- 数学基础
  - 线性代数：
    - 谱定理（物理意义）：SVD实际就是将一个矩阵分解为两个旋转+一个缩放，对于对称矩阵，两个旋转矩阵是一样的，进一步拆开，就是特征值与特征向量（正交）的乘积；

$$A = U\Lambda U^T = \sum_{i=1}^{n} \lambda_i u_i u_i^T, \Lambda = \text{diag}(\lambda_1, \cdots, \lambda_n)$$

    - 瑞利商：利用SVD物理意义理解——A分解后，旋转矩阵不会改变长度，只有缩放矩阵（这里是个对角矩阵，特征值），所以根据特征值大小，界定范围

Given a symmertric matrix $A \in S^n$,

Physical meaning of SVD!

$$\boxed{\lambda_{\min}(A) \le \frac{x^T A x}{x^T x} \le \lambda_{\max}(A), \forall x \ne 0}$$

$$\lambda_{max}(A) = \max_{x:\|x\|_2=1} x^T A x$$

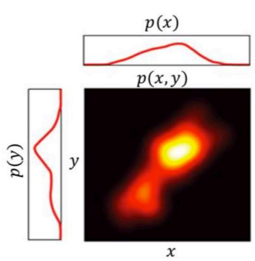
$$\lambda_{min}(A) = \min_{x:\|x\|_2=1} x^T A x$$

  - 概率论：
    - 边缘概率（可以理解为全概率公式）：对联合概论的一个变量进行积分，即不考虑一个变量的影响；

Continuous:

$$p(x) = \int p(x,y)\,dy$$

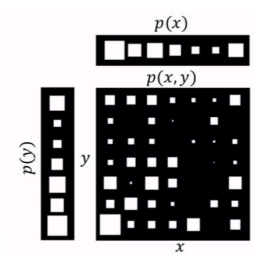
$$p(y) = \int p(x,y)\,dx$$



Discrete:

$$p(x) = \sum_y p(x,y)$$

$$p(y) = \sum_x p(x,y)$$



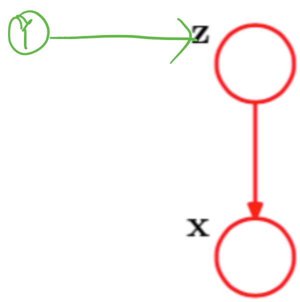
    - 条件概率：还是关心某个变量，例如y的，只是这时y值是确定的，即只考虑这个y = y1，不考虑其他y值，来看此时的联合概率分布；

$$P(x|Y=y^*) = \frac{p(x, Y=y^*)}{\int p(x, Y=y^*)dx} = \frac{p(x, Y=y^*)}{p(Y=y^*)}$$

      - 变形后，得到贝叶斯公式！——p(x, y) = p(xly)p(y) = p(ylx)p(x)

○ 图论：

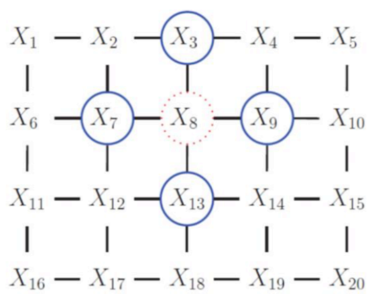
- 有向图：节点代表变量，边代表联系；如果没有联系，p(x,y) = p(x)p(y)，相互独立；
  - 马尔可夫假设：一个变量x，在给定父节点情况下，独立于它的非子节点！因为前驱节点的信息已经浓缩到父节点了，不需要再去参考前驱节点；



若z已知，则 x 独立于 Y

the joint distribution is $p(x,z) = p(z)p(x|z)$

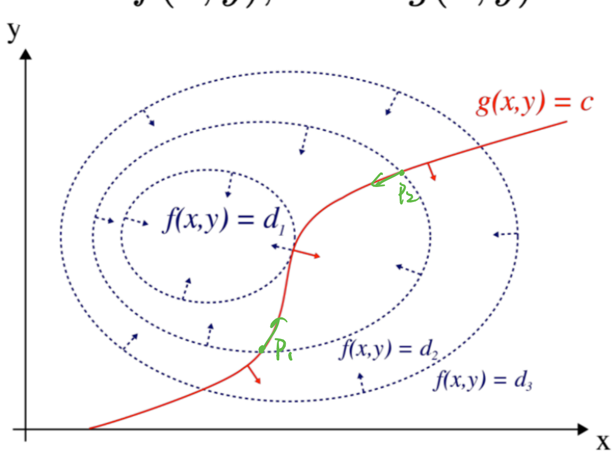
- 无向图：MRF马尔可夫随机场/MN马尔可夫网络；同样的，给定所有父节点，则该变量与其他节点独立；



○ 优化方法： 拉格朗日乘数法（多元微积分）：等高线的梯度于切线垂直啦！

- https://blog.csdn.net/Queen0911/article/details/100611797

$$\max \ f(x,y), \ s.t.: \ g(x,y) = 0$$



- 可以看到，最值点处，等高线于条件约束相切！
- 在点p1、p2处，虽然满足约束，但是仍然可以向梯度，沿着红线约束方向前进，直到到达最值点，在红线上如何移动，都不是朝着梯度方向，不能增大了！
- 所以最值点一定满足：等高线梯度方向 于 条件约束梯度方向 在同一条线上（方向可不同），即：
- 引入拉格朗日乘数 $$\nabla_{x,y} f = \lambda \nabla_{x,y} g$$ 法：

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda g(x, y)$$

  - 整理求解得 $$\nabla_{x,y,\lambda} \mathcal{L}(x, y, \lambda) = 0.$$ 到：可以看到即满足约束条件，又满足最值点的条件！

$$\nabla_{x,y,\lambda} \mathcal{L}(x, y, \lambda) = 0 \iff \begin{cases} \nabla_{x,y} f(x, y) = \lambda \nabla_{x,y} g(x, y) \\ g(x, y) = 0 \end{cases}$$

  - 拉格朗日乘数法——专门用于求解带等式约束条件的优化问题！