CMPS 242: Machine Learning: Fall 2016: Homework 1 Solutions

Due: 11th October 2016

Question 1 (1 + 1 points): You move into a new neighborhood, and there are three other houses in your cul-de-sac. When talking to your new neighbors, lets call them Alice, Bob, and Cathy, you find out that Alice has three pets. Since you don't know if Alice is a cat or a dog lover, you can assume that the probability that Alice has a cat is $\frac{1}{2}$ and the probability that she has a dog is $\frac{1}{2}$. Alice reveals to you that one of her pets is a cat.

a) What is the probability that at least one of Alice's pets is a dog?

> Since we already know that one of Alice's pets is a cat, this means that one of her pets is "fixed" and hence we are only left with the cases of
>
> **Cat**-Dog-Dog, **Cat**-Cat-Dog, **Cat**-Dog-Cat, **Cat**-Cat-Cat
>
> By simply counting the cases, we see that there are 3 out of 4 cases in which Alice has a Dog, hence the probability is $P = \frac{3}{4}$.

After getting to know your neighbors more, you find out the following: Alice has two dogs and a cat, Bob has a cat, a dog, and a hamster, while Cathy has four dogs. However, since you are busy with school, you never got a chance to see any of the pets. During your evening walk, you see a dog without a leash, and you want to alert your neighbors.

b) Whose door should you knock first and why?

> We first need to find out the probability that an observed pet is a dog. For this we use the sum rule.

$$
\begin{aligned}
P(\text{Animal} = \text{Dog}) &= \sum_{\text{owners}} P(\text{Animal} = \text{Dog}|\text{owner}) \\
&= \frac{1}{3} \cdot \frac{2}{3} + \frac{1}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{4}{4} = \frac{2}{3}
\end{aligned}
$$

Now, we can use Bayes' rule to compute the posterior probability for the dog belonging to each owner

$$
\begin{aligned}
P(\text{owner}|\text{animal}) &= \frac{P(\text{animal}|\text{owner})P(\text{owner})}{P(\text{animal})} \\
P(\text{Alice}|\text{dog}) &= \frac{P(\text{dog}|\text{Alice})P(\text{Alice})}{P(\text{dog})} = \frac{1}{3} \\
P(\text{Bob}|\text{dog}) &= \frac{P(\text{dog}|\text{Bob})P(\text{Bob})}{P(\text{dog})} = \frac{1}{6} \\
P(\text{Cathy}|\text{dog}) &= \frac{P(\text{dog}|\text{Cathy})P(\text{Cathy})}{P(\text{dog})} = \frac{1}{2}
\end{aligned}
$$

You should knock on Cathy's door because the dog is most likely hers.

Question 2 (1 + 1 + 1 + 2 points): Recall the 1-d Gaussian distribution that we studied in the class.

a) Prove that the Gaussian distribution is well normalized. In other words show that $\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu, \sigma^2\right) dx = 1$

In other words, we have to show that

$$
\int_{-\infty}^{\infty} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}}}_{\text{normalization ct.}} \underbrace{\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)}_{\text{we need to massage this}} dx = 1
$$

Define

$$
\theta = \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right) dx
$$

Let's do a first change of variables to $y = x - \mu$ which has $dy = dx$ and simplifies theta to

$$
\theta = \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy
$$

We then obtain θ^2 as

$$
\begin{aligned}
\theta^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2\sigma^2}\right)\exp\left(-\frac{z^2}{2\sigma^2}\right) dydz \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(-\frac{y^2+z^2}{2\sigma^2}\right) dydz
\end{aligned}
$$

The next step is to do a further change of variables to polar coordinates $x = r\cos\phi$ and $y = r\sin\phi$. In order to do the change of variables correctly, we must compute the Jacobian which is given by

$$
J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \phi} \end{vmatrix} = \begin{vmatrix} \cos\phi & -r\sin\phi \\ \sin\phi & r\cos\phi \end{vmatrix} = r
$$

