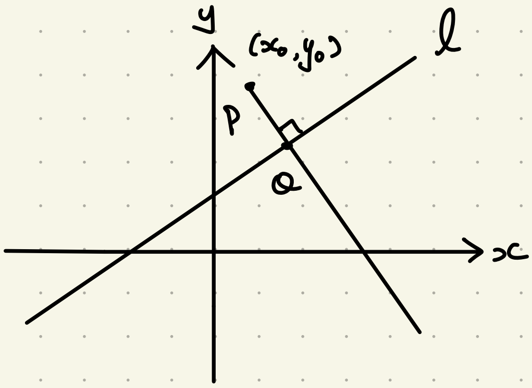
1. 证明：点 (u,v) 到一条线 (a,b,c) 的距离为：$|au+bv+c|$，这里 $a^2+b^2=1$ （15分）



设点 $P(x_0,y_0)$ 和直线 l (a,b,c) 的距离为点P到直线 l 的垂线段的长。

设点P到直线 l 的垂线为 l'，垂足为Q。

$$Ax+By+C=0$$
$$Ax+C=-By$$
$$y=-\frac{A}{B}x-\frac{C}{B} \quad -(1)$$

$\therefore$ l 的斜率为 $-\frac{A}{B}$.

$\because l\perp l'$ $\therefore$ l' 斜率为 $\frac{B}{A}$. $\therefore$ l' 方程为 $y-y_0=\frac{B}{A}(x-x_0) \quad -(2)$

将(1)代入(2)

$$-\frac{A}{B}x-\frac{C}{B}-y_0=\frac{B}{A}(x-x_0)$$
$$-\frac{A}{B}x-\frac{C}{B}-y_0=\frac{B}{A}x-\frac{B}{A}x_0$$
$$-\frac{C}{B}-y_0+\frac{B}{A}x_0=\left(\frac{B}{A}+\frac{A}{B}\right)x$$
$$x=\frac{-\frac{C}{B}-y_0+\frac{B}{A}x_0}{\frac{B^2+A^2}{AB}}$$
$$x=\frac{-ABy_0+B^2x_0-AC}{A^2+B^2}$$

$$(1)\quad \frac{A}{B}x=-\frac{C}{B}-y$$
$$x=-\frac{C}{A}-\frac{B}{A}y \quad -(3)$$

将(3)代入(2)

$$y-y_0=\frac{B}{A}\left(-\frac{C}{A}-\frac{B}{A}y-x_0\right)$$
$$y-y_0=-\frac{BC}{A^2}-\frac{B^2}{A^2}y-\frac{B}{A}x_0$$
$$y+\frac{B^2}{A^2}y=-\frac{BC}{A^2}-\frac{B}{A}x_0+y_0$$
$$\left(1+\frac{B^2}{A^2}\right)y=-\frac{BC}{A^2}-\frac{AB}{A^2}x_0+\frac{A^2y_0}{A^2}$$
$$\left(\frac{A^2+B^2}{A^2}\right)y=-\frac{BC}{A^2}-\frac{AB}{A^2}x_0+\frac{A^2y_0}{A^2}$$
$$y=\frac{A^2y_0-ABx_0-BC}{A^2+B^2}$$

$\therefore$ 点Q坐标 $\left(\frac{B^2x_0-ABy_0-AC}{A^2+B^2},\frac{A^2y_0-ABx_0-BC}{A^2+B^2}\right)$

$$|PQ|^2=\left(\frac{B^2x_0-ABy_0-AC}{A^2+B^2}-x_0\right)^2+\left(\frac{A^2y_0-ABx_0-BC}{A^2+B^2}-y_0\right)^2$$
$$=\left(\frac{-A^2x_0-ABy_0-AC}{A^2+B^2}\right)^2+\left(\frac{-B^2y_0-ABx_0-BC}{A^2+B^2}\right)^2$$
$$=\frac{A^2(Ax_0+By_0+C)^2+B^2(By_0+Ax_0+C)^2}{(A^2+B^2)^2}$$
$$=\frac{(Ax_0+By_0+C)^2}{A^2+B^2}$$

$$\therefore |PQ|=\frac{|Ax_0+By_0+C|}{\sqrt{A^2+B^2}}$$

$\because A^2+B^2=1$, $P(u,v)$,

$\therefore d=|PQ|=|Au+Bv+C|$ #

2. 简述EM算法的基本原理和流程（以高斯混合模型求解为例）（15分）

EM算法的基本原理：在概率模型中寻找最大似然估计或最大后验估计的算法.
其中，此概率模型依赖于无法观测的隐性变量

以高斯混合模型求解为例，现已知 $\begin{cases}\text{各个类别的分布模型}\\ \text{采样数据}\end{cases}$

，通过EM算法求未知 $\begin{cases}\text{每个采样数据所属类别}\\ \text{各个模型参数}(\mu,\Sigma,\pi)\end{cases}$ μ为均值，Σ为方差，π为高斯分布的权重（选择模型的先验概率），

而每个采样数据的所属类别为隐变量，EM算法可用于解数据缺失的参数估计问题.

