Ravi Patel
A11850926

COGS 118A Spring 2016 - Assignment #2
Due: April 19, 2017, 11:59pm
Grade: ____ out of 100 points
Instructions:
Please answer the questions below and insert your answers to create a pdf file (you can have the hand-written version and scan it to create a pdf file); submit your file to TED (ted.ucsd.edu) by 11:59pm, 04/19/2016. You may search information online but you have to use your own words to answer the questions.
Late policy: 5% of the total points will be deducted for the first late day, then every 10% of the total points will be deducted for every extra day past due.

1 (10 points) Probability & Events

Given three events A, B, C with $P(A) = 1/2$, $P(B) = 1/3$, $P(C) = 1/5$, and $P(A, B, C) = 1/30$, can we determine whether A, B, C are all independent? Why?

A, B, C are independent iff $A \perp\!\!\!\perp B$, $B \perp\!\!\!\perp C$, & $C \perp\!\!\!\perp A$. However, $A \perp\!\!\!\perp B$ iff $Pr(A, B) = Pr(A) \cdot P(B)$ iff $Pr(A|B) = P(A)$ iff $Pr(B|A) = Pr(B)$.
From the provided information, it is not enough to determine if $A \perp\!\!\!\perp B$. Henceforth, it is inconclusive to determine whether A, B, C are all independent.

2 (20 points) Conditional Probability

Below describes some standard definition about a test:

1. True positive rate (correctly identified): $P(test + |sick+)$ Sick people correctly diagnosed as sick
2. False positive rate (incorrectly identified): $P(test + |sick-)$ Healthy people incorrectly identified as sick
3. True negative rate (correctly rejected): $P(test - |sick-)$ Healthy people correctly identified as healthy
4. False negative rate (incorrectly rejected): $P(test - |sick+)$ Sick people incorrectly identified as healthy

A mammogram tests for breast cancer. The true positive rate of a particular mammogram is 97% ($P(test + |cancer+)$). The true negative rate is also pretty good, at 93% ($P(test - |cancer-)$). If the frequency of breast cancer in the population is $P(cancer+) = 0.05\%$, and people are randomly selected to receive mammograms.

$Pr(test + | cancer -) = 1 - Pr(test - | cancer -) = 0.07$
$Pr(test - | cancer +) = 1 - Pr(test + | cancer +) = 0.03$

2.1

What is the probability of having breast cancer, given they got positive test results, $P(cancer+|test+)$?
Hint: you can simply use the conditional probability equation to derive this.

$$\Pr(cancer+|test+) = \frac{\Pr(test+|cancer+)\cdot\Pr(cancer+)}{\Pr(test+|cancer+)\cdot\Pr(cancer+)+\Pr(test+|cancer-)\cdot\Pr(cancer-)}$$

$$= \frac{(.97)(0.05)}{(.97)(0.05)+(0.07)(0.95)} = \frac{0.0485}{.115} = 0.422$$

2.2

What is the probability of not having breast cancer, given they got negative test results, $P(cancer-|test-)$?

$$\Pr(cancer-|test-) = \frac{\Pr(test-|cancer-)\cdot\Pr(cancer-)}{\Pr(test-|cancer-)\cdot\Pr(cancer-)+\Pr(test-|cancer+)\cdot\Pr(cancer+)}$$

$$= \frac{(.93)(.95)}{(.93)(.95)+(.03)(.05)} = \frac{0.8835}{0.885} = .998$$

2.3

Compute *precision*, *recall*, and $F-value = \frac{2\times precision\times recall}{precision+recall}$.

Precision $= \Pr(cancer+|test+) = 0.422$

Recall $= \Pr(test+|cancer+) = 0.97$

$$\Rightarrow F\text{-}value = \frac{2\cdot Precision\cdot recall}{Precision+recall} = \frac{2\cdot(0.422)(.97)}{0.422+0.97} = \frac{0.818}{1.392} = 0.588$$

3 (25 points) Communication Channel

In a communication channel shown in Figure below signal 3 is sent more frequently than 1, with $P(X=3) = 3\times P(X=1)$; $P(X=2) = 2\times P(X=1)$.
Hint: Use the conditional probability equation discussed in the class. Consider α, β, and γ as fixed value, and carry α, β, and γ in your answers, if needed.

