CPSC 633-600 (Total 100)
Homework #3

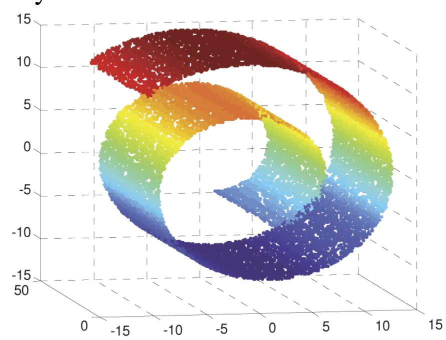
See course web page for the **due date**.

Instructor: Yoonsuck Choe

April 6, 2013

1 Dimensionality reduction

Problem 1 (Written: 5 pts): Can PCA be an effective method for analyzing data sets like the Swiss role data set (below)? Explain why or why not.



Problem 2 (Written: 10 pts): Explain why it is important to select a proper ϵ value in Isomap. For example, what would happen when ϵ is too large? Discuss in the context of the Swiss roll data set shown above.

2 Local Methods

Problem 3 (Written: 10 pts): The SOM, given an input vector $\vec{x}$ and the best matching unit index $i(\vec{x})$, the learning rule for the reference vector for unit j is:

$$\vec{w}_j \leftarrow \vec{w}_j + \eta h(j, i(\vec{x}))(\vec{x} - \vec{w}_j)$$

(1) When $\eta = 1$ and the neighborhood function covers the entire SOM map uniformly, i.e., $h(j, i(\vec{x})) = 1$, what does the above update rule reduce to (just plug in those values above)?

(2) When you train the SOM map using the above rule with inputs $\vec{x}_1, \vec{x}_2, ...\vec{x}_n$, in that sequence, what would $\vec{w}_j$ be after you have gone through the n inputs exactly one time?

(3) In the above case, you will see that $\vec{w}_j$ for all j will be the same. How can you avoid this by changing certain assumption in (1)?

Problem 4 (Written: 5 pts): In radial basis function networks, among (a) the RBF units, (b) output units, and (c) RBF-to-output connections, which part is associated with "one shot" learning, i.e., the part that does not require iterative update using a learning rule?

3 MDL

Problem 5 (Written: 10 pts): (1) Given four messsages (A, B, C, D), without any knowledge of the probability of each message, how many bits would be needed to represent each message? (2) Now consider the following probability: P(A) = 0.5, P(B) = 0.25, P(C)=0.125, P(D)=0.125, calculate the average bits per message (i.e., expected length of the code), and compare to (1).

4 Conditional Independence

Problem 6 (Written: 10 pts): Consider three binary variables $a, b, c \in \{0, 1\}$ having the joint probability distribution given in the table below. Show by direct evaluation that this distribution has the property that a and b are dependent, so that $P(a, b) \neq P(a)p(b)$, but that they become independent when conditioned on c, so that $P(a, b|c) = P(a|c)p(b|c)$ for both $c = 0$ and $c = 1$ [from C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006].

Table 1: Joint Probability

a	b	c	P(a,b,c)
0	0	0	0.192
0	0	1	0.144
0	1	0	0.048
0	1	1	0.216
1	0	0	0.192
1	0	1	0.064
1	1	0	0.048
1	1	1	0.096

Problem 7 (Written: 10 pts): How is the above result related to the concept of conditional independence $P(a|b, c) = P(a|c)$, a is independent from b given c? (derive from $P(a, b|c) = P(a|c)p(b|c)$: **DO NOT** simply plug in the probability values from above).

5 Naive Bayes Classifier

Given two attributes a_1 and a_2 and the class v, consider the following conditional probabilities:

a_1	a_2	$P(a_1, a_2 \vert v = \oplus)$
1	1	1/8
1	2	3/8
2	1	3/8
2	2	1/8

a_1	a_2	$P(a_1, a_2 \mid v = \ominus)$
1	1	5/8
1	2	~~1/8~~ 0
2	1	1/8
2	2	2/8

Assume that $P(v = \oplus) = 1/2$ and $P(v = \ominus) = 1/2$.

Problem 8 (Written: 10 pts): Calculate the following probabilities (some can be directly taken from the tables above):

(1) $P(a_1 = 2, a_2 = 1|v = \oplus)$

(2) $P(a_1 = 2, a_2 = 1|v = \ominus)$

(3) $P(a_1 = 2|v = \oplus)$

(4) $P(a_1 = 2|v = \ominus)$

(5) $P(a_2 = 1|v = \oplus)$

(6) $P(a_2 = 1|v = \ominus)$

(7) $P(a_1 = 2, a_2 = 1)$

Problem 9 (Written: 10 pts): With the above,

(1) Calculate $P(v = \oplus|a_1 = 2, a_2 = 1)$ using Bayes rule:

$$\frac{P(a_1 = 2, a_2 = 1|v = \oplus)P(\oplus)}{P(a_1 = 2, a_2 = 1)}$$

(2) Repeat (1) for class $v = \ominus$.

(3) Based on (1) and (2), what should be the decision?

(4) Calculate $P(v = \oplus|a_1 = 2, a_2 = 1)$ using Naive Bayes:

$$\frac{P(a_1 = 2|v = \oplus)P(a_2 = 1|v = \oplus)P(\oplus)}{P(a_1 = 2, a_2 = 1)}$$

(5) Repeat (4) for class $v = \ominus$.

(6) Based on (4) and (5), what should be the decision?

(7) Do the decisions in (3) and (6) differ? Explain why.

6 Bayesian Belief Network

Problem 10 (Written: 10 pts): Based on Table 1 (problem 6) draw a Bayesian Belief Network (BBN) that correctly represents the conditional independence relation.

Problem 11 (Written: 10 pts): For each node in the BBN, calculate the conditional probability table, again from Table 1.

Problem 12 (Program: 5 pts): [optional: 5 point extra credit] Write a program to do Monte Carlo estimation of the conditional probability $P(a|b)$ (generate three sets with 10, 100, and 1000 examples). Compare the simulation results from the three sets, and also calculate (by hand) the true $P(a|b)$ from Table 1. Are the results comparable?