Problem Set 1

MAS 622J/1.126J: Pattern Recognition and Analysis

Due Thursday, 23 September 2010

[Note: All instructions to plot data or write a program should be carried out using Matlab. In order to maintain a reasonable level of consistency and simplicity we ask that you do not use other software tools.]

If you collaborated with other members of the class, please write their names at the end of the assignment.

Problem 1: Why? [5 points]

Limit your answer to problem 1 to a page.

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.

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.

Solution: Refer to examples discussed in lecture.

Problem 2: Probability Warm-Up [20 points]

Let X and Y be discrete random variables. Let μ_X denote the expected value of X and σ_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 $\sigma_X^2 = \mathrm{E}[X^2] - \mu_X^2$.

c. Show that independent implies uncorrelated.

d. Show that uncorrelated does not imply independent.

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

f. 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 $s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

g. Let X_1 and X_2 be independent and identically distributed (i.i.d) continuous random variables. Can $\Pr[X_1 \leq X_2]$ be calculated? If so, find its value. If not, explain. Hint 1: Remember that for a continuous variable $\Pr[X_1 = k] = 0$, for any value of k. Hint 2: Remember the definition of i.i.d. variables.

h. 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.

Solution:

a. The following is for continuous random variables. A similar argument holds for continuous random variables.

$$\begin{aligned}
\mathrm{E}[aX + bY] &= \sum_{xy}(ax + by)\,p(x, y) \\
&= a\sum_{xy} x\,p(x, y) + b\sum_{xy} y\,p(x, y) \\
&= a\sum_{x} x\,p(x)\,dx + b\sum_{y} y\,p(y)\,dy \\
&= a\mathrm{E}[X] + b\mathrm{E}[Y]
\end{aligned}$$

b. Making use of the definition of variance and the previous part, we have:

$$\begin{aligned}
\sigma_X^2 &= \mathrm{E}[(X - \mu_X)^2] \\
&= \mathrm{E}[X^2 - 2\mu_X X + \mu_X^2] \\
&= \mathrm{E}[X^2] - \mathrm{E}[2\mu_X X] + \mathrm{E}[\mu_X^2] \\
&= \mathrm{E}[X^2] - 2\mu_X \mathrm{E}[X] + \mu_X^2 \\
&= \mathrm{E}[X^2] - 2\mu_X\mu_X + \mu_X^2 \\
&= \mathrm{E}[X^2] - 2\mu_X^2 + \mu_X^2 \\
&= \mathrm{E}[X^2] - \mu_X^2
\end{aligned}$$

c. In order to check if two discrete random variables X and Y are uncorrelated, we have to proof $\sigma_{xy} = 0$ (the same holds for continuous random variables.)

From the previous question:

$$\sigma_{X,Y}^2 = \mathrm{E}[XY] - \mu_X\mu_Y$$

If two variables are independent:

$$\begin{aligned}
\mathrm{E}[XY] &= \sum_{xy} xy\, p(x,y) \\
&= \sum_{xy} xy\, p(x)\, p(y) \\
&= \sum_{x} x\, p(x) \sum_{y} y\, p(y) \\
&= \mathrm{E}[X]\, \mathrm{E}[Y]
\end{aligned}$$

Finally,

$$\sigma^2_{X,Y} = \mathrm{E}[X]\,\mathrm{E}[Y] - \mu_X \mu_Y = 0$$

d. To proof this, we need to find one case where 1) $p(xy) \neq p(x)p(y)$ and 2) $\sigma_{xy} = 0$ are satisfied. One possible solution is as follows:

Suppose we have the discrete random variables X and Y, and we observed all possibilities:

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

If we look at the case where $x = 1$ and $y = 1$, 1) is satisfied:

$$\begin{aligned}
p(x = 1\, y = 1) &= \frac{2}{6} = \frac{1}{3} \\
p(x = 1)p(y = 1) &= \frac{2}{6}\frac{4}{6} = \frac{2}{9}
\end{aligned}$$