$\Pr(x=1) = 1/6$ $\quad\Pr(x=3) = 3/6$ $\quad$ Let $\Pr(x=1) = \delta$

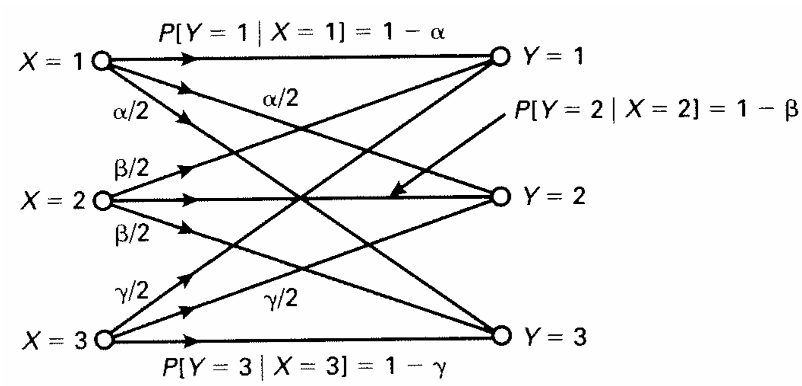
$\Pr(x=2) = 2/6$ $\quad\Rightarrow \delta = 1/6$

3.1

When $Y=1$ is observed; what is the conditional probability that $X=1$ was sent?

$$\Pr(x=1|y=1) = \frac{\Pr(y=1|x=1)\cdot\Pr(x=1)}{\Pr(y=1)} = \frac{\Pr(y=1|x=1)\cdot\Pr(x=1)}{\sum_{k=1}^{3}\Pr(y=1|x=k)\cdot\Pr(x=k)}$$

$$\Pr(x=1|y=1) = \frac{(1-\alpha)\cancel{\delta}}{(1-\alpha)\cancel{\delta} + \beta\cancel{\delta} + \frac{3}{2}\gamma\cancel{\delta}} = \frac{(1-\alpha)}{(1-\alpha)+\beta+\frac{3}{2}\gamma}$$



3.2

What is $P(X=1, X=1), P(X=2, X=1), P(X=3, X=1)$?

$\Pr(X=1, X=1) = \Pr(X=1) = \delta = 1/6$ $\quad \Pr(X=3, X=1) = 0.$

$\Pr(X=2, X=1) = 0$

3.3

What is $P(X=1|X=1), P(X=2|X=1), P(X=3|X=1)$?

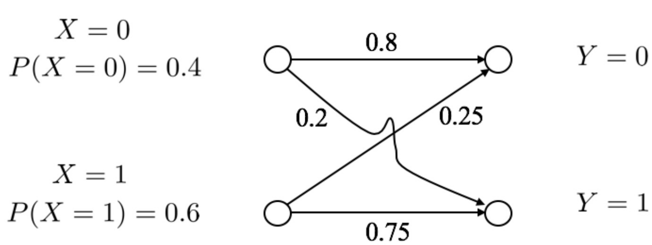
$\Pr(X=1|X=1) = 1$

$\Pr(X=2|X=1) = \frac{\Pr(X=2, X=1)}{\Pr(X=1)} = \frac{0}{\delta} = 0$

$\Pr(X=3|X=1) = \frac{\Pr(X=3, X=1)}{\Pr(X=1)} = \frac{0}{\delta} = 0$

4 (15 points) Binary Communication System

For the binary communication system shown in Figure below compute P(Y=0) and P(Y=1).



