IID.

$(x_1, x_2 \dots x_n | \theta) = f_D(x_1|\theta) f_D(x_2|\theta) \cdots f(x_n|\theta)$

多元高斯密度函数：

均值0，方差1的一元正态分布。

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

均值u，方差δ^2时，标准化：$z = \frac{x-u}{\delta}$，标准化后方差为1。标准化的意义在于将数据点x到均值u的距离转化为数据点x到均值的距离等于多少个总体的标准差δ，消除了数据分布差异和量纲对概率计算的影响：

$$f(x) = \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{z^2}{2}} = \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{(x-u)^2}{2\delta^2}}$$

$u)^2$

高斯分布概率密度计算核心在于计算数据点到中心的距离，并且除以标准差将这个绝对距离转化为相对距离，然后通过距离平方的指数衰减计算概率密度。

先从各维度不相关的多元正态分布入手，数据点通过d维的列向量描述 $x = [x_1, x_2, \dots, x_d]^T$，各维度均值方差分别为$u_1, u_2 \dots, u_d$，$\delta_1, \delta_2 \dots, \delta_d$描述：

$$f(x) = \frac{1}{(\sqrt{2\pi})^d \delta_1 \delta_2 \cdots \delta_d} e^{-\frac{1}{2}\left[\left(\frac{x_1-u_1}{\delta_1}\right)^2 + \left(\frac{x_2-u_2}{\delta_2}\right)^2 + \cdots + \left(\frac{x_d-u_d}{\delta_d}\right)^2\right]}$$

前面多出的项是为了让概率之和为1，该方程也可写为：

$$f(x) = \frac{1}{(\sqrt{2\pi})\delta_1} e^{-\frac{1}{2}\left(\frac{x_1-u_1}{\delta_1}\right)^2} \frac{1}{(\sqrt{2\pi})\delta_2} e^{-\frac{1}{2}\left(\frac{x_2-u_2}{\delta_2}\right)^2} \cdots \frac{1}{(\sqrt{2\pi})\delta_d} e^{-\frac{1}{2}\left(\frac{x_d-u_d}{\delta_d}\right)^2}$$

各维度之间不相关的多元正态分布概率密度其实就是各个维度正态分布概率密度函数的乘积，因为各变量之间互不相关，联合概率密度等于各自概

率密度的乘积

$$d^2(x,u)=\left(\frac{x_1-u_1}{\delta_1}\right)^2+\left(\frac{x_2-u_2}{\delta_2}\right)^2+\cdots+\left(\frac{x_d-u_d}{\delta_d}\right)^2$$

$$=[x_1-u_1, x_2-u_2, \cdots x_d-u_d]\begin{bmatrix}\frac{1}{\delta_1^2} & 0 & \cdots & 0\\ 0 & \frac{1}{\delta_2^2} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \frac{1}{\delta_d^2}\end{bmatrix}\cdot\begin{bmatrix}x_1-u_1\\ x_2-u_2\\ \vdots\\ x_d-u_d\end{bmatrix}=(\bar{x}-\bar{u})^T\Sigma^{-1}(\bar{x}-\bar{u})$$

$$\therefore f(x)=\frac{1}{(\sqrt{2\pi})^d(\delta_1\delta_2\cdots\delta_d)}\exp\left(-\frac{1}{2}d(x,u)^2\right)$$

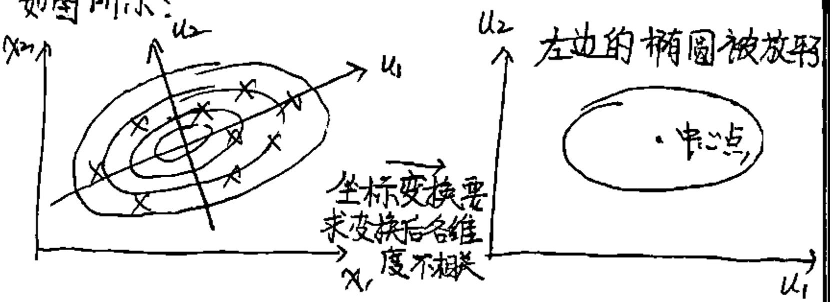
$$=\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}}e^{-\frac{1}{2}(\bar{x}-\bar{u})^T\Sigma^{-1}(\bar{x}-\bar{u})}$$