Now it is easy to verify that 2) is also satified:

$$\sigma^2_{X,Y} = \mathrm{E}[XY] - \mu_X \mu_Y = 0 - 0\frac{4}{6} = 0$$

e. Given that $Z = aX + bY$ and that X and Y are uncorrelated, we have

$$\begin{aligned}
\sigma^2_Z &= \mathrm{E}[(Z - \mu_Z)^2] \\
&= \mathrm{E}[Z^2] - \mu_Z^2 \\
&= \mathrm{E}[(aX + bY)^2] - (a\mu_X + b\mu_Y)^2 \\
&= \mathrm{E}[a^2X^2 + 2abXY + b^2Y^2] - (a^2\mu_X^2 + 2ab\mu_X\mu_Y + b^2\mu_Y^2) \\
&= a^2\mathrm{E}[X^2] + 2ab\mathrm{E}[XY] + b^2\mathrm{E}[Y^2] - a^2\mu_X^2 - 2ab\mu_X\mu_Y - b^2\mu_Y^2 \\
&= a^2(\mathrm{E}[X^2] - \mu_X^2) + 2ab(\mathrm{E}[XY] - \mu_X\mu_Y) + b^2(\mathrm{E}[Y^2] - \mu_Y^2) \\
&= a^2\sigma_X^2 + 2ab\sigma^2_{XY} + b^2\sigma_Y^2 \\
&= a^2\sigma_X^2 + b^2\sigma_Y^2,
\end{aligned}$$

where only the last equality depends on X and Y being uncorrelated.

f. Using the result of (a) and the fact $\mathrm{E}[X_i] = \mu_X$,

$$\mathrm{E}[\bar{X}] = \mathrm{E}[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[X_i] = \frac{1}{n}\; n\;\mu_X = \mu_X$$

Also, using the result of (d) and the fact $\mathrm{Var}[X_i] = \sigma_X^2$

$$\mathrm{Var}[\bar{X}] = \mathrm{Var}[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}[X_i] = \frac{1}{n^2}\; n\;\sigma_X^2 = \sigma_X^2/n$$

g. Given that X_1 and X_2 are continuous random variables, we know that $\Pr[X_1 = x] = 0$ and $\Pr[X_2 = x] = 0$ for any value of x. Thus,

$$\Pr[X_1 \leq X_2] = \Pr[X_1 < X_2].$$

Given that X_1 and X_2 are i.i.d., we know that replacing X_1 with X_2 and X_2 with X_1 will have no effect on the world. In particular, we know that

$$\Pr[X_1 < X_2] = \Pr[X_2 < X_1].$$

However, since probabilities must sum to one, we have

$$\Pr[X_1 < X_2] + \Pr[X_2 < X_1] = 1.$$

Thus,

$$\Pr[X_1 \leq X_2] = \frac{1}{2}.$$

h. For discrete random variables, unlike the continuous case above, we need to know the distributions of X_1 and X_2 in order to find $\Pr[X_1 = x]$ and $\Pr[X_2 = x]$. Thus, the argument we used above fails. In general, it is not possible to find $\Pr[X_1 \leq X_2]$ without knowledge of the distributions of both X_1 and X_2.

Problem 3: Teatime with Gauss and Bayes [20 points]

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 $\mu = 0$, $\alpha = 15$, and $\beta = 3$. Plot $p(y)$ and $p(y|x = 9)$ for a reasonable range of y. What is the difference between these two distributions?

Solution:

a. To find $p(y)$, simply factor $p(x,y)$ and then integrate over x:

$$\begin{aligned}
p(y) &= \int_{-\infty}^{\infty} p(x,y)\,dx \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\alpha\beta} e^{-\left(\frac{(y-\mu)^2}{2\alpha^2}+\frac{(x-y)^2}{2\beta^2}\right)}\,dx \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\alpha\beta} e^{-\frac{(y-\mu)^2}{2\alpha^2}}\, e^{-\frac{(x-y)^2}{2\beta^2}}\,dx \\
&= \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{(y-\mu)^2}{2\alpha^2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\beta^2}} e^{-\frac{(x-y)^2}{2\beta^2}}\,dx \\
&= \frac{1}{\sqrt{2\pi\alpha^2}} e^{-\frac{(y-\mu)^2}{2\alpha^2}} \\
&= \mathcal{N}(\mu,\alpha^2)
\end{aligned}$$

