Istanbul Technical University- Fall 2007

BBL514E Pattern Classification

Homework 1

Purpose: Better understanding of PR/ML basics. Use of Matlab.

Total worth: 10% of your grade.

Handed out: Tuesday, Sept 18, 2007.

Due: Monday, Oct 15, 2007 9:55am. (please submit via ninova!)

Instructor: Zehra Cataltepe (cataltepe@itu.edu.tr)

Policy: Collaboration in the form of discussions is acceptable, but you should write your own answer/code by yourself. Cheating is highly discouraged for it could mean a zero or negative grade from the homework.

If a question is not clear, please let me know (via email, during office hour or in class).

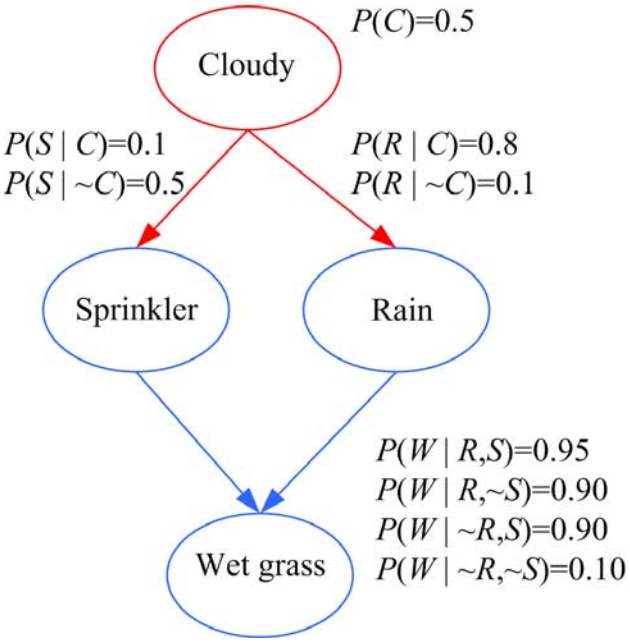
Submission Instructions: Please submit through the class ninova site.
Please zip and upload all your files using filename studentID_HW1.zip. You must provide all matlab functions you wrote with your zipped file. Functions you do not submit may cause you lose a portion of your grade. You must also include a .doc or pdf file with answers to the questions and how to call your matlab functions for each question.

QUESTIONS:

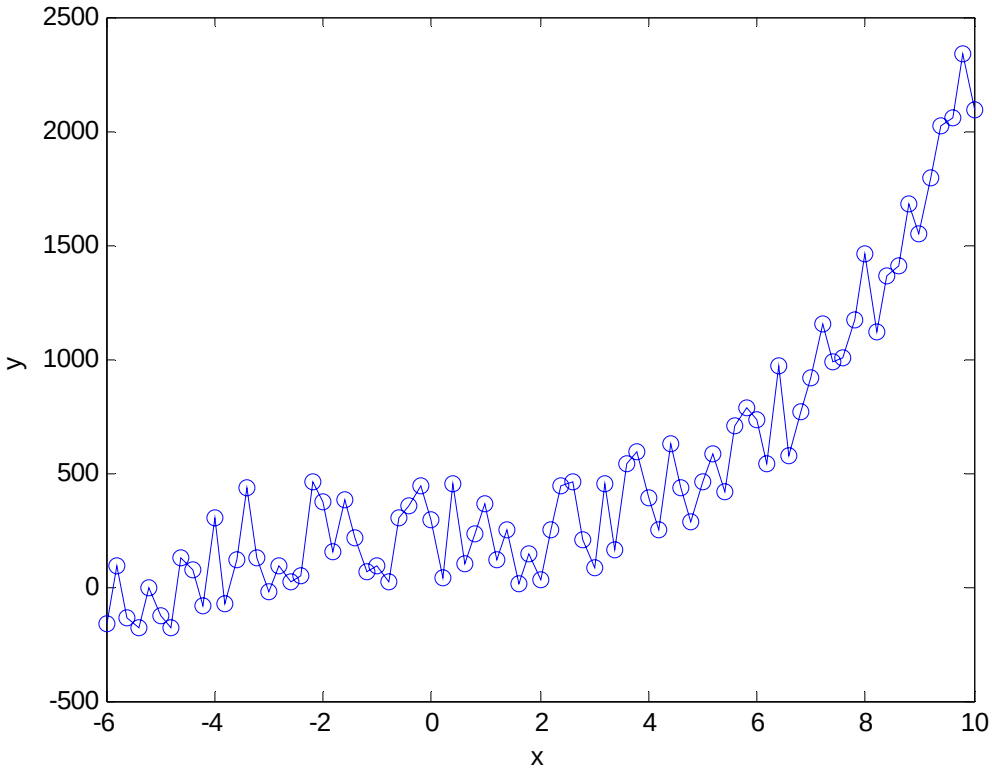
Q1) [20 points][Central Limit Theorem]
Q1a) [8 points] Consider a univariate random variable x distributed according to $U\left(10-\frac{\sqrt{3}}{2},10+\frac{\sqrt{3}}{2}\right)$ (i.e. minimum and maximum values x can take are $10-\frac{\sqrt{3}}{2}$ and $10+\frac{\sqrt{3}}{2}$ respectively.)
Q1aa) Using matlab, draw N=1 samples of x from the distribution, compute the mean and plot the histogram of the means for 500 different experiments.
Q1ab) Repeat Q1aa) for N=4 samples.
Q1ac) Repeat Q1aa) for N=10 samples.
Q1ad) Repeat Q1aa) for N=100 samples.

