Problem set #7

Bayesian Methods

1. Consider a set o iid observations $x_1, \ldots, x_N$ with exponential distribution $p(x_i) = \frac{1}{\alpha} e^{-\frac{x_i}{\alpha}}$. We want to estimate the parameter α assuming that it is a random variable with prior distribution $p(\alpha) = \frac{1}{\beta} e^{-\frac{\alpha}{\beta}}$. Determine
   a) a posteriori distribution of α.
   b) MMSE estimate of α
   c) MAP estimate of α
2. Consider a set o iid observations $x_1, \ldots, x_N$ with normal distribution $N(\mu, \sigma^2)$ with unknown parameter μ (σ is known). Assume that μ is a random variable with normal distribution $N(\mu_0, \sigma_0)$. Determine
   a) a posteriori distribution of μ
   b) MMSE estimate of μ
   c) MAP estimate of μ
3. Assume that the sensor model belongs to the *exponential family* i.e., $p(x|\theta) = h(x)g(\theta)exp(t(x)c(\theta))$. Prove that if the prior distribution is *conjugate prior*

$$p(\theta) = g(\theta)^d \exp(bc(x))$$

   then the a posteriori distribution has the same expression (with different parameters b, d)

$$p(\theta) = g(\theta)^{\tilde{d}} \exp(\tilde{b}c(x)) \qquad \tilde{d} = d + 1, \;\; \tilde{b} = b + t(x)$$

   If we use conjugate priors, the computation of the a posteriori distribution is specially simples.
4. Prove that the following distributions belong to the exponential family and determine the conjugate priors for each of them
   a) exponential distribution
   b) normal distribution with unknown mean
   c) binomial distribution
   d) Rayleigh distribution
5. *(denoising)* Suppose we observe a sum of two unknown signals of length N, $y = x + w$, $x, y, w \in \mathbb{R}^N$. Assume that x has a normal distribution $N(0, R)$ and w is a white Gaussian noise $N(0, \sigma^2 I)$. Determine the a posteriori distribution of x and the MAP estimate of x.
6. Solve this problem using the ML method. Explain why it does not work.

7. Let x be an unknown (scalar) variable with Gaussian distribution $N(x_0, \sigma_0^2)$. Suppose we make a linear coorupted with additive noise $y = cx + v$ where $v \sim N(0, \sigma^2)$. What is the a posteriori distribution of x a the its MAP estimate. (see result bellow)

8. Repeat the previous problem assuming that x, y are random vectors. The prior distribution is now $N(x_0, P)$ and the noise v has a normal distribution $N(0, R)$.

9. *(Kalman filter)* Suppose that the variable we want to estimate changes during time. We know a prior distribution for $x(n)$, $N(\hat{x}(n), P(n)))$. Then $x(n)$ changes according to a linear stochastic model $x(n+1) = Ax(n) + w(n+1)$ where $w(n+1) \sim N(0, Q)$. Finally we observe a variable $y(n+1) = Cx(n+1) + v(n+1)$

   determine the distribution of $x(n+1)$ after the first step.

   determine the distribution of $x(n+1)$ after observing $y(n+1)$

10. *(multiple models)* An observation x was produced by one of the following sensor models $p(x|\theta_1), p(x|\theta_2)$ (we do not know which). Let $k \in \{0, 1\}$ denote the model which generated x and let $P_i = Pr\{k = 1\}$ be the prior distribution of k. Determine the a posteriori distribution of k.

    (this is called the *MAP classifier* and it is optimal in the sense that it minimizes the probability of decision error)

Inference with Gaussian variables: If the random vector $z = (x, y)$ has a joint normal distribution with mean and covariance

$$E\{z\} = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} \qquad Cov\{z\} = \begin{bmatrix} P_{xx} & P_{xy} \\ P_{yx} & P_{yy} \end{bmatrix}$$

the distribution of vector x given vector y is $N(\hat{x}, P)$ such that

$$\hat{x} = \bar{x} + P_{xy} P_{yy} (y - \bar{y})$$
$$P = P_{xx} P_{yy}^{-1} P_{yx}$$

Bom trabalho!
Jorge Salvador Marques