Σ 是协方差矩阵，里面第i行第j列元素表示第i个变量与第j个变量的协方差，由于假设各个维度不相关 $\therefore$ 协方差只有在对角线位置有值. $\delta_1\cdot\delta_2\cdots\delta_d=|\Sigma|^{\frac{1}{2}}$

$\boxed{\text{维度之间互相关的多元高斯分布}}$

化归思想. 找到倾斜的椭圆分布的长轴方向 u_1 和短轴方向 u_2. 计算数据点在这两个轴上的坐标(变换到这两个方向后，新维度之间显然不相关). 以长轴为x轴，短轴为y轴建立新的坐标系，经过变换后相当于将倾斜的椭圆放平，此时数据的各个维度之间不相关，就可以用前面各维度不相关的高斯分布解了。

如图所示：



如果变换的方向 u_1 和 u_2 用列向量 $\bar{u}_1=\begin{bmatrix}u_1^1\\u_1^2\end{bmatrix}$，$\bar{u}_2=\begin{bmatrix}u_2^1\\u_2^2\end{bmatrix}$
那么数据的投影长度可以用点积来计算为：

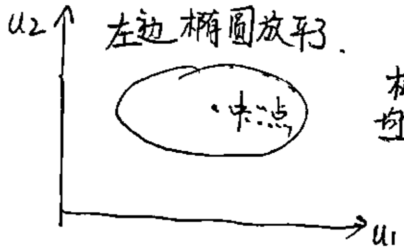
$\bar{u}_1^T\cdot\bar{X}=[u_1^1,u_1^2]\begin{bmatrix}x_1\\x_2\end{bmatrix}$ 投影长度代表了数据点在 u_1 方向上的坐标，这个过程可以用矩阵变化表示为：

$\bar{Y}=\begin{bmatrix}u_1^T\cdot\bar{X}\\u_2^T\cdot\bar{X}\end{bmatrix}=\begin{bmatrix}u_1^T\\u_2^T\end{bmatrix}\bar{X}=U^T\bar{X}$. 其中 $X=\begin{bmatrix}x_1\\x_2\end{bmatrix}$
代表源空间的数据坐标，$U=[\bar{u}_1,\bar{u}_2]$ 每一列代表一个投影方向。$\bar{u}_1,\bar{u}_2$ 代表变换方向的单位向量，$\bar{u}_1^T\cdot\bar{X}$ 代表将数据点 $\bar{X}$ 投影到 u_1 方向上的长度。
∵ $\bar{u}_1,\bar{u}_2$ 都是单位向量，且相互垂直，∴ U 是一个正交矩阵 $U^T=U^{-1}$
现在数据的各个维度已经去相关，即可用多元正态分布计算。首先需将数据标准化(消除量纲)

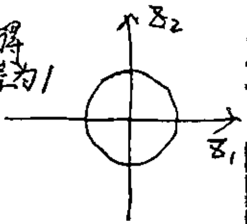
$$Z=\begin{bmatrix}\frac{Y_1-u_{Y_1}}{\delta_{Y_1}}\\\frac{Y_2-u_{Y_2}}{\delta_{Y_2}}\end{bmatrix}=\begin{bmatrix}\frac{1}{\delta_{Y_1}}, & 0\\0, & \frac{1}{\delta_{Y_2}}\end{bmatrix}\begin{bmatrix}Y_1-u_{Y_1}\\Y_2-u_{Y_2}\end{bmatrix}$$

$$=D(\bar{Y}-\bar{u}_Y)=D(U^T\bar{X}-U^T\bar{u}_X)$$

$$=DU^T(\bar{X}-\bar{u}_X)$$



标准化使得均值为0,方差为1



相对距离的平方,

$$d^2(z,u) = z_1^2 + z_2^2 = [z_1 \ \ z_2]\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \bar{z}^T \bar{z}$$

$$d^2(x,u) = \bar{z}^T\bar{z} = (DU^T(\bar{X}-\bar{u}_x))^T(DU^T(\bar{X}-\bar{u}_x))$$

$$= \underline{(\bar{X}-\bar{u}_x)^T U D^T D U^T (\bar{X}-\bar{u}_x)}$$

$$\Sigma_Y^{-1} = D^T D = \begin{bmatrix} \frac{1}{\delta_{Y_1}^2} & 0 \\ 0 & \frac{1}{\delta_{Y_2}^2} \end{bmatrix}$$