E-step：计算期望E，利用对隐变量的现有估计值，计算其最大似然估计值

M-step：最大化M，通过最大化在E-step中的最大似然估计值计算各参数值.
此参数值会用于下一个E-step.

E-step

已知样本集 $Y=(y_1, y_2, \cdots, y_T)$，引入隐变量 $\gamma_{t,k}$（表示样本y_t属于第k个模型），
设总计有M个模型（M个类），T个样本

此时完整数据为 $(y_t, \gamma_{t,1}, \gamma_{t,2}, \cdots, \gamma_{t,M})$，$t=1,2,T$

若 y_t 属于第1个模型，则 $\gamma_{t,1}=1$，$\gamma_{t,2}=\gamma_{t,3}=\cdots=\gamma_{t,M}=0$，$(y_t, 1, 0, \ldots, 0)$

完全数据的似然函数 $p(y,\gamma\mid\mu,\Sigma,\pi)=\prod_{t=1}^{T}p(y_t,\gamma_{t,1},\cdots\gamma_{t,M}\mid\mu,\Sigma,\pi)$

$$=\prod_{t=1}^{T}\prod_{k=1}^{M}(\pi_k N(y_t;\mu_k,\Sigma_k))^{\gamma_{t,k}}$$

$N(y_t\mid\mu_k,\Sigma_k)$为第k个高斯模型的概率密度函数.
$\Rightarrow$ 选定第k个模型后，模型产生y_t的概率.

$$=\prod_{k=1}^{M}\pi_k^{\gamma_{t,k}}\prod_{t=1}^{T}(N(y_t;\mu_k,\Sigma_k))^{\gamma_{t,k}}.$$

完全数据的对数似然函数

目标：寻找 (μ_*, Σ_*, π_*)，
使 $\ln p(y,\gamma\mid\mu,\Sigma,\pi)$最大 ★.

$$\ln p(y,\gamma\mid\mu,\Sigma,\pi)$$

$$=\sum_{k=1}^{M}\left(\sum_{t=1}^{T}\gamma_{t,k}\right)\ln\pi_k+\sum_{t=1}^{T}\gamma_{t,k}\left(-\ln(2\pi)-\frac{1}{2}\ln|\Sigma_k|-\frac{1}{2}(y_t-\mu_t)^T(\Sigma_k)^{-1}(y_t-\mu_t)\right)$$

$\because$ 隐变量 $\gamma_{t,k}$ 未知

$\therefore$ 先给定起始参数 μ_0, Σ_0, π_0，或第i次迭代参数 μ_i, Σ_i, π_i，对γ进行估计

最大化Q函数：

前提.

$$Q(\mu,\Sigma,\pi,\mu_i,\Sigma_i,\pi_i) = E_\gamma[\ln p(y,\gamma|\mu,\Sigma,\pi) \,|\, Y,\underline{\mu_i,\Sigma_i,\pi_i}]$$

$$= \sum_{k=1}^{M}\left(\sum_{t=1}^{T} E(\gamma_{t,k}|y_t,\mu_i,\Sigma_i,\pi_i)\right)\ln\pi_k + \sum_{t=1}^{T} E(\gamma_{t,k}|y_t,\mu_i,\Sigma_i,\pi_i)\left(-\ln(2\pi)-\frac{1}{2}\ln|\pi_k| - \frac{1}{2}(y_t-\mu_t)^T(\Sigma_k)^{-1}(y_t-\mu_t)\right)$$

隐变量的期望值估计 前提

$$E(\gamma_{t,k}|\underline{y_t,\mu_i,\Sigma_i,\pi_i}) = p(\gamma_{t,k}=1|y_t,\mu_i,\Sigma_i,\pi_i)$$

$$= \frac{\pi_{i,k}N(y_t;\mu_{i,k},\Sigma_{i,k})}{\sum_{k=1}^{M}\pi_{i,k}N(y_t;\mu_{j,k},\Sigma_{i,k})}$$

将 $E(\gamma_{t,k}|y_t,\mu_i,\Sigma_i,\pi_i)$ 的值代入Q函数，此时Q函数无隐变量，可最大化Q函数.