$$\Pr(Y=0) = \Pr(Y=0|X=0)\cdot\Pr(X=0) + \Pr(Y=0|X=1)\cdot\Pr(X=1)$$
$$= (.8)(.4) + (.25)(.6)$$
$$= .32 + .15 = .47$$

$$\Pr(Y=1) = \Pr(Y=1|X=1)\cdot\Pr(X=1) + \Pr(Y=1|X=0)\cdot\Pr(X=0)$$
$$= (0.75)(0.6) + (0.2)(0.4)$$
$$= .45 + .08 = .53$$

5 (15 points) Maximum Likelihood Estimation

By setting the derivatives of the log likelihood function equal to zero

$$\ln p(\boldsymbol{x}|\mu, \sigma^2) = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n - \mu)^2 - \frac{N}{2}\ln\sigma^2 - \frac{N}{2}\ln(2\pi)$$

with respect to μ and σ,

other terms can be dropped due to the property of argmax.

$$\mu_{ML} = \arg\min_{\mu} - \sum_{n=1}^{N} \ln p(x_n; \mu, \sigma) = \arg\max_{\mu} \sum_{n=1}^{N} \ln p(x_n; \mu, \sigma) = \sum_{n=1}^{N}(x_n - \mu)^2$$

$$\sigma_{ML} = \arg\min_{\sigma} - \sum_{n=1}^{N} \ln p(x_n; \mu_{ML}, \sigma) = \arg\max_{\sigma} \sum_{n=1}^{N} \ln p(x_n; \mu_{ML}, \sigma) = \frac{-1}{2\sigma^2}\sum_{n=1}^{N}(x_n - \mu_{ML})^2 - \frac{N}{2}\ln\sigma^2$$

verify the following results using the maximum likelihood estimation concept:

$$\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

$$\sigma_{ML} = [\frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{ML})^2]^{\frac{1}{2}}$$

$$\frac{d\mu_{ML}}{d\mu} = -2\sum_{n=1}^{N}(x_n - \mu) \Rightarrow \frac{d\mu_{ML}}{d\mu} = 0 \text{ when } -2\sum_{n=1}^{N}(x_n - \mu) = 0 \Leftrightarrow \sum_{n=1}^{N}(x_n - \mu) = 0 \Leftrightarrow \left(\sum_{n=1}^{N} x_n\right) - N\mu = 0$$

$$\Updownarrow$$

$$\mu = \frac{1}{N}\sum_{n=1}^{N} x_n$$

$$\therefore \mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

$$\frac{d\sigma_{ML}}{d\sigma} = \frac{\sum_{n=1}^{N}(x_n - \mu_{ML})^2}{\sigma^3} - \frac{N}{\sigma} \Rightarrow \frac{d\sigma_{ML}}{d\sigma} = 0 \text{ when } \frac{\sum_{n=1}^{N}(x_n - \mu_{ML})^2}{\sigma^3} - \frac{N}{\sigma} = 0 \Leftrightarrow \sum_{n=1}^{N}(x_n - \mu_{ML})^2 - \sigma^2 N = 0 \Leftrightarrow \sum_{n=1}^{N}(x_n - \mu_{ML})^2 = N\sigma^2$$

$$\Updownarrow$$

$$\sigma = \left[\frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{ML})^2\right]^{1/2}$$

$$\therefore \sigma_{ML} = \left[\frac{1}{N}\sum_{n=1}^{N}(x_n - \mu_{ML})^2\right]^{1/2}$$

6 (15 points) Least Squares Estimation

Please download the file **data.mat** from the course website, and load it into your environment (Python or MATLAB). Now do the following procedures, paste your source code and show the result in your report:

Python:

(0) Import Python packages and load data

```
import scipy.io as sio
import matplotlib.pyplot as plt
import numpy as np
```

```
data = sio.loadmat('data.mat')
x = data['x'].reshape([-1, 1])
y = data['y'].reshape([-1, 1])
```

(1) Plot the data using plot function.

```
plt.plot(x, y)
plt.grid()
```

(2) Create the matrix

```
X = np.hstack((np.ones((len(x),1)),np.power(x,1)))
```

