Problem Set 1

MAS 622J/1.126J: Pattern Recognition and Analysis

Due Monday, 15 September 2008

[Note: All instructions to plot data or write a program should be carried out using either Python accompanied by the `matplotlib` package or Matlab. Feel free to use either or both, but in order to maintain a reasonable level of consistency and simplicity we ask that you do not use other software tools.]

Problem 1: Why?

a. Describe an application of pattern recognition related to your research. What are the features? What is the decision to be made? Speculate on how one might solve the problem. Limit your answer to a page.

b. In the same way, describe an application of pattern recognition you would be interested in pursuing for fun in your life outside of work.

Problem 2: Probability Warm-Up

Let X and Y be random variables. Let $\mu_X \equiv \mathrm{E}[X]$ denote the expected value of X and $\sigma_X^2 \equiv \mathrm{E}[(X - \mu_X)^2] = \mathrm{E}[X^2] - \mu_X^2$ denote the variance of X. a and b are constant values. Use excruciating detail to answer the following:

a. Show $\mathrm{E}[aX + bY] = a\mathrm{E}[X] + b\mathrm{E}[Y]$.

b. Show that independent implies uncorrelated.

c. Show that uncorrelated does not imply independent.

d. Let $Z = aX + bY$. Show that if X and Y are uncorrelated, then $\sigma_Z^2 = a^2\sigma_X^2 + b^2\sigma_Y^2$.

e. Let X_i $(i = 1, ..., n)$ be random variables independently drawn from the same probability distribution with mean μ_X and variance σ_X^2. For the sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, show the following: (i) $\mathrm{E}[\overline{X}] = \mu_X$. (ii) $\mathrm{Var}[\overline{X}]$ (variance of the sample mean) $= \sigma_X^2/n$. Note that this is different from the sample variance $s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

f. The conditional expected value $\mathrm{E}(X|Y)$ is a random variable in its own right, whose value depends on the value of Y. Notice that the conditional expected value of X given the event Y = y is a function of y. If we write $E(X|Y=y) = g(y)$ then the random variable $\mathrm{E}[X|Y = y] = \sum_x xP(x|Y=y)$ is just $g(Y)$. Show $\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X|Y]]$ and $\mathrm{E}[Y] = \mathrm{E}[\mathrm{E}[Y|X]]$.

g. For a real value function f and discrete random variables X and Y, $\mathrm{E}[f(X,Y)] = \sum_x \sum_y f(x,y)P(x,y)$. Show $\mathrm{E}[f(X,Y)] = \mathrm{E}[\mathrm{E}[f(X,Y)|Y]]$.

h. Let X_1 and X_2 be independent and identically distributed continuous random variables. Can $\Pr[X_1 \leq X_2]$ be calculated? If so, find its value. If not, explain.

i. Let X_1 and X_2 be independent and identically distributed discrete random variables. Can $\Pr[X_1 \leq X_2]$ be calculated? If so, find its value. If not, explain.

Problem 3: Teatime with Gauss and Bayes

Let $p(x,y) = \frac{1}{2\pi\alpha\beta} e^{-\left(\frac{(y-\mu)^2}{2\alpha^2} + \frac{(x-y)^2}{2\beta^2}\right)}$.

a. Find $p(x)$, $p(y)$, $p(x|y)$, and $p(y|x)$. In addition, give a brief description of each of these distributions.

b. Let $\mu = 0$, $\alpha = 20$, and $\beta = 2.5$. Plot $p(y)$ and $p(y|x = 10.5)$ for a reasonable range of y. What is the difference between these two distributions?

Problem 4: Covariance Matrix

Let $\Lambda_X = \begin{bmatrix} 37 & -15 \\ -15 & 37 \end{bmatrix}$.

a. Verify that Λ_X is a valid covariance matrix.

b. Find the eigenvalues and eigenvectors of Λ_X by hand. Show all your work.

c. Write a program to find and verify the eigenvalues and eigenvectors of Λ_X.

d. We provide 200 data points sampled from the distribution $\mathcal{N}(0, \Lambda_X)$. Download the dataset from the course website and plot the data points. Project the data onto the covariance matrix eigenvectors and plot the transformed data. What is the difference between the two plots?