M-step

$$\mu_{i+1},\Sigma_{i+1},\pi_{i+1} = \arg\max Q(\mu,\Sigma,\pi,\mu_i,\Sigma_i,\pi_i)$$

对Q函数求导，其导数为0，可得：

$$\mu_{i+1,k} = \frac{\sum_{t=1}^{T} E(\gamma_{t,k}|y_t,\mu_i,\Sigma_i,\pi_i)y_t}{E(\gamma_{t,k}|y_t,\mu_i,\Sigma_i,\pi_i)},\quad k=1,2,\cdots,M$$

$$\Sigma_{i+1,k} = \frac{\sum_{t=1}^{T} E(\gamma_{t,k}|y_t,\mu_i,\Sigma_i,\pi_i)(y_t-\mu_{i,k})^2}{E(\gamma_{t,k}|y_t,\mu_i,\Sigma_i,\pi_i)},\quad k=1,2,\cdots,M$$

$$\pi_{i+1,k} = \frac{E(\gamma_{t,k}|y_t,\mu_i,\Sigma_i,\pi_i)}{M},\quad k=1,2,\cdots,M$$

参考资料：https://blog.csdn.net/lin_limin/article/details/81048411

3. 用伪代码写出Mean-shift的算法流程（以图像分割为例），并分析影响算法性能的主要因素。（15分）

图像分割 ⇒ 对图像中每个像素点给定一个类别的标签

图像特征点可提取至少5维特征，即 (x, y, R, G, B)。RGB为非均匀颜色空间，因此我们可以选择转化为 (x, y, L, U, V) 来保证颜色的均匀性。

Meanshift 算法会寻找概率密度最大的点。由于不同的点最终会收敛于不同的峰值，因此这些点会组成一类，图像分割的任务完成。

执行过程：

① 迭代空间构建，用 (x, y, R, G, B) 和不同的窗口半径构成5维空间球体，此球体以点P为圆心 $(x_p, y_p, R_p, G_p, B_p)$。

② 求所有在这个球体空间内的点相对中心点P的色彩向量之和 (W)，并将点P移至向量W的终点，此点作为新中心点1。重复步骤1直到中心点 P_n 不再移动。

③ 让输出图像对应的初始点P的色彩值更新为本轮迭代终点 P_n 的色彩值。

④ 对输入图像的每个点都依次执行步骤1,2,3，完成均值漂移滤波。

Mean Shift 聚类流程

① 在未标记的数据点中随机选择一个点作为中心点。

② 找出所有离中心点距离窗口半径之内的所有点，记为集合M，这些点为簇c。球内点属于这个类的概率加1。

③ 计算从中心点开始到集合M所有点的向量，并相加，得到向量shift.

④ 中心点沿着shift方向移动.

⑤ 重复步骤②,③,④，直到shift的大小非常小，得新的中心点.

⑥ 若簇c的中心点与其他簇的中心点距离小于阈值，则合并为1类，否则c为新类。

影响算法性能的主要因素：

① 特征子的选取 ② 窗口半径的大小

伪代码

Input: I, Picture features (x, y, R, G, B)
Φ, Kernel
B, Bandwidth (窗口半径)
t_1, threshold.

Output: Λ, Picture segmentation (x, y, k), $k \in \{N_1, N_2, \cdots, N_k\}$

$i = 0$
do
  $C_i = (I / \Lambda)$ (random) $\Rightarrow$ 随机采样部分样本来降低复杂度.
    do
      for $\{P_i\}$ in sphere of C with bandwidth B do
        $\vec{S}: \sum f(P_i, C, \Phi)$ $\Rightarrow$ $\vec{S}$: 漂移向量.
        $c(P_i, i) = c(P_i, i) + 1$
      end for
      $P = P + S$ $\Rightarrow$ S: 向量
    while $\|S\| > t_1$ $\Rightarrow$ 移动距离 > 阈值.
    for $k \in \{1, 2, \cdots, j-1\}$ do
      if $\|C_k - C_i\| < B$ then
        merge (C_k, C_i)
      end if.
    end for
while not $I / \Lambda = \Phi$
return Λ