The integral goes to 1 because it is of the form of a probability distribution integrated over the entire domain. To find $p(x|y)$, divide $p(x,y)$ by $p(y)$:

$$\begin{aligned}
p(x|y) &= \frac{p(x,y)}{p(y)} \\
&= \frac{1}{\sqrt{2\pi\beta^2}} e^{-\frac{(x-y)^2}{2\beta^2}} \\
&= \mathcal{N}(y,\beta^2)
\end{aligned}$$

Finding $p(x)$ and $p(y|x)$ follows essentially the same procedure, but the algebra is more involved and requires completing the square in the exponent.

$$\begin{aligned}
p(x) &= \int_{-\infty}^{\infty} p(x,y)\,dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\alpha\beta} e^{-\left(\frac{(y-\mu)^2}{2\alpha^2}+\frac{(x-y)^2}{2\beta^2}\right)}\,dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\alpha\beta} e^{-\left(\frac{\beta^2(y-\mu)^2+\alpha^2(x-y)^2}{2\alpha^2\beta^2}\right)}\,dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\alpha\beta} e^{-\left(\frac{\beta^2y^2-2\beta^2\mu y+\beta^2\mu^2+\alpha^2x^2-2\alpha^2xy+\alpha^2y^2}{2\alpha^2\beta^2}\right)}\,dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\alpha\beta} e^{-\left(\frac{(\alpha^2+\beta^2)y^2-2(\alpha^2x+\beta^2\mu)y+(\beta^2\mu^2+\alpha^2x^2)}{2\alpha^2\beta^2}\right)}\,dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\alpha\beta} e^{-\left(\frac{y^2-2\frac{\alpha^2x+\beta^2\mu}{\alpha^2+\beta^2}y+\frac{\beta^2\mu^2+\alpha^2x^2}{\alpha^2+\beta^2}}{2\frac{\alpha^2\beta^2}{\alpha^2+\beta^2}}\right)}\,dy
\end{aligned}$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\alpha\beta} e^{-\left(\frac{y^2 - 2\frac{\alpha^2 x + \beta^2\mu}{\alpha^2+\beta^2} y + \left(\frac{\alpha^2 x + \beta^2\mu}{\alpha^2+\beta^2}\right)^2 - \left(\frac{\alpha^2 x + \beta^2\mu}{\alpha^2+\beta^2}\right)^2 + \frac{\beta^2\mu^2 + \alpha^2 x^2}{\alpha^2+\beta^2}}{2\frac{\alpha^2\beta^2}{\alpha^2+\beta^2}}\right)} dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\alpha\beta} e^{-\left(\frac{\left(y - \frac{\alpha^2 x + \beta^2\mu}{\alpha^2+\beta^2}\right)^2 - \left(\frac{\alpha^2 x + \beta^2\mu}{\alpha^2+\beta^2}\right)^2 + \frac{\beta^2\mu^2 + \alpha^2 x^2}{\alpha^2+\beta^2}}{2\frac{\alpha^2\beta^2}{\alpha^2+\beta^2}}\right)} dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\alpha\beta} e^{-\left(\frac{\left(y - \frac{\alpha^2 x + \beta^2\mu}{\alpha^2+\beta^2}\right)^2}{2\frac{\alpha^2\beta^2}{\alpha^2+\beta^2}}\right)} e^{-\left(\frac{\frac{\beta^2\mu^2 + \alpha^2 x^2}{\alpha^2+\beta^2} - \left(\frac{\alpha^2 x + \beta^2\mu}{\alpha^2+\beta^2}\right)^2}{2\frac{\alpha^2\beta^2}{\alpha^2+\beta^2}}\right)} dy \\
&= \frac{1}{2\pi\alpha\beta}\sqrt{2\pi\frac{\alpha^2\beta^2}{\alpha^2+\beta^2}} e^{-\left(\frac{\frac{\beta^2\mu^2 + \alpha^2 x^2}{\alpha^2+\beta^2} - \left(\frac{\alpha^2 x + \beta^2\mu}{\alpha^2+\beta^2}\right)^2}{2\frac{\alpha^2\beta^2}{\alpha^2+\beta^2}}\right)} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\frac{\alpha^2\beta^2}{\alpha^2+\beta^2}}} e^{-\left(\frac{\left(y - \frac{\alpha^2 x + \beta^2\mu}{\alpha^2+\beta^2}\right)^2}{2\frac{\alpha^2\beta^2}{\alpha^2+\beta^2}}\right)} dy \\
&= \frac{1}{\sqrt{2\pi(\alpha^2+\beta^2)}} e^{-\left(\frac{\frac{\beta^2\mu^2 + \alpha^2 x^2}{\alpha^2+\beta^2} - \left(\frac{\alpha^2 x + \beta^2\mu}{\alpha^2+\beta^2}\right)^2}{2\frac{\alpha^2\beta^2}{\alpha^2+\beta^2}}\right)} \\
&= \frac{1}{\sqrt{2\pi(\alpha^2+\beta^2)}} e^{-\left(\frac{(\alpha^2+\beta^2)(\beta^2\mu^2+\alpha^2x^2) - (\alpha^2 x + \beta^2\mu)^2}{2\alpha^2\beta^2(\alpha^2+\beta^2)}\right)} \\
&= \frac{1}{\sqrt{2\pi(\alpha^2+\beta^2)}} e^{-\left(\frac{\alpha^2\beta^2\mu^2 + \alpha^4x^2 + \beta^4\mu^2 + \alpha^2\beta^2x^2 - \alpha^4x^2 - 2\alpha^2\beta^2\mu x - \beta^4\mu^2}{2\alpha^2\beta^2(\alpha^2+\beta^2)}\right)} \\
&= \frac{1}{\sqrt{2\pi(\alpha^2+\beta^2)}} e^{-\left(\frac{\alpha^2\beta^2x^2 - 2\alpha^2\beta^2\mu x + \alpha^2\beta^2\mu^2}{2\alpha^2\beta^2(\alpha^2+\beta^2)}\right)} \\
&= \frac{1}{\sqrt{2\pi(\alpha^2+\beta^2)}} e^{-\left(\frac{(x-\mu)^2}{2(\alpha^2+\beta^2)}\right)} \\
&= \mathcal{N}(\mu, \alpha^2+\beta^2)
\end{aligned}
$$