(3) Compute the least squares line over the given data by:

```
X_t = X.transpose((1,0))
sol = np.dot(np.linalg.inv(np.dot(X_t,X)),np.dot(X_t,y))
```

(4) Overlay the computed least square line over the given data:

```
plt.hold(True)
plt.plot(x,sol[0]+sol[1]*x)
```

(5) Assign a title to this figure:

```
plt.title('Least square line fitting')
plt.xlabel('x')
plt.ylabel('y')
```

(6) Copy and paste the figure in your report.

MATLAB:

(1) Plot the data using plot function.

```
>> plot(x,y); grid;
```

(2) Create the matrix

```
>> A = [x.^0 x.^1]}
```

(3) Compute the least squares line over the given data by:

```
>> sol = inv(A'*A)*A'*y;
```

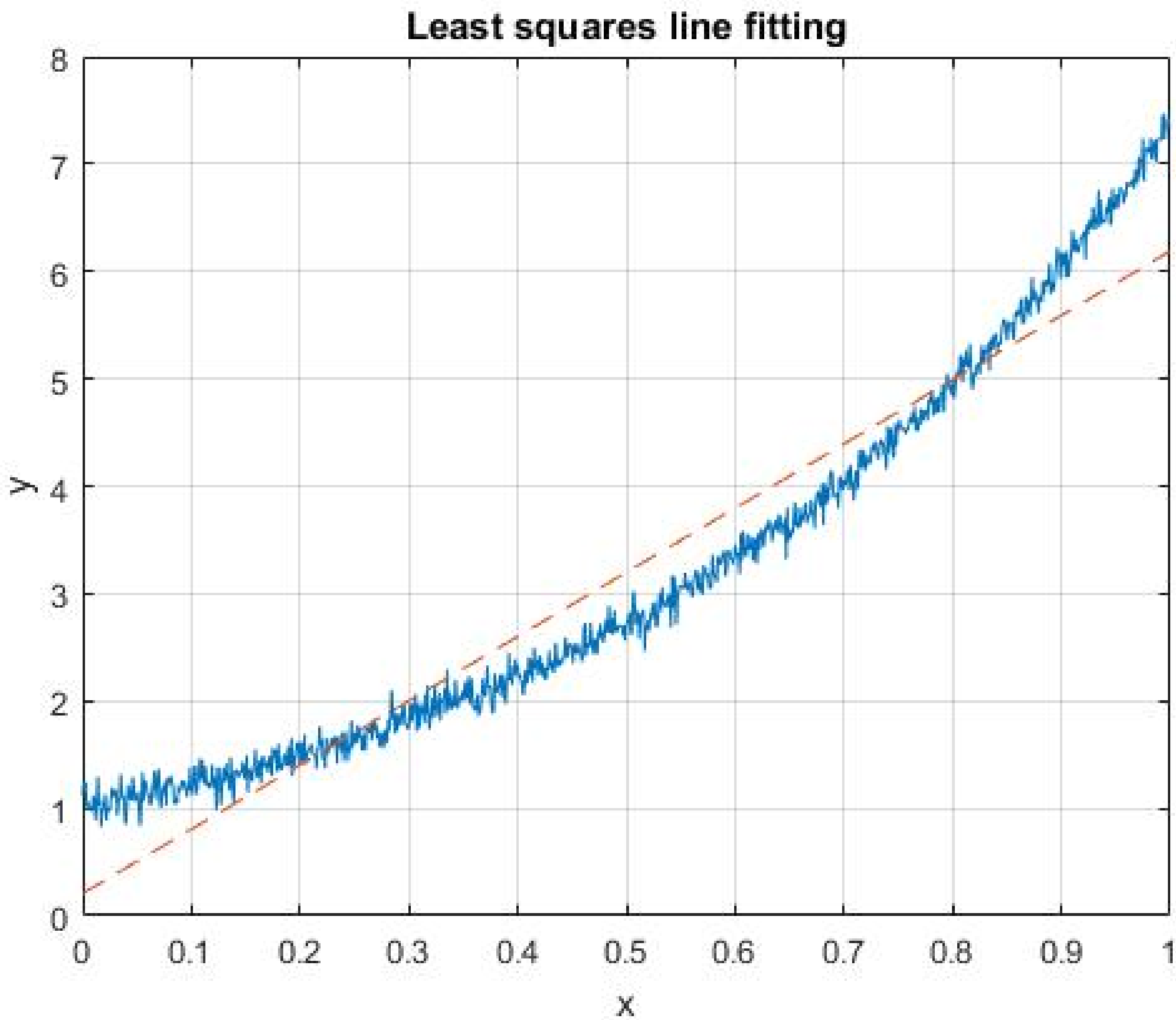
(4) Overlay the computed least square line over the given data:

```
>> hold on; plot(x, sol(1)+sol(2)*x, '--');
```

