Assignment 2

说明
作业用中文撰写，鼓励使用LaTex.
文档按"学号_姓名.pdf"命名提交.
本次作业截止时间为2020年10月27日，请到课程网站及时提交。

Question 1

Let x have a uniform density

$$p(x|\theta) \sim U(0,\theta) = \begin{cases} 1/\theta & 0 \leq \mathrm{x} \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

1. Suppose that n samples $\mathrm{D} = \{\mathrm{x}_1, \ldots, \mathrm{x}_\mathrm{n}\}$ are drawn independently according to $\mathrm{p}(\mathrm{x}|\theta)$. Show that the maximum likelihood estimate for θ is max[D], i.e., the value of the maximum element in D.
2. Suppose that n = 5 points are drawn from the distribution and the maximum value of which happens to be $\max_{\mathrm{k}} \mathrm{x}_\mathrm{k} = 0.6$. Plot the likelihood $\mathrm{p}(\mathrm{x}|\theta)$ in the range $0 \leq \theta \leq 1$. Explain in words why you do not need to know the values of the other four points.

Question 2

Assume we have training data from a Gaussian distribution of known covariance Σ but unknown mean μ. Suppose further that this mean itself is random, and characterized by a Gaussian density having mean m_0 and covariance Σ_0.

1. What is the MAP estimator for μ?
2. Suppose we transform our coordinates by a linear transform $x' = Ax$, for nonsingular matrix A, and accordingly for other terms. Determine whether your MAP estimator gives the appropriate estimate for the transformed mean μ'. Explain.

Question 3

Consider data $\mathrm{D} = \left\{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}3\\3\end{pmatrix}, \begin{pmatrix}2\\ *\end{pmatrix}\right\}$, sampled from a two-dimensional (separable) distribution $p(x_1, x_2) = p(x_1)p(x_2)$, with (1). As usual, $*$ represents a missing feature value.

$$p(x_1) \sim \begin{cases} \frac{1}{\theta_1} e^{-\theta_1 x_1} & \text{if } x_1 \geq 0 \\ 0 & \text{otherwise} \end{cases} \text{ and } p(x_2) \sim U(0, \theta_2) \begin{cases} \frac{1}{\theta_2} & \text{if } 0 \leq x_2 \leq \theta \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

1. Start with an initial estimate $\theta^0 = \binom{2}{4}$ and analytically calculate $Q(\theta, \theta^0)$— the E step in the EM algorithm. Be sure to consider the normalization of your distribution.
2. Find the θ that maximizes your $Q(\theta, \theta^0)$ — the M step.

Question 4

Consider training an HMM by the Forward-backward algorithm, for a single squence 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 $\hat{b}_{jk}$?

Question 5

Consider a normal $p(x) \sim N(\mu, \sigma^2)$ and Parzen-window function $\varphi(x) \sim N(0,1)$. Show that the Parzen-window estimate

$$p_n(x) = \frac{1}{nh_n} \sum_{i=1}^{n} \varphi\left(\frac{x - x_i}{h_n}\right)$$

has the following properties:

1. $\bar{p}(x) \sim N(\mu, \sigma^2 + h_n^2)$
2. $Var[p_n(x)] \simeq \frac{1}{2nh_n\sqrt{\pi}} p(x)$

Question 6

Explore the effect of r on the accuracy of nearest-neighbor search based on partial distance. Assume we have a large number n of points randomly placed in a d-dimensional hypercube. Suppose we have a test point x, also selected randomly in the hypercume, and find its nearest neighbor. By definition, if we use the full d-dimensional Euclidean distance, we are guaranteed to find its nearest neighbor. Suppose though we use the partial distance

$$D_r(x, x') = \left(\sum_{i=1}^{r} (x_i - x_i{}')^2\right)^{1/2}$$

1. Plot the probability that a partial distance search finds the true closest neighbor of an arbitrary point x as a function of r for fixed n ($1 \leq r \leq d$) for $d = 10$.
2. Consider the effect of r on the accuracy of a nearest-neighbor classifier. Assume we have $n/2$ prototypes from each two categories in a hypercube of length 1 on a side. The density for each category is separable into the product of (linear) ramp functions, highest at one side, and zero at the other side of the range. Thus the density for category ω_1 is highest at $(0, 0, \ldots 0)^t$ and zero at $(1, 1, \ldots, 1)^t$, while the density for ω_2 is highest at $(1, 1, \ldots, 1)^t$ and zero at $(0, 0, \ldots 0)^t$. State by inspection the Bayesian decision boundary.