Q1b) [8 points] Repeat Q1a), but instead of uniform density, use N(10,1) (i.e. x is normally distributed with mean 10 and variance 1.)
Q1c) [4 points] What are the differences and similarities between the plots you see in 1a) and 1b)? Why?
Hint: Scale x and y axis of plots in Q1a) and b) to the same range in order to be able to compare them.

Q2) [20 points] [Bayesian Networks]
Compute P[W] for the Bayesian Network shown below.
(C: Cloudy, S: Sprinkler was on, R: It rained, W: Grass is wet, ~ opposite of an event, for example, ~W: Grass is not wet.)



Q3: [30 points] [10-fold cross validation, pseudo inverse, overfitting]



Given the x (input) and y (output) values in the file X.txt

a) [3 points] Partition all data randomly into 10 folds and produce 10 different training-validation set pairs.

b) [2 points] Normalize your training inputs and outputs by using training sample mean and std deviation (see matlabExercises.m).

c) [10 points]For each of the 10 training sets, compute the hypothesis that minimizes the training error $E(w) = \frac{1}{N}\sum_{t=1}^{N}\left(y^t - g(x^t, w)\right)^2$

for the following hypotheses classes:

i. $g(x,w) = w_0 + x*w_1$

ii. $g(x,w) = w_0 + x*w_1 + x^2*w_2 + x^3*w_3$

iii. $g(x,w) = w_0 + x*w_1 + x^2*w_2 + x^3*w_3 + x^4*w_4 + x^5*w_5$

iv. $g(x,w) = w_0 + x*w_1 + x^2*w_2 + x^3*w_3 + x^4*w_4 + x^5*w_5 + \ldots. x^{50}*w_{50}$

Hint: For pseudo inverse see the first week lecture notes on review of algebra.

d) [10 points] For the hypothesis that minimizes the training error, plot

   a. the mean and std/sqrt(10) (i.e. errorbar) of training and validation errors (use errorbar() function of matlab). (x axis=1..4 (hypothesis class), y axis=mean error over 10 folds)

   b. Plot also the training input and outputs and the minimum train error hypothesis outputs for each hypothesis class above. (4 plots, 10 hypotheses on each plot.)

e) [5 points] Which hypothesis class would you choose among i...iv and why?

Q4: [30 points] [Bayesian Decision Theory]

Assume a discriminant function of the form:

$$g_i(x) = \ln p(x \mid w_i) + \ln P(w_i)$$

which achieves the minimum error classification. Assume that $x \in R$, for class 1: $x \sim N(\mu_1, \sigma)$ and for class 2: $x \sim N(\mu_2, \sigma)$. Also assume that $P(w_1) = P(w_2)$.

a) [10 points] Derive the discriminant functions $g_1(x)$ and $g_2(x)$.
b) [8 points] What is equation for the separating surface?
c) [8 points] For $\mu_1 = 5$, $\mu_2 = 15$ and $\sigma = 5$, plot the pdf's of the two classes inputs and also the separating surface.
d) [4 points] If $P(w_1)$ increases to 0.8, where would the separating surface be?