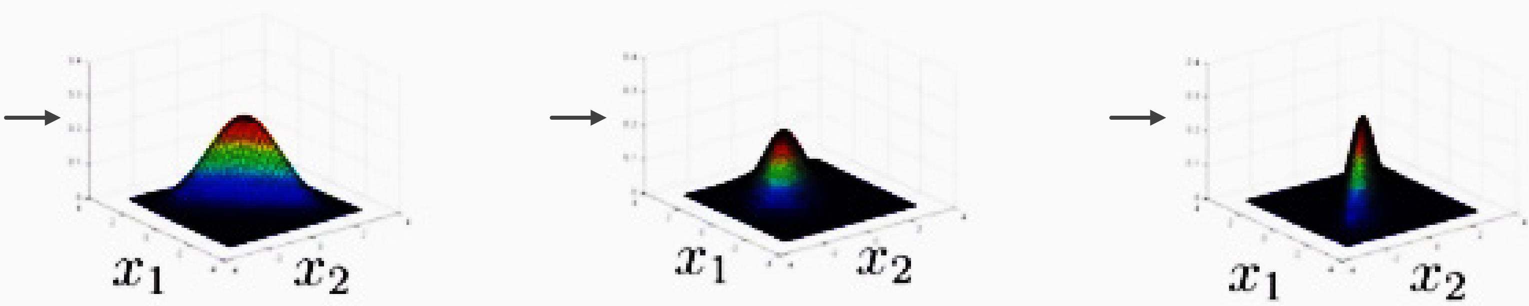
- $\mu = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \Sigma = \begin{bmatrix} 1 & -0.3 \\ -0.3 & 1 \end{bmatrix}$

anaomaly detection using the multivariate gaussian distribution

multivariate gaussian (normal) distribution

parameters μ, Σ $\quad\quad$ $\mu \in \mathbb{R}^2$ and $\sigma \in \mathbb{R}^2$

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



parameter fitting:

given training set $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\} \leftarrow x \in \mathbb{R}^n$

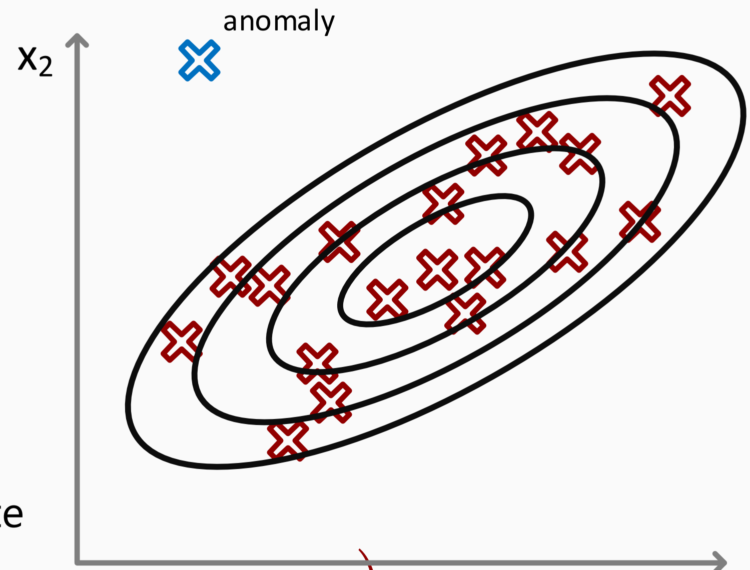
$$\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)} \quad \text{and} \quad \Sigma = \frac{1}{m}\sum_{i=1}^{m}\left(x^{(i)}-\mu\right)\left(x^{(i)}-\mu\right)^T$$

anamoly detection with multvariate gaussian

1) Fit the model p(x) by setting

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

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



2) given a new example x, compute

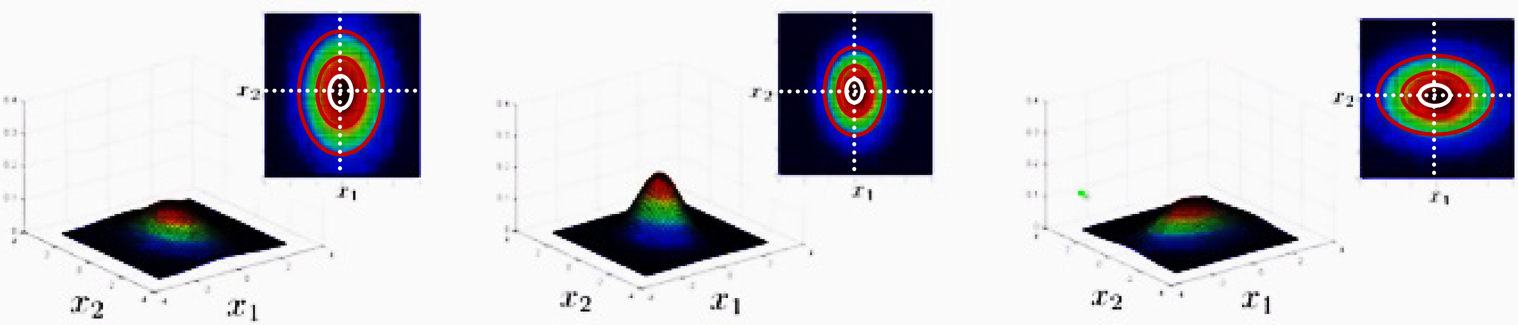
$$p(x;\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}\,|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x-\mu)^{T}\,\Sigma^{-1}(x-\mu)\right)$$

flag an anomaly if $p(x) < \varepsilon$

relationship to the original single variable model

original model: $p(x) = p(x_1;\mu_1,\sigma_1^2) \times p(x_2;\mu_2,\sigma_2^2) \times \ldots \times p(x_n;\mu_n,\sigma_n^2)$

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



the original model corresponds to multivariate gaussians, where the contours of the gaussian are always axis aligned.

$$\text{constraint: } \Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_1^2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_n^2 \end{bmatrix}$$

original model

$$p(x_1; \mu_1, \sigma_1^2) \times \ldots \times p(x_n; \mu_n, \sigma_n^2)$$

manually create features to capture anomalies where x_1, x_2 take unusual combinations of values

$$\frac{x_1}{x_2} = x_5 \text{ or } \frac{(x_4)^2}{x_3} = x_6 \text{ etc...}$$

computationally cheaper and scales better to larger n's

$$n = 10{,}000, n = 100{,}000$$

sufficient model even if m (training set size) is small

multivariate gaussian

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

automatically captures correlations between multiple features

$$\Sigma = \mathbb{R}^{n \times n} \qquad \Sigma^{-1}$$

computationally more expensive

$$\Sigma \sim \frac{n^n}{2}$$

must have $m > n$ or else Σ is noninvertible $\rightarrow m \gg n$

Consider applying anomaly detection using a training set $\{x^{(1)}, \ldots, x^{(m)}\}$ where $x^{(i)} \in \mathbb{R}^n$. Which of the following statements are true? Check all that apply.

- [x] The original model $p(x_1; \mu_1, \sigma_1^2) \times \cdots \times p(x_n; \mu_n, \sigma_n^2)$ corresponds to a multivariate Gaussian where the contours of $p(x; \mu, \Sigma)$ are axis-aligned.

Correct Response

- [ ] Using the multivariate Gaussian model is advantageous when m (the training set size) is very small (m < n).

Correct Response

- [x] The multivariate Gaussian model can automatically capture correlations between different features in x.

Correct Response

- [x] The original model can be more computationally efficient than the multivariate Gaussian model, and thus might scale better to very large values of n (number of features).

Correct Response