We plug this into our θ^2 and get that

$$
\begin{aligned}
\theta^2 &= \int_0^{\infty}\int_0^{2\pi} \exp\left(-\frac{r^2}{2\sigma^2}\right) r d\theta dr \\
&= 2\pi \int_0^{\infty} \exp\left(-\frac{r^2}{2\sigma^2}\right) r dr
\end{aligned}
$$

One more change of variables for $r^2 = u$ so that $dr = \frac{1}{2r}\, du$.

$$
\begin{aligned}
\theta^2 &= 2\pi \int_0^{\infty} \exp\left(-\frac{u}{2\sigma^2}\right) \frac{1}{2} du \\
&= \pi(-2\sigma^2)\exp\left(-\frac{u}{2\sigma^2}\right)\Bigg|_0^{\infty} \\
&= 2\pi\sigma^2
\end{aligned}
$$

It then follows that $\theta = \sqrt{2\pi\sigma^2}$. Plugging this back into our integral shows that

$$
\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) dx = \frac{\theta}{\sqrt{2\pi\sigma^2}} = \frac{\sqrt{2\pi\sigma^2}}{\sqrt{2\pi\sigma^2}} = 1
$$

b) Formally show that for the Gaussian distribution

$$\mathbb{E}\left[x\right] = \mu$$
$$\operatorname{var}\left[x\right] = \sigma^2$$

This integration is very similar to the one we had above. We start from the definition of expectation

$$\begin{aligned}
\mathbb{E}[x] &= \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) x dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left(x-\mu\right)^2\right) x dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} y^2\right) (y+\mu) dy \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} y^2\right) y dy \\
&\quad + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} y^2\right) \mu dy
\end{aligned}$$

The first term disappears because

$$\begin{aligned}
&\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} y^2\right) y dy = \\
&\quad = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} y^2\right) y dy \\
&\qquad + \int_{0}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} y^2\right) y dy \\
&\quad = -\int_{0}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} y^2\right) y dy \\
&\qquad + \int_{0}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} y^2\right) y dy = 0
\end{aligned}$$

If you're confused as to why we could do this, check the properties of odd functions under integration. The second term is just $\mu \int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) dx = \mu$. This then shows that

$$\mathbb{E}[x] = \mu$$

For the variance, we cheat a bit by massaging the normalized Gaussian that we proved in part a), and then take the following derivative

$$
\begin{aligned}
\frac{\partial}{\partial \sigma^2} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}\left(x-\mu\right)^2\right) dx &= \frac{\partial}{\partial \sigma^2}\sqrt{2\pi\sigma^2} \\
\frac{1}{2(\sigma^2)^2} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}\left(x-\mu\right)^2\right) \left(x-\mu\right)^2 dx &= \frac{2\pi}{2\sqrt{2\pi\sigma^2}} \\
\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma^2}\left(x-\mu\right)^2\right) \left(x-\mu\right)^2 dx &= \sigma^2 \\
\mathbb{E}[(x-\mu)^2] &= \sigma^2 \\
\mathbb{E}[x^2] - 2\mu\mathbb{E}[x] + \mathbb{E}[x]^2 &= \sigma^2 \\
\mathbb{E}[x^2] - 2\mu^2 + \mu^2 &= \sigma^2 \\
\mathbb{E}[x^2] - \mathbb{E}[x]^2 &= \sigma^2 \\
var[x] &= \sigma^2
\end{aligned}
$$

c) Show that the maximum likelihood (ML) estimate for the Gaussian is given by

$$
\begin{aligned}
\mu_{\mathrm{ML}} &= \frac{1}{N}\sum_{n=1}^{N} x_n \\
\sigma^2_{\mathrm{ML}} &= \frac{1}{N}\sum_{n=1}^{N} \left(x_n - \mu_{ML}\right)^2
\end{aligned}
$$

The log likelihood, $\mathcal{L}$, is given by

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

We first find the MLE value for μ. Thus

$$\frac{\partial \mathcal{L}}{\partial \mu} = -\frac{1}{2\sigma^2}\sum_{n=1}^{N}(-2)\left(x_n-\mu\right) = 0$$

