ColumbiaX: Machine Learning: Week1: multivariate gaussian distribution

probabilistic model vs non-probabilistic model

Difference is that the presence of probability distribution with this model. here is an example of probabilistic model focused on how the data of dimension d is generated.

Multivariate Gaussian distribution

The density function is :

$$p(x|\mu,\sigma) = \frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{det(\sigma)}} \exp^{-\frac{1}{2}(x-\mu)^T\sigma^{-1}(x-\mu)}$$

where μ is mean vector across dimension 2 of m x d dimensional dataset matrix, its dimension is d too. and σ is a covariance matrix of d x d dimension which is positive definite and symmetric. Let’s go deeper into the formula above in detail.

1 what is a covariance matrix ?

for a dataset with m x d dimension where m is the amount of observations and d is the length of the observation. Intuitively we have d variables (regard features as variables) . **covariance** is defined as follows:

$$Cov(\mathbf{X},\mathbf{Y}) = \frac{\sum_{i=1}^{m}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{y}_i-\bar{\mathbf{y}})}{n-1} = \mathbf{E}[(\mathbf{X}-\bar{\mathbf{x}})(\mathbf{Y}-\bar{\mathbf{y}})]$$

$\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ is the mean value of each feature dimension, here we only consider case d = 2. What does the covariance mean? It depicts the correlation between 2 features. if 2 features are positively correlated, the conv(x,y) has a positive sign, ,if negatively correlated, then a negative sign of conv(x,y), if conv(x,y) == 0 , it means the 2 features have no correlation. the higher the absolute value , the higher the correlation between two features. To eliminate the effect of dimension(量纲), correlation coefficient is induced.

$$\mathbf{p} = \frac{Cov(\mathbf{X},\mathbf{Y})}{dev(X)dev(Y)}$$

where $dev(\mathbf{X}) = \sqrt{\mathbf{E}((\mathbf{X} - \bar{\mathbf{x}})^2)}$ represents the standard deviation of X. correlation coefficient can be treated as normalized covariance. Intuitive and easy-to-understand explanation of covariance and correlation coefficient can be found here [Intuitively understanding covariance and correlation coefficient](). In short, covariance gives a measure of how much two variables are linearly related to each other. Correlation normalize the contribution of each variable in order to measure only how much the variables are related, rather than also being affected by the scale of the separate variables. Independence is a distinct property from covariance. Independence requires both linear and non-linear independence. But zero covariance only represents linear independence.

But covariance can only represent up to 2 variables(features), to characterize multi-variate covariance, covariance matrix is induced.

2 What is linear independence and non-independence?

3 The defination of positive definite matrix

要判断一个矩阵是否为正定、半正定、负定矩阵，首先要了解二次型的概念。
二次，即代表在函数中，变量的阶数都是二阶齐次的。

> $$\mathbf{f}(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n) = \mathbf{a}_{11}\mathbf{x}_1^2 + \mathbf{a}_{22}\mathbf{x}_2^2 + ... + \mathbf{a}_{nn}\mathbf{x}]_\mathbf{n}^\mathbf{2} + \sum_{\mathbf{i,j=1}}^{\mathbf{n}} \mathbf{a_{ij}}\mathbf{x_i}\mathbf{x_j}(\mathbf{i}\neq\mathbf{j})$$

该函数可以进一步用矩阵来表示

> $$\mathbf{f} = \mathbf{X}^T\mathbf{A}\mathbf{X}$$
>
> $$X = (\mathbf{x_1}, \mathbf{x_2}, \mathbf{...}, \mathbf{x_n})^T, \mathbf{A}_{ij} = \mathbf{a}_{ij}$$

若$\forall_{i,j}\mathbf{a}_{ij}$为实数，则$\mathbf{f}$称为实二次型。
研究实二次型的目的在于研究一个二次方程所表示的曲线的几何形状，首先就要将方程化为标准方程，然后才可讨论其几何形状。 标准型定义为只含有平方项的二次型。

如果对任何一组不全为0元素的$\mathbf{X}$，都有$f > 0$，则称f为正定二次型，其对应矩阵称为正定矩阵。

判定条件：

> 1) n元实二次型$\mathbf{f}$为正定的充分必要条件是：标准型的n个系数全为正。
> 2) 实对称矩阵$\mathbf{A}$为正定矩阵的充分必要条件是$\mathbf{A}$的特征值全为正。
> 3) n阶实对称矩阵$\mathbf{A}$为正定矩阵的充分必要条件是$\mathbf{A}$的各阶顺序主子式都为正值。

半正定矩阵则弱化为对于任何一组不全为0的$\mathbf{X}$，都有$f \geq 0$.