Σ_Y 是去相关后数据的协方差矩阵.因为是对角阵,它的逆等于对角元素取倒数

∵ 变换后数据各维度不相关,也就是变换后的协方差矩阵是对角阵,即:

$$\Sigma_Y = \begin{bmatrix} \delta_{Y_1}^2 & 0 \\ 0 & \delta_{Y_2}^2 \end{bmatrix}$$

$$\Sigma_Y = E[(\bar{Y}-\bar{u}_Y)(\bar{Y}-\bar{u}_Y)^T]$$

$$= E[U^T(\bar{X}-\bar{u}_x)(\bar{X}-\bar{u}_x)^T U]$$

$$= U^T E[(\bar{X}-\bar{u}_x)(\bar{X}-\bar{u}_x)^T] U$$

$$= U^T \Sigma_X U$$

即 $\Sigma_X = U\Sigma_Y U^T$, U是正交矩阵, $U^{-1} = U^T$

$$\therefore \ UD^TDU^T = U\Sigma_Y^{-1}U^T$$

And ∵ $(U\Sigma_Y^{-1}U^T)(U\Sigma_Y U^T) = I$

∴ $(U\Sigma_Y^{-1}U^T)\Sigma_X = I$

∴ $UD^TDU^T = \Sigma_X^{-1}$

∴ $d^2(X,u) = (\bar{X}-\bar{u}_X)\Sigma_X^{-1}(\bar{X}-\bar{u}_X)$

我们在计算过程中得到最终零均值，方差为1的Z，相当于对原坐标X做了一次变换；

$Z = DU^T(\bar{X}-\bar{u}_X)$

因此，概率密度函数在源空间做全空间积分的时候需要做换元变换，整体减小了$|DU^T|$，

∵ $UD^TDU^T=\Sigma_X^{-1}$，∴ $|DU^T| = \sqrt{|\Sigma_X^{-1}|} = |\Sigma_X|^{-\frac{1}{2}}$

∴ 为保证概率密度函数全空间积分为1，需要乘上$|\Sigma_X|^{-\frac{1}{2}}$，还需要除以$(\sqrt{2\pi})^d$，这一项是在计算$e^{x^2}$的积分时引入的，每个维度都会有，∴是d次方

∴ $$f(X) = \frac{1}{\sqrt{2\pi}}e^{-\frac{Z^2}{2}} = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_X|^{\frac{1}{2}}}e^{-\frac{(\bar{X}-\bar{u}_X)^T\Sigma_X^{-1}(\bar{X}-\bar{u}_X)}{2}}$$

主要思想，通过线性变换将数据的各维度去相关，再将去相关后的数据标准化，在推导概率分布的过程中，可以消去这个变换，只需求源空间的协方差矩阵就可以了。

假设平面上有一点A，这个点客观存在，一旦A指定，它的概率大小P就已经确定了，现我们添加一个坐标系，添加坐标系使得P(A)可以被量化 $P(A)=f(u_1,u_2)$ $P(A)=f(u_1,u_2)$ 使用其他坐标系量化 $P(A)=f(v_1,v_2)$ $P(A)=f(v_1,v_2)$，不管使用哪个

坐标系，A点，概率始终不变 $\therefore f(u_1, u_2) = f(v_1, v_2)$

Example

$$\Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

$$U = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \quad U^T = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

$$(\Sigma)_{new} = U^T \Sigma U = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix} \cdot \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

$$= \begin{bmatrix} \delta_1^2 & 0 \\ 0 & \delta_2^2 \end{bmatrix}$$

$$\Rightarrow \quad \theta = \frac{\pi}{4}$$

$$\therefore (\Sigma)_{new} = \begin{bmatrix} 1.8 & 0 \\ 0 & 0.2 \end{bmatrix}$$

新的坐标是原坐标系经过 $\theta=\frac{\pi}{4}$ 旋转而来，在新坐标系下，输入元素变得不相关，x_1方向方差为1.8，分布比较宽，x_2方向的方差为0.2，分布比较窄。