Taking out the constant multipliers, this means that

$$\sum_{n=1}^{N}\left(x_n - \mu_{ML}\right) = 0 \Leftrightarrow \sum_{n=1}^{N} x_n = N\mu_{ML} \Leftrightarrow \mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n$$

Applying the same procedure for the variance, leads to

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \sigma^2} &= \frac{1}{2(\sigma^2)^2}\sum_{n=1}^{N}(x_n-\mu)^2 - \frac{N}{2}\frac{\sigma^2}{(\sigma^2)^2} \\ &= \frac{1}{2(\sigma^2)^2}\left(\sum_{n=1}^{N}(x_n-\mu)^2 - N\sigma^2\right) = 0\end{aligned}$$

It then directly follows that

$$\sigma^2_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{ML})^2$$

d) Show that the ML estimate of the mean is unbiased but the variance is biased:

$$\begin{aligned}\mathbb{E}\left[\mu_{\mathrm{ML}}\right] &= \mu \\ \mathbb{E}\left[\sigma^2_{\mathrm{ML}}\right] &= \left(\frac{N-1}{N}\right)\sigma^2\end{aligned}$$

In order to find the expectation of μ_{ML} we plug in the above formula i.e.

$$\mathbb{E}[\mu_{ML}] = \frac{1}{N}\sum_{n}^{N}\mathbb{E}[x_n] = \frac{1}{N}\sum_{n}^{N}\mu = \mu$$

Same approach for the variance, namely

$$\begin{aligned}\mathbb{E}[\sigma^2_{ML}] &= \frac{1}{N}\mathbb{E}\left[\sum_{n=1}^{N}(x_n-\mu_{ML})^2\right] \\ &= \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\left[x_n^2 - 2x_n\mu_{ML} + \mu_{ML}^2\right] \\ &= \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\left[x_n^2 - 2x_n\frac{1}{N}\sum_{m}^{N}x_m + \frac{1}{N}\sum_{m}^{N}x_m\frac{1}{N}\sum_{o}^{N}x_o\right]\end{aligned}$$

We assume that since we know that $E[x_n^2] = \mu^2 + \sigma^2$ and that for $m \neq n$, $E[x_n x_m] = E[x_n]E[x_m] = \mu^2$, it follows that the above is equal to

$$\begin{aligned}
\mathbb{E}[\sigma^2_{ML}] &= \frac{1}{N}\sum_{n=1}^{N}\left[\mu^2+\sigma^2-2(\mu^2+\frac{1}{N}\sigma^2)+(\mu^2+\frac{1}{N}\sigma^2)\right]\\
&= \mu^2+\sigma^2-2(\mu^2+\frac{1}{N}\sigma^2)+(\mu^2+\frac{1}{N}\sigma^2)\\
&= \frac{N-1}{N}\sigma^2
\end{aligned}$$

Question 3 (1 + 1 + 1 points): a) Show that if X and Y are independent random variables, then

$$p(X,Y) = p(X)\cdot p(Y)$$
$$\mathbb{E}[XY] = \mathbb{E}[X]\cdot\mathbb{E}[Y]$$

Using the product rule, $p(X,Y) = p(X|Y)\cdot p(Y)$.
Since X and Y are independent, $p(X|Y) = p(X)$.
Therefore, $p(X,Y) = p(X)\cdot p(Y)$.