Problem 5: Distribution Linearity

Let X_1 and X_2 be i.i.d. according to

$$p(x_i) = \begin{cases} 1, \text{for } 0 \leq x_i \leq 1 \\ 0, \text{otherwise} \end{cases} \quad \text{for } i = 1, 2$$

Let $Y = X_1 + X_2$.

a. Find an expression for $p(y)$. Plot $p(y)$ for some reasonable range of y.

b. Find an expression for $p(x_1|y)$. Plot $p(x_1|y)$ as a function of x_1 with y treated as a known parameter for some reasonable value of y and some reasonable range of x_1.

c. Repeat the parts above, this time letting X_1 and X_2 be i.i.d. according to $\mathcal{N}(0, 1)$.

d. What was the point of this problem? Hint: check out the title.

Problem 6: Probabilistic Modeling

Let $x \in \{0, 1\}$ denote a person's affective state ($x = 0$ for "positive-feeling state", and $x = 1$ for "negative-feeling state"). The person feels positive with probability θ_1. Suppose that an affect-tagging system (or a robot) recognizes her feeling state and reports the observed state (variable y) to you. But this system is unreliable and obtains the correct result with probability θ_2.

a. Represent the joint probability distribution $P(x, y|\theta)$ for all x, y (a 2x2 matrix) as a function of the parameters $\theta = (\theta_1, \theta_2)$.

b. The Maximum Likelihood estimation criterion for the parameter θ is defined as:

$$\widehat{\theta}_{ML} = \arg\max_{\theta} L(t_1, ..., t_n; \theta) = \arg\max_{\theta} \prod_{i=1}^{n} p(t_i|\theta)$$

where we have assumed that each data point t_i is drawn independently from the same distribution so that the likelihood of the data is $L(t_1, ..., t_n; \theta) = \prod_{i=1}^{n} p(t_i|\theta)$. Likelihood is viewed as a function of the parameters, which depends on the data. Since the above expression can be technically challenging, we maximize the log-likelihood $\log L(t_1, ..., t_n; \theta)$ instead of likelihood. Note that any monotonically increasing function (i.e., log function) of the likelihood has the same maxima. Thus,

$$\widehat{\theta}_{ML} = \arg\max_{\theta} \log L(t_1, ..., t_n; \theta) = \arg\max_{\theta} \sum_{i=1}^{n} \log p(t_i|\theta)$$

Suppose we get the following joint observations $t = (x, y)$.

x	y
1	0
1	1
0	0
1	1
1	0
0	1
0	0

What are the maximum-likelihood (ML) values of θ_1 and θ_2? (*Hint.* Since $P(x, y|\theta) = P(y|x, \theta_2)P(x|\theta_1)$, the estimation of the two parameters can be done separately in the log-likelihood criterion.)

Problem 7: Monty Hall

To get credit for this problem, you must not only write your own correct solution, but also write a computer simulation (in either Matlab or Python) of the process of playing this game:

Suppose I hide the ring of power in one of three identical boxes while you weren't looking. The other two boxes remain empty. After hiding he ring of power, I ask you to guess which box it's in. I know which box it's in and, after you've made your guess, I deliberately open the lid of an empty box, which is one of the two boxes you did not choose. Thus, the ring of power is either in the box you chose or the remaining closed box you did not choose. Once you have made your initial choice and I've revealed to you an empty box, I then give you the opportunity to change your mind – you can either stick with your original choice, or choose the unopened box. You get to keep the contents of whichever box you finally decide upon.

- What choice should you make in order to maximize your chances of receiving the ring of power? Explain your answer.
- Write a simulation. There are two choices in this game for the contestant in this game: (1) choice of box, (2) choice of whether or not to switch. In your simulation, first let the host choose a random box to place the ring of power. Show a trace of your program's output for a single game play, as well as a cumulative probability of winning for 1000 rounds of the two policies (1) to choose a random box and then switch and (2) to choose a random box and not switch.