语音识别Homework 2

王蔚达 2151300

Problem 1: Teacher-mood-model

Your school teacher gave three different types of daily homework assignments:

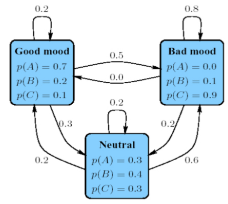
- A: took about 5 minutes to complete
- B: took about 1 hour to complete
- C: took about 3 hours to complete

Your teacher did not reveal openly his mood to you daily, but you kne that your teacher was either in a bad, neutral, or a good mood for a whole day.

Mood changes occurred only overnight.

Model parameters:

- Observation $\Sigma = \{A, B, C\}$
- Set of states $S = \{good, neutral, bad\}$
- Transition probabilities between any two states a_{ij}
- Emission probabilities within each state $b_i(x)$



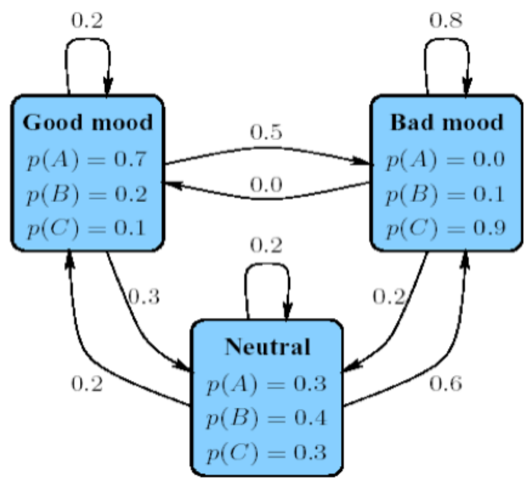
One week, your teacher gave the following homework assignments:

Monday	Tuesday	Wednesday	Thursday	Friday
A	C	B	A	C

QUESTIONS

What did his mood curve look like most likely that week?

- Searching for the most probable path – Viterbi algorithm



① 初始化

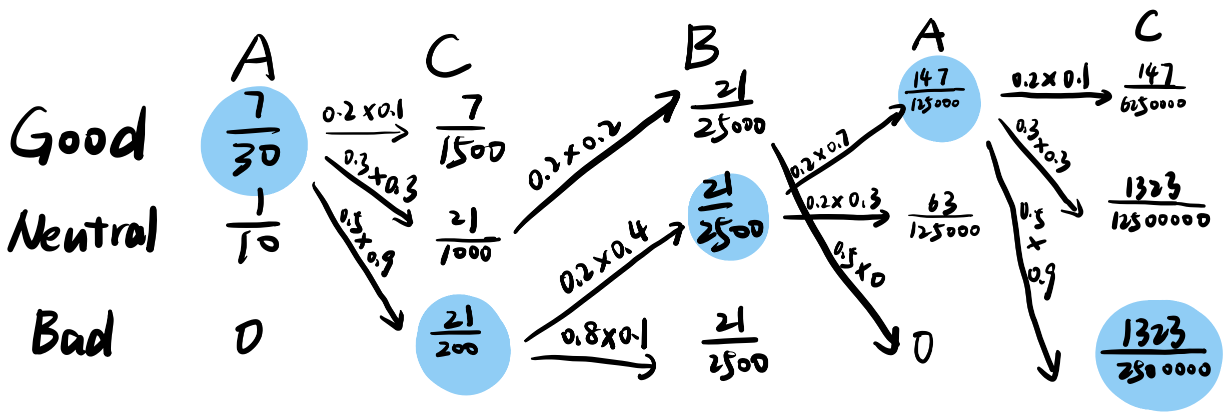
Good 且为A $= \frac{1}{3} \times 0.7 = \frac{7}{30}$

Neutral 且为A $= \frac{1}{3} \times 0.3 = \frac{1}{10}$

Bad 且为A $= \frac{1}{3} \times 0 = 0$

② 概率转移矩阵(利用Viterbi算法)

根据 $V_t^j = b_j(C)\max_i(V_{t-1}^i a_{ij})$ 计算得下表



③ 心情曲线

由回推,其心情曲线如下:

周一	周二	周三	周四	周五
A	C	B	A	C
good	bad	neutral	good	bad

Problem 2 (EM for a 1D Laplacian Mixture Model):

In this problem you will derive the EM algorithm for a *one-dimensional* Laplacian mixture model. You are given n observations $x_1, \dots, x_n \in \mathbb{R}$ and we want to fit a mixture of m Laplacians which has the following density

$$f(x) = \sum_{j=1}^{m} \pi_j f_L(x; \mu_j, \beta_j),$$

where $f_L(x; \mu_j, \beta_j) = \frac{1}{2\beta_j} e^{-\frac{1}{\beta_j}|x-\mu_j|}$, and the mixture weights π_j are a convex combination, i.e. $\pi_j \geq 0$ and $\sum_{j=1}^{m} \pi_j = 1$. For simplicity, assume that the scale parameters $\beta_j > 0$ are known beforehand and thus *fixed*.

