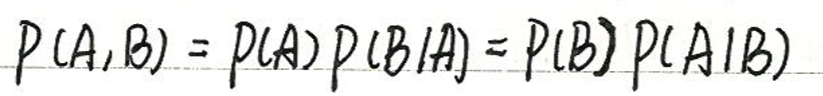
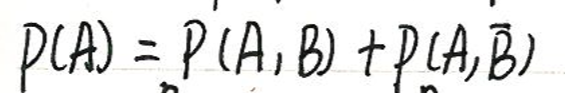
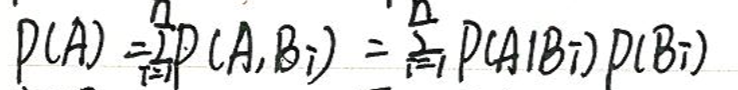
联合概率：A，B同时发生的概率为A的概率乘以A发生条件下B的概率



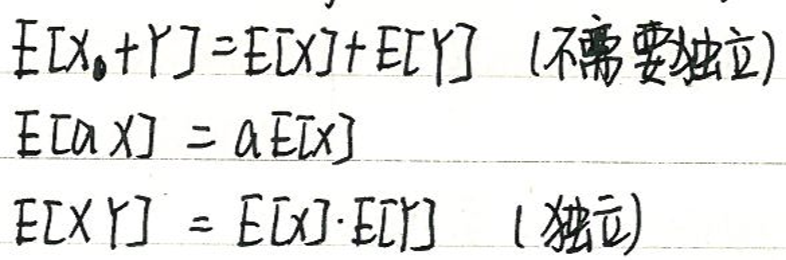
A的概率为AB同时发生的概率加A发生B不发生的概率



A发生的概率为A和其他事件同时发生的概率之和

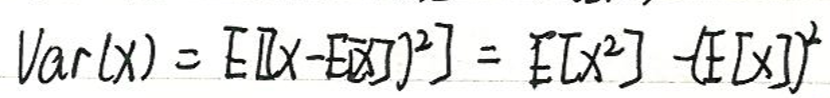


期望：x的期望为x(i)\*Px(i)之和（连续型则求积分）

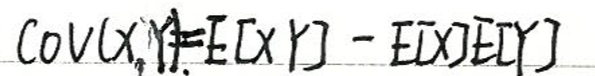


1. x和y的期望等于x的期望+y的期望
2. a\*x的期望等于a\*（x的期望）
3. x和y同时出现的期望（需要x、y互相独立）等于x的期望乘以y的期望

方差：（x-x的期望）的平方的期望=x的平方的期望-x的期望的平方



协方差：xy同时发生的期望减x的期望乘y的期望



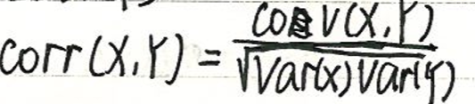
协方差：1/n\*[∑(xi-Ex)(yi-Ey)]

协方差矩阵

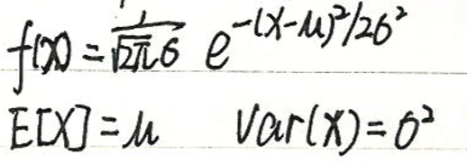
[ cov(x,x) cov(x,y)

Cov(y,x)cov(y,y)]

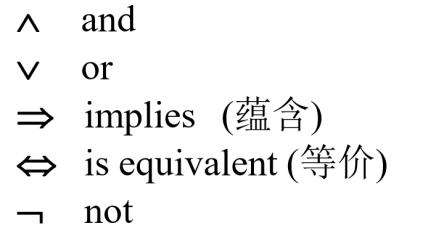
相关系数：X,y的协方差除以x的方差乘y的方差再开方



正态分布：



Propositional logic命题逻辑



真值表：

如果P为false，则P=>任意Q都正确

只有P和Q真值相等才能互推正确

Lec 9

神经网络

Net input Ij = x1\*w1j+x2\*w2j+x3\*w3j+……+bias（j）

Output Oj = 1/（1+e^Ij）

Lec 10

支持度：同时包含前项与后项的概率

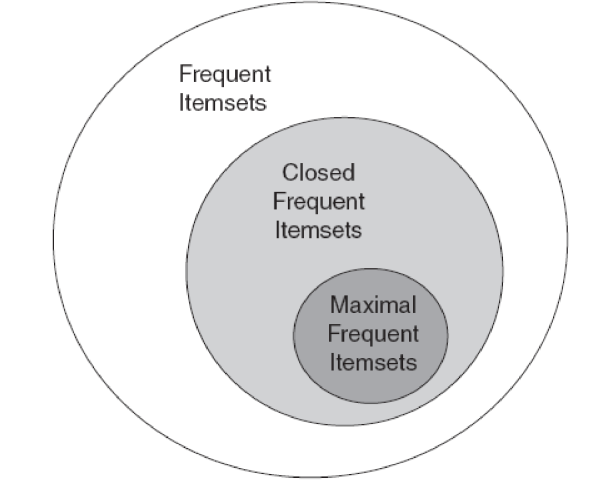
置信度：包含前项的同时包含后项的概率

频繁项集：支持度大于最小支持度的项集

最大频繁项集：直接超集不频繁

闭项集：直接超集支持度与该项集不同

Apriori：由于k项集的子集不频繁时该项集也不频繁，所以可以直接从k-1频繁项集中自连接得到k候选项集，再修剪去不频繁的项集得到k频繁项集



Lec 11

k-means：不断求中心点直到收敛，分为k个聚类

DBSCAN计数包括点本身、包括圆上的点（任何点的密度都取决于半径）

核心点：在半径Eps内周围点数在Minpts或以上的点

边界点：在核心点Eps距离内的点

噪声点：既不是核心也不是边界

Lec 12

Hierarchical Clustering层次化聚类

Agglomerative凝聚式

Divisive分裂式

计算距离矩阵

Loop

合并最近两个聚类

更新距离矩阵以反映新聚类与原始聚类间距离

Loop

只剩下一个聚类

Single link 所有都取最小值

Complete link：两个簇任意两点距离最大值的最小的两个簇合并