Binghamton University, State University of New York
Department of Electrical and Computer Engineering

EECE 629 – MACHINE PATTERN RECOGNITION (Fall 2017)

Assignment #1

Due Date: 10/03/2017 (Tuesday)

Note: This is an individual assignment, no group submission is accepted.

1. Suppose that we have three colored boxes *r* (red), *b* (blue), and *g* (green). Box *r* contains 3 apples, 4 oranges, and 3 limes; box *b* contains 1 apple, 1 orange, and 0 limes; and box *g* contains 3 apples, 3 oranges, and 4 limes. If a box is chosen at random with probabilities $p(r)=0.2$, $p(b)=0.2$, $p(g)=0.6$, and a piece of fruit is removed from the box (with equal probability of selecting any of the items in the box), then what is the probability of selecting an apple? If we observe that the selected fruit is in fact an orange, what is the probability that it came from the green box?

2. Consider two binary variables *x* and *y* having the joint distribution given in the table below. Please evaluate the following quantities, which *H*[.] denotes the *entropy* and *I*[.] is called the *mutual information*.
   a. $H[x]$
   b. $H[y]$
   c. $H[y|x]$
   d. $H[x|y]$
   e. $H[x,y]$
   f. $I[x,y]$

		y	
		0	1
x	0	1/3	1/3
	1	0	1/3

   Draw a diagram to show the relationship between these various quantities.

3. Please evaluate the *Kullback-Leibler* divergence (refer to Section 1.6.1 in Bishop book) between two Gaussians $p(x) = \mathcal{N}(x|\mu, \Sigma)$ and $q(x) = \mathcal{N}(x|m, \mathbf{L})$.

4. Consider the linear basis function model in Section 3.1 (Bishop book), and suppose that we have already observed *N* data points, so that the posterior distribution over w is given by

$$\mathrm{p(w|t)} = \mathcal{N}(\mathbf{w}|\mathbf{m}_{\mathrm{N}}, \mathbf{S}_{\mathrm{N}})$$

   where

$$\mathbf{m}_{\mathrm{N}} = \mathbf{S}_N(\mathbf{S}_0^{-1}\mathbf{m}_o + \beta\Phi^T\mathbf{t})$$
$$\mathbf{S}_{\mathrm{N}}^{-1} = \mathbf{S}_0^{-1} + \beta\Phi^T\Phi$$

   This posterior can be regarded as the prior for the next observation. By considering an additional data point (x_{N+1}, t_{N+1}), and by completing the square in the exponential, show that the resulting posterior distribution is again given by the equations above but with $\mathbf{S}_\mathrm{N}$ replaced by $\mathbf{S}_{\mathrm{N}+1}$ and $\mathbf{m}_\mathrm{N}$ replaced by $\mathbf{m}_{\mathrm{N}+1}$.

5. Computer Problem.

	ω_1			ω_2			ω_3		
point	x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
1	0.42	-0.087	0.58	-0.4	0.58	0.089	0.83	1.6	-0.014
2	-0.2	-3.3	-3.4	-0.31	0.27	-0.04	1.1	1.6	0.48
3	1.3	-0.32	1.7	0.38	0.055	-0.035	-0.44	-0.41	0.32
4	0.39	0.71	0.23	-0.15	0.53	0.011	0.047	-0.45	1.4
5	-1.6	-5.3	-0.15	-0.35	0.47	0.034	0.28	0.35	3.1
6	-0.029	0.89	-4.7	0.17	0.69	0.1	-0.39	-0.48	0.11
7	-0.23	1.9	2.2	-0.011	0.55	-0.18	0.34	-0.079	0.14
8	0.27	-0.3	-0.87	-0.27	0.61	0.12	-0.3	-0.22	2.2
9	-1.9	0.76	-2.1	-0.065	0.49	0.0012	1.1	1.2	-0.46
10	0.87	-1.0	-2.6	-0.12	0.054	-0.063	0.18	-0.11	-0.49

Using the dataset above, consider Gaussian density models in different dimensions.

a. Write a program to find the maximum likelihood value $\hat{\mu}$ and $\hat{\sigma}^2$. Apply your program individually to each of the three features x_i of category w_1 in the table.

b. Modify your program to apply to two-dimensional Gaussian data $p(\mathrm{x}) \sim \mathcal{N}(\mu, \Sigma)$. Apply your data to each of the three possible pairings of two features for w_1.

c. Modify your program to apply to three-dimensional Gaussian data. Apply your data to the full three-dimensional data for w_1.

d. Assume your three-dimensional model is separable, so that $\Sigma = diag(\sigma_1^2, \sigma_2^2, \sigma_3^2)$. Write a program to estimate the mean and the diagonal components of Σ. Apply your program to the data in w_2.

e. Compare your results for the mean of each feature μ_i calculated in the above ways. Explain why they are the same or different.

f. Compare your results for the variance of each feature σ_i^2 calculated in the above ways. Explain why they are the same or different.