- Introduce latent variables so that we can apply the EM procedure.
- Analogously to the previous question, write down the steps of the EM procedure for this model. If some updates cannot be written analytically, give an approach on how to compute them.

  (*Hint: Recall a property of functions that makes them easy to optimize.*)

EM算法由E步与M步构成

① E步(期望步骤)

假设隐藏变量为 $y_i \in \{1,2,3,\dots,m\}$，观测值为 x_i

即 y_i 是状态(state)，x_i 是观测值 (observation).

依题意知 $P(x_i|y_i=j) = f_L(x_i|\mu_j,\beta_j)$，$P(y_i=j) = \pi_j$

$$P(x_i) = \sum_{s=1}^{m} P(y_i=s) P(x_i|y_i=s) = \sum_{s=1}^{m} \pi_s f_L(x_i|\mu_s,\beta_s)$$

(全概率公式)

由此可以得到后验概率 $q_j(x_i)$ (给定 x_i 属于 j 状态

$$q_j(x_i) = P(y_i=j|x_i) = \frac{P(x_i|y_i=j)P(y_i=j)}{\sum_{s=1}^{m} P(x_i|y_i=s)P(y_i=s)} = \frac{\pi_j f_L(x_i|\mu_j,\beta_j)}{\sum_{s=1}^{m} \pi_s f_L(x_i|\mu_s,\beta_s)}$$

② M步 (最大化步骤)

该问题的对数极大似然估计为

$$\max_\theta \log \prod_{i=1}^{n} \left(\sum_{j=1}^{m} P(x_i, y_i=j|\theta) \right)$$

$$= \max_\theta \sum_{i=1}^{n} \log \sum_{j=1}^{m} P(x_i, y_i=j|\theta)$$

$$= \max_\theta \sum_{i=1}^{n} \log \sum_{j=1}^{m} q_j(x_i) \frac{P(x_i, y_i=j|\theta)}{q_j(x_i)}$$

$$\geq \max_\theta \sum_{i=1}^{n} \sum_{j=1}^{m} q_j(x_i) \log \frac{P(x_i, y_i=j|\theta)}{q_j(x_i)}$$

由 Jensen不等式知，当且仅当 $q_j(x_i) \propto P(x_i, y_i=j|\theta)$ 取等，即①中结果

在M step中可理解为已知x_i去优化θ（其中β固定，故优化μ）

故优化函数为

$$\theta^* = \arg\max g(\theta),\ g(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{m} q_j(x_i)\log P(x_i, y_i = j \mid \theta)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} q_j(x_i)\log\left(\pi_j \frac{1}{2\beta_j} e^{-\frac{1}{\beta_j}|x_i-\mu_j|}\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} q_j(x_i)\left(\log\pi_j - \frac{1}{\beta_j}|x_i-\mu_j| - \log 2\beta_j\right) \quad (1)$$

又有约束$\sum_{j=1}^{m}\pi_j = 1$（$q_j(x_i)$可视为由①得的常量）

故有 $\mathcal{L}(\pi,\mu,\lambda) = \sum_{i=1}^{n}\sum_{j=1}^{m} q_j(x_i)\left(\log\pi_j - \frac{1}{\beta_j}|x_i-\mu_j| - \log 2\beta_j\right) + \lambda\left(\sum_{j=1}^{m}\pi_j - 1\right)$

$$\frac{\partial\mathcal{L}}{\partial\pi_j} = \frac{\sum_{i=1}^{n} q_j(x_i)}{\pi_j} + \lambda = 0 \quad\Rightarrow\quad \pi_j = \frac{-\sum_{i=1}^{n} q_j(x_i)}{\lambda}$$

$\because \sum_{j=1}^{m}\pi_j = 1 \quad \therefore \lambda = -\sum_{i=1}^{n}\sum_{j=1}^{m} q_j(x_i) = -n$

对于任意一个μ_j，要使得(1)式最大，

则需使得以下优化问题成立.

$$\max_{\mu_j} -\sum_{i=1}^{n}\frac{q_j(x_i)}{\beta_j}|x_i-\mu_j|$$

以上是个凸优化问题，且涉及到绝对值函数（会分段）

故当 $\mu_j = \text{median}\{x_1, x_2, \dots, x_n\}$时，取得全局优解.

部分参考自：https://las.inf.ethz.ch/courses/introml-f18/hw/hw6_solutions.pdf