(5) Assign a title to this figure:

```
>> title('Least squares line fitting'); xlabel('x'); ylabel('y');
```

(6) Copy and paste the figure in your report.



7 (5 points) Bonus Question

Starting with the equation for a Gaussian distribution, $p(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

Let $c = \frac{1}{\sigma\sqrt{2\pi}}$

7.1

Show that the mean of $p(x; 0, \sigma)$ is indeed at zero using the equation: mean = E[X] $= \int_{-\infty}^{+\infty} xp(x)dx$.

Hint: change variables to $y = x^2/2$, $dy = xdx$.

$$E[X] = \int_{-\infty}^{+\infty} x P_r(x)\, dx = c\int_{-\infty}^{+\infty} x \exp\left\{\frac{-x^2}{2\sigma^2}\right\} dx =$$

let $y = \frac{x^2}{2\sigma^2} \Rightarrow \sigma^2 dy = x dx \Rightarrow c\int_{-\infty}^{+\infty} x \exp\left\{\frac{-x^2}{2\sigma^2}\right\} dx = \sigma c \int_{+\infty}^{+\infty} e^{-y} dy = 0$

$$\therefore \quad E[X] = \int_{-\infty}^{+\infty} x P_r(x) dx = 0$$

$\blacksquare$

7.2

Show that the most likely variable (with the highest probability density) is also μ for $p(x; \mu, \sigma)$.

Hint 1: The derivative of a function is zero at the peak.

Hint 2: The peak of the log of the normal distribution $\log p(x; \mu, \sigma)$ and that of the original normal distribution $p(x; \mu, \sigma)$ are at the same x value.

By MLE, the most likely variable is when $x^* = \underset{x}{\operatorname{argmax}}\, P_r(x; \mu, \sigma) = \underset{x}{\operatorname{argmax}} \left[\ln(P_r(x; \mu, \sigma))\right]$

$$x^* = \underset{x}{\operatorname{argmax}}\left[\ln\left(P_r(x;\mu,\sigma)\right)\right] = \underset{x}{\operatorname{argmax}}\left[\ln\left(c \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}\right)\right] = \underset{x}{\operatorname{argmax}}\left[\ln(c) - \frac{(x-\mu)^2}{2\sigma^2}\right] = \underset{x}{\operatorname{argmin}}\left[\frac{(x-\mu)^2}{2\sigma^2}\right]$$

Let $f(x) = \frac{(x-\mu)^2}{2\sigma^2} \Rightarrow x^* = \underset{x}{\operatorname{argmin}}\, f(x)$.

$\frac{df(x)}{dx} = \frac{x-\mu}{\sigma^2} \Rightarrow \frac{df(x)}{dx} = 0$ when $\frac{x-\mu}{\sigma^2} = 0 \iff x = \mu$.

$\therefore$ by MLE, $x^* = \mu$

$\therefore$ The most likely variable is also μ for $P_r(x; \mu, \sigma)$