To find $p(y|x)$ we simply divide $p(x, y)$ by $p(x)$. In finding $p(x)$, we already know the form of $p(y|x)$ (see the longest line in the derivation of $p(x)$ above):

$$
p(y|x) = \frac{p(x, y)}{p(x)}
$$

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

$$= \mathcal{N}\left(\frac{\alpha^2 x+\beta^2\mu}{\alpha^2+\beta^2}, \frac{\alpha^2\beta^2}{\alpha^2+\beta^2}\right)$$

Note that all the above distibutions are Gaussian.

b. The following Matlab code produced Figure 1:

```
m = 0.0
a = 15.0
b = 3
x = 9

y = -100:1:100
mean = ((a^2)*x + (b^2)*m)/(a^2 + b^2)
var = ((a*b)^2)/(a^2 + b^2)
p_y_given_x = (1.0/sqrt(2*pi*var))*exp(-((y-mean).^2)/(2*var))
var2 = a^2 + b^2
p_y = (1.0/sqrt(2*pi*var2))*exp(-((y-m).^2)/(2*var2))

hold off
plot(y,p_y_given_x,'b')
hold on
plot(y,p_y,'r')
legend('p(y|x)', 'p(y)')
sy = size(y)
axis([y(1),y(sy(2)),0,0.2])
xlabel('y')
text(-70,0.14, '\mu=0')
text(-70,0.12, '\alpha=15')
text(-70,0.1, '\beta=3')
```

Problem 4: Covariance Matrix [15 points]

Let $\mathbf{\Sigma} = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$.

a. Find the eigenvalues and eigenvectors of $\mathbf{\Sigma}$ by hand (include all calculations.) Verify your computations with MATLAB function *eig*.

b. Verify that $\mathbf{\Sigma}$ is a valid covariance matrix.

