# CPSC 633-600 (Total 100 + 20 points extra credit)
# Homework #4

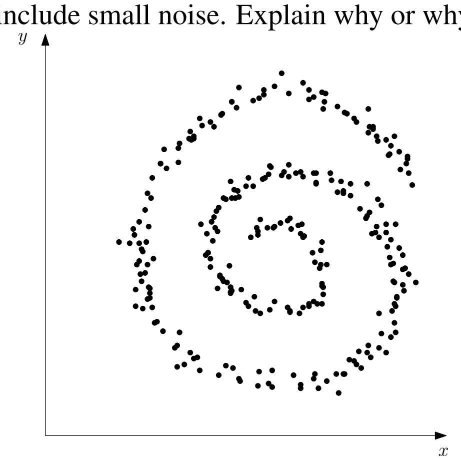
See course web page for the **due date**.

Instructor: Yoonsuck Choe

April 8, 2015 - REVISED (4/9 1pm)

## 1 Dimensionality reduction

**Problem 1 (Written: 5 pts):** Can PCA be an effective method for analyzing data sets like the data set (below)? Note that the data may include small noise. Explain why or why not.




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

## 2 Local Methods

**Problem 3 (Written: 15 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) The learning rate is fixed $\eta = 1$.

(2) Let $h(j, i(\vec{x})) = 1$, for the best matching unit $j = i(\vec{x})$, $h(j, i(\vec{x})) = 2/3$ for its immediate neighbor ($j = i(\vec{x}) \pm 1$) and $h(j, i(\vec{x})) = 1/3$ for its second-order neighbor ($j = i(\vec{x}) \pm 2$). For all the rest, $h(j, i(\vec{x})) = 0$.

(3) Consider a 1-D SOM with 7 units with the following weight vectors. Plot the vectors and connect them according to the order given below ($\vec{w}_1$ connected to $\vec{w}_2$, etc.).

| | $w_1$ | $w_2$ |
|---|---|---|
| $\vec{w}_1$ | 0 | 3 |
| $\vec{w}_2$ | 9 | 8 |
| $\vec{w}_3$ | 1 | 5 |
| $\vec{w}_4$ | 4 | 3 |
| $\vec{w}_5$ | 3 | 0 |
| $\vec{w}_6$ | 0 | 7 |
| $\vec{w}_7$ | 7 | 4 |

(4) Given an input vector $\vec{x} = (5, 3)$, plot how the weight vectors change after one iteration of training. Plot in the same graph as (3) above.

**Problem 4 (Written: 10 pts):** In radial basis function networks, among (a) the RBF units, (b) output units, and (c) RBF-to-output connections, which part is associated the most with "local" in "local learning"?

## 3 Conditional Independence

**Problem 5 (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 $b$, so that $P(a, b|c) = P(a|c)p(b|c)$ for both $c = 0$ and $c = 1$ [adapted from C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006].

| a | b | c | P(a,b,c) |
|---|---|---|---|
| 0 | 0 | 0 | 0.36 |
| 0 | 0 | 1 | 0.04 |
| 0 | 1 | 0 | 0.04 |
| 0 | 1 | 1 | 0.06 |
| 1 | 0 | 0 | 0.18 |
| 1 | 0 | 1 | 0.12 |
| 1 | 1 | 0 | 0.02 |
| 1 | 1 | 1 | 0.18 |

Tabulate your results in the following format:

| a | b | p(a,b) | p(a) p(b) |
|---|---|---|---|
| 0 | 0 | | |
| 0 | 1 | | |
| 1 | 0 | | |
| 1 | 1 | | |

| a | b | c | p(a,b\|c) | p(a\|c) p(b\|c) |
|---|---|---|---|---|
| 0 | 0 | 0 | | |
| 0 | 0 | 1 | | |
| 0 | 1 | 0 | | |
| 0 | 1 | 1 | | |
| 1 | 0 | 0 | | |
| 1 | 0 | 1 | | |
| 1 | 1 | 0 | | |
| 1 | 1 | 1 | | |

**Problem 6 (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 $P(a|b, c) = P(a|c)$ from $P(a, b|c) =$

$P(a|c)p(b|c)$: **DO NOT** simply plug in the probability values from above).

# 4 Naive Bayes Classifier

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

| $a_1$ | $a_2$ | $P(a_1, a_2 \| v = \oplus)$ |
|---|---|---|
| 1 | 1 | 0.12 |
| 1 | 2 | 0.08 |
| 2 | 1 | 0.08 |
| 2 | 2 | 0.72 |

| $a_1$ | $a_2$ | $P(a_1, a_2 \| v = \ominus)$ |
|---|---|---|
| 1 | 1 | 0.04 |
| 1 | 2 | 0.16 |
| 2 | 1 | 0.40 |
| 2 | 2 | 0.40 |

Assume that $P(v = \oplus) = 0.5$ and $P(v = \ominus) = 0.5$.

**Problem 7 (Written: 10 pts):** Calculate the following probabilities (some can be directly taken from the tables above):

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

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

(3) $P(a_1 = 1|v = \oplus)$

(4) $P(a_1 = 1|v = \ominus)$

(5) $P(a_2 = 2|v = \oplus)$

(6) $P(a_2 = 2|v = \ominus)$

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

**Problem 8 (Written: 10 pts):** With the above,

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

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

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

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

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

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

(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 this is a good news or bad news for naive Bayes.

# 5 Bayesian Belief Network

**Problem 9 (Written: 10 pts):** Based on the table in problem 5 draw a Bayesian Belief Network (BBN) that correctly represents the conditional independence relation.

**Problem 10 (Written: 10 pts):** For each node in the BBN, calculate the probability table (prior or conditional probability) based on problem 5.

**Problem 11 (Program: 20 pts):** **[optional: 20 point extra credit]** Write a program to do Monte Carlo estimation of the conditional probability $P(a|b)$. Generate three set of samples, with 10, 100, and 1000 instances, respectively, and empirically estimate $P(a|b)$. Compare the simulation results from the three sets, and also calculate (by hand) the true $P(a|b)$ from the table in problem 6. Are the results comparable?