E1 213 Pattern Recognition and Neural Networks

Practice Problems: Set 2

1. Suppose a class conditional density is normal with mean μ and variance 10. Suppose we assume that the class conditional density is normal with unknown mean and variance 1. Suppose we have large amount of data and do a ML estimate for μ. Would the estimated mean be close to the true mean? What can be said about the error rate of the resulting Bayes classifier?

2. The Bernoulli random variable that we considered takes values 0 and 1. Sometimes we want binary random variables that take values $+1$ or -1. Suppose X is a random variable that takes values $+1$ and -1 with probabilities p and $(1-p)$ respectively. Write the mass function of X with p as a parameter and derive the ML estimate for p. Can you justify the final answer in terms of 'sample mean'? Denote by μ the mean of X. Also write the mass function of X with μ as the parameter.

3. Suppose X is a discrete random variable with mass function

$$f(x \mid p) = (1-p)^{x-1}p, \ x = 1, 2, \cdots$$

where p (with $0 < p < 1$) is the parameter. (This is geometric random variable). What is the ML estimate for p? Suppose we want a Bayesian estimate for p. What is the conjugate prior? What would be the MAP estimate? Is $\sum_i x_i$ a sufficient statistic for p?

4. We want to estimate θ, which is the probability of heads of a coin. The data consists of N tosses of which N_1 are heads. Suppose we want a Bayesian estimate. Suppose our prior density is

$$\begin{aligned} f(\theta) &= 0.5 \ \text{ if } \theta = 0.5 \\ &= 0.5 \ \text{ if } \theta = 0.6 \\ &= 0 \ \text{ otherwise} \end{aligned}$$

Guess the MAP estimate for θ and provide a justification. Then derive the MAP estimate of θ to verify your intuition.

5. Consider a 2-class problem where class conditional densities are normal. We have a large number of feature vectors in our training set.

Hoewever, we do not have class labels for any of these feature vectors. Can we still learn the class conditional densities and implement a Bayes classifier?

6. Let Z be a K-dimensional random vector with joint mass function given by
$$f(\mathbf{z}) = \prod_{i=1}^{K} \rho_i^{z_i}$$
where $\mathbf{z} = [z_1 \ \cdots, \ z_K]^T$ with $z_i \in \{0, 1\}$ and $\sum_i z_i = 1$. The ρ_i are the parameters. (Note that $\rho_i \geq 0$ and $\sum_i \rho_i = 1$). Let X be a random variable whose conditional density conditioned on $\mathbf{z}$ is given by
$$f(x \mid \mathbf{z}) = \prod_{i=1}^{K} (\phi(x \mid \mu_i, \sigma_i))^{z_i}$$
where $\phi(x \mid \mu_i, \sigma_i)$ is a guassian density function with mean μ_i and variance σ_i^2. Show that the (marginal) density of X is a mixture of K Gaussians.

7. We considered the EM algorithm in the context of ML estimation in the sense that it is meant for maximizing the log likelihood. Suppose we want to obtain the MAP estimate and want to make use of hidden or latent variables. As in the notation used in class, let $\mathbf{x}$ be the observed data and let $\mathbf{z}$ be the hidden data. We want to maximize $\ln(f(\theta \mid \mathbf{x}))$ which is the log of the posterior density. Show that we can use the EM algorithm where the E-step computes the same $Q(\theta, \theta^{(k)})$ as earlier but now the M-step maximizes, over θ, $Q(\theta, \theta^{(k)}) + \ln(f(\theta))$ where $f(\theta)$ is the prior density.