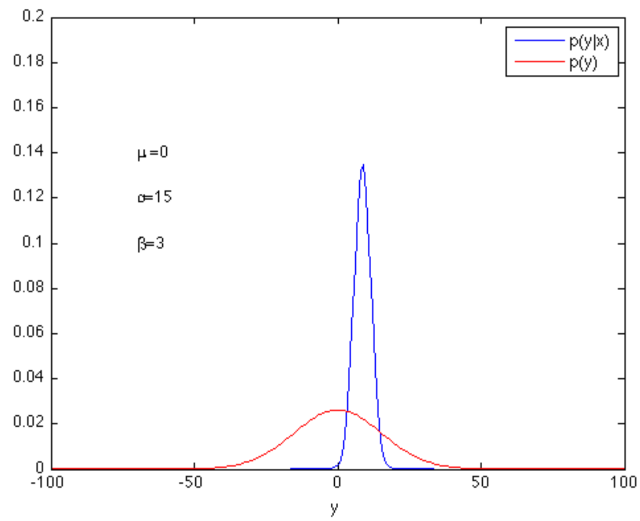

Figure 1: The marginal p.d.f. of y and the p.d.f. of y given x for a specific value of x. Notice how knowing x makes your knowledge of y more certain.

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

Solution:

a. We can find the eigenvectors and eigenvalues of $\mathbf{\Sigma}$ by starting with the definition of an eigenvector. Namely, a vector $\mathbf{e}$ is an eigenvector of $\mathbf{\Sigma}$ if it satisfies

$$\mathbf{\Sigma e} = \lambda \mathbf{e}$$

for some constant scalar λ, which is called the eigenvalue corresponding to $\mathbf{e}$. This can be rewritten as

$$(\mathbf{\Sigma} - \lambda I)\mathbf{e} = \mathbf{0}.$$

This is equivalent to

$$\det(\mathbf{\Sigma} - \lambda I) = 0.$$

Thus, we require that

$$(5-\lambda)^2 - 4^2 = 0$$

By inspection, this is true when $\lambda = 9$ and $\lambda = 1$.

To find the eigenvectors, we plug the eigenvalues back into the equation above to get

$$(\mathbf{\Sigma} - 9I)\mathbf{e} = \begin{bmatrix} 5-9 & 4 \\ 4 & 5-9 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -4 & 4 \\ 4 & -4 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

which gives $a = b$. Normalized, this results in the eigenvector

$$\mathbf{e}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

Similarly, $\lambda = 1$ gives

$$(\mathbf{\Sigma} - 1I)\mathbf{e} = \begin{bmatrix} 5-1 & 4 \\ 4 & 5-1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

which gives $a = -b$. Normalized, this results in the eigenvector

$$\mathbf{e}_1 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}.$$

b. The matrix $\mathbf{\Sigma}$ is a valid covariance matrix if it is symmetric and positive semi-definite. Clearly, it is symmetric, since $\mathbf{\Sigma}^T = \mathbf{\Sigma}$. One way to prove it is positive semi-definite is to show that all its eigenvalues are non-negative. This is indeed the case, as shown in the previous question.

c. The following Matlab program generated Figure **??**:

The second plot in Figure **??** shows the data rotated to align with the eigenvectors of the data's covariance matrix.

Problem 5: Probabilistic Modeling [20 points]

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 $\theta = (\theta_1, \theta_2)$.

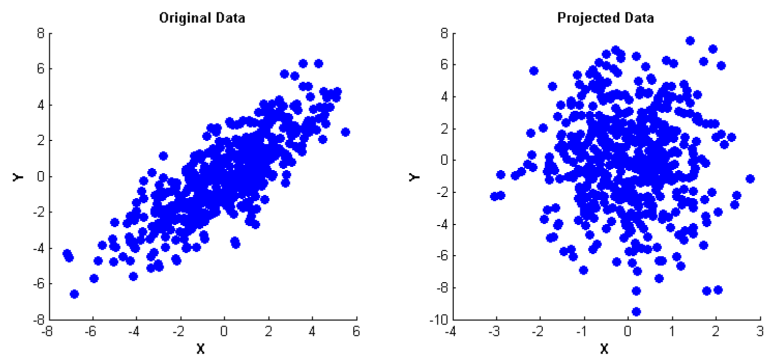

Figure 2: The original data and the data transformed into the coordinate system defined by the eigenvectors of their covariance matrix.

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
0	0
0	1
0	0
1	1

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 6: Ring Problem [20 points]

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? Justify your answer using Bayes' rule.
- 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.