UCAS

PATTERN RECOGNITION

Assignment 2

October 11, 2019

说　明

1、请选完成所有证明题

2、本次课程的编程题将留在第三次作业中

3、作业用中文撰写，提交时，务必注明题号，鼓励使用LaTeX撰写

4、文档按“学号_姓名.pdf”命名，打包成“学号_姓名.zip”提交

5、本次作业截止时间为2019年10月22日，请在课程网站上及时提交

Question 1

Two random variables $\mathbf{x}$ and $\mathbf{y}$ are called 'statistically independent' if $p(\mathbf{x}, \mathbf{y}|\omega) = p(\mathbf{x}|\omega)p(\mathbf{y}|\omega)$.

(a). Prove that if $x_i - \mu_i$ and $x_j - \mu_j$ are statistically independent (for $i \neq j$), then σ_{ij} as defined in $\sigma_{ij} = \mathbb{E}[(x_i - \mu_i)(x_j - \mu_j)]$ is 0.

(b). Prove that the converse is true for the Gaussian case.

Question 2

Maximum likelihood methods apply to estimates of prior probabilities as well. Let samples be drawn by successive, independent selections of a state of nature ω_i with unknown probability $P(\omega_i)$. Let $z_{ik} = 1s$ if the state of nature for the k-th sample is ω_i and $z_{ik} = 0$ otherwise.

(a). Show that

$$P(z_{i1}, ..., z_{in}|P(\omega_i)) = \prod_{k=1}^{n} P(\omega_i)^{z_{ik}}(1 - P(\omega_i))^{1-z_{ik}}$$

(b). Show that the maximum likelihood estimate for $P(\omega_i)$ is

$$\hat{P}(\omega_i) = \frac{1}{n}\sum_{k=1}^{n} z_{ik}.$$

Question 3

One measure of the difference between two distributions in the same space is the *Kullback-Leibler divergence* of Kullback-Leibler 'distance':

$$D_{KL}(p_1(\mathbf{x}), p_2(\mathbf{x})) = \int p_1(\mathbf{x}) \ln \frac{p_1(\mathbf{x})}{p_2(\mathbf{x})} d\mathbf{x}.$$

(This 'distance', does not obey the requisite symmetry and triangle inequalities for a metric.) Suppose we seek to approximate an arbitrary distribution $p_2(\mathbf{x})$ by a normal $p_1(\mathbf{x}) \sim \mathcal{N}(\mu, \mathbf{\Sigma})$. Show that the values that lead to the smallest Kullback-Leibler divergence are the obvious ones:

$$\mu = \mathbb{E}_2(\mathbf{x}), \mathbf{\Sigma} = \mathbb{E}_2[(\mathbf{x} - \mu)(\mathbf{x} - \mu)^t],$$

where the expectation taken is over the density p_2

Question 4

Let the sample mean $\hat{\mu}_n$ and the sample covariance matrix C_n for a set of n samples $x1, ..., x_n$ (each of which is d-dimensional) be defined by

$$\hat{\mu}_n = \frac{1}{n}\sum_{k=1}^{n} x_k$$

and

$$C_n = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \hat{\mu}_n)(x_k - \hat{\mu}_n)^t.$$

(a). Show that alternative, recursive' techniques for calculating $\hat{}_n$ and C_n based on the successive addition of new samples x_{n+1} can be derived using the recursion relations

$$\hat{\mu}_{n+1} = \hat{\mu}_n + \frac{1}{n+1}(x_{n+1} - \hat{\mu}_n)$$

and

$$C_{n+1} = \frac{n-1}{n}C_n + \frac{1}{n+1}(x_{n+1} - \hat{\mu}_n)(x_{n+1} - \hat{\mu}_n)^t$$

(b). What is the computational complexity of finding $\hat{\mu}_n$ and C_n by these recursive methods?

Question 5

Consider $D = \{(1,1)^t, (3,3)^t, (2,*)^t\}$ sampled from a two-dimensional (separable) distribution $p(x_1, x_2) = p(x_1)p(x_2)$, with

$$p(x_1) = \begin{cases} \frac{1}{\theta_1} e^{-\theta_1 x_1} & \text{if } x_1 \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$p(x_2) = \begin{cases} \frac{1}{\theta_2} & \text{if } x_1 \leq x_2 \leq \theta_2 \\ 0 & \text{otherwise} \end{cases}$$

As usual, $*$ represents a missing feature value. Use EM method to:

(a). Start with an initial estimate $\theta^0 = (2, 4)^t$, calculate Q
(b). Find the θ that maximizes Q

Question 6

Consider training an HMM by the Forward-backward algorithm, for a single sequence of length T where each symbol could be one of c values. What is the computational complexity of a single revision of all values $\hat{a}_{ij}$ and $\hat{b}_{jk}$?