By definition of the expected value of random variables (assuming the discrete case, continuous is similar),

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_{X,Y} XY\cdot p(X,Y)\\
&= \sum_{X,Y} XY\cdot p(X)\cdot p(Y)\\
&= \sum_{X} X\ p(X)\cdot\sum_{Y} Y\ p(Y)\\
&= \mathbb{E}[X]\cdot\mathbb{E}[Y]
\end{aligned}$$

b) Write some code to simulate a large number (say 10,000) of coin tosses with two independent coins, and empirically verify the above results. Write a short (5 - 10 sentence) summary of your experiment and what you observed.

```
# case of independent tosses
import random
import numpy as np
```

```python
N = 10000
tries = 0
both_heads = 0
tail_tail = 0
coin1_heads = 0
coin2_heads = 0
while tries < N:
    tries += 1
    coin1 = random.randint(0, 1)
    coin2 = random.randint(0, 1)
    if coin1 == 0:
        coin1_heads += 1
        #print('H')
    if coin2 == 0:
        coin2_heads += 1
        #print('H')
    if coin1 == 0 and coin2 == 0:
        both_heads += 1
    if coin1 == 1 and coin2 == 1:
        tail_tail += 1

pXY = both_heads/float(tries)
pX = coin1_heads/float(tries)
pY = coin2_heads/float(tries)

# to verify p(X,Y) = P(X) . P(Y)
print('P(coin1=H, coin2=H)     = %f' % (pXY))
print('P(coin1=H)              = %f' % (pX))
print('P(coin2=H)              = %f' % (pY))
print('P(coin1=H) . P(coin2=H) = %f' % (pX * pY))
print('')

# to verify E(XY) = E(X). E(Y)
print('E[coin1] . E[coin2]     = %f' % ((1-pX)*(1-pY)))
print('E[coin1 . coin2]        = %f' % ...
    (tail_tail/float(tries)))
```

```
>> P(coin1=H, coin2=H)     = 0.257100
>> P(coin1=H)              = 0.500900
>> P(coin2=H)              = 0.506000
>> P(coin1=H) . P(coin2=H) = 0.253455

>> E[coin1] . E[coin2]     = 0.246555
>> E[coin1 . coin2]        = 0.250200
```

Note on how above expectation was computed:

```
E(coin1)    = (0 * p(coin1 = H)) + (1 * p(coin1 = T)) = (1 - pX)
E(coin2)    = (0 * p(coin2 = H)) + (1 * p(coin2 = T)) = (1 - pY)
E(coin1 . coin2)  =    [0 * 0 * p(coin1=H, coin2=H)]
                     + [1 * 1 * p(coin1=T, coin2=T)]
                     + [0 * 1 * p(coin1=H, coin2=T)]
                     + [1 * 0 * p(coin1=T, coin2=H)]
```

```
= [1 * 1 * p(coin1=T, coin2=T)]
= tail_tail/float(tries)
```

c) Now let your coin tosses depend on each other (e.g., the value that coin 2 takes depends on the value of coin 1) and empirically verify that the above results do not hold. Write a short (5 - 10 sentence) summary of your experiment and what you observed.

```python
# case of dependent tosses
# making coin2 take the same value as coin1
import random
import numpy as np
N = 10000
tries = 0
both_heads = 0
tail_tail = 0
coin1_heads = 0
coin2_heads = 0
while tries < N:
    tries += 1
    coin1 = random.randint(0, 1)
    #coin2 takes the same value as coin1
    if coin1 == 0:
        coin1_heads += 1
        coin2_heads += 1
        both_heads += 1
    if coin1 == 1:
        tail_tail += 1
pXY = both_heads/float(tries)
pX = coin1_heads/float(tries)
pY = coin2_heads/float(tries)

print('P(coin1=H, coin2=H)     = %f' % (pXY))
print('P(coin1=H)              = %f' % (pX))
print('P(coin2=H)              = %f' % (pY))
print('P(coin1=H) . P(coin2=H) = %f' % (pX * pY))
print('')

print('E[coin1] . E[coin2]     = %f' % ((1-pX)*(1-pY)))
print('E[coin1 . coin2]        = %f' % ...
    (tail_tail/float(tries)))
```

```
>> P(coin1=H, coin2=H)     = 0.500000
>> P(coin1=H)              = 0.500000
>> P(coin2=H)              = 0.500000
>> P(coin1=H) . P(coin2=H) = 0.250000

>> E[coin1] . E[coin2]     = 0.250000
>> E[coin1 . coin2]        = 0.500000
```