PURDUE
UNIVERSITY

Homework 7

Fall 2020
(Due: Nov 20, 2020, Friday)

Name: Elias Talcott Email: etalcott@purdue.edu

Homework is due at 11:59pm (midnight) Eastern Time. Please print this homework, write your solution, and scan the solution. Submit your homework through Gradescope. No late homework will be accepted.

Exercise 1.
Use Law of Total Expectation to compute the followings:

(a) $\mathbb{E}[\cos(X+Y)]$, where $X \sim \mathcal{N}(0,1)$, and $Y \mid X \sim \text{Uniform}[x-\pi, x+\pi]$

(b) $\mathbb{E}[Xe^Y]$, where $X \sim \text{Uniform}[3,7]$, and $Y \mid X \sim \mathcal{N}(0,x^2)$

$$\mathbb{E}[\cos(X+Y)] = \iint_{\mathbb{R}^2} \cos(x+y) f_{XY}(x,y)\,dx\,dy = \iint_{\mathbb{R}^2} \cos(x+y) f_{Y|X}(y|x) f_X(x)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} \mathbb{E}[\cos(x+y) \mid X=x] f_X(x)\,dx = \int_{-\infty}^{\infty}\left(\cancel{\int_{x-\pi}^{x+\pi} \cos(x+y) f_{Y|X}(y|x)\,dy}^{\,0}\right) f_X(x)\,dx$$

$$\Rightarrow \mathbb{E}[\cos(x+y)] = 0$$

b) $$\mathbb{E}[Xe^Y] = \int_{-\infty}^{\infty} \mathbb{E}[Xe^Y \mid X=x] f_X(x)\,dx$$

$$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} xe^y f_{Y|X}(y|x)\,dy\right) f_X(x)\,dx = \int_{-\infty}^{\infty} x\left(\int_{-\infty}^{\infty} e^y \frac{1}{\sqrt{2\pi x^2}} e^{\frac{-(y-0)^2}{2x^2}}\,dy\right) f_X(x)\,dx$$

$$= \int_{-\infty}^{\infty} x\left(\underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi x^2}} e^{\frac{-(y-x^2)^2}{2x^2}}\,dy}_{N(x^2,x^2)\Rightarrow 1}\, e^{\frac{x^2}{2}}\right) f_X(x)\,dx = \int_3^7 xe^{\frac{x^2}{2}} f_X(x)\,dx = \frac{1}{4}\int_3^7 xe^{\frac{x^2}{2}}\,dx$$

$u = \frac{x^2}{2}$

$\frac{du}{dx} = x$, $du = x\,dx$ $\rightarrow \frac{1}{4}\int_{9/2}^{49/2} e^u\,du = \frac{1}{4}\left[e^u\right]_{9/2}^{49/2} \rightarrow \mathbb{E}[Xe^Y] = \frac{e^{\frac{49}{2}} - e^{\frac{9}{2}}}{4}$

Exercise 2.
Let X and Y be two independent random variables with densities

$$f_X(x) = \begin{cases} xe^{-x}, & x \geq 0, \\ 0, & x < 0, \end{cases} \quad \text{and} \quad f_Y(y) = \begin{cases} ye^{-y}, & y \geq 0, \\ 0, & y < 0. \end{cases}$$

Find the PDF of $Z = X + Y$.

$$F_Z(z) = \mathbb{P}[Z \leq z] = \mathbb{P}[X+Y \leq z]$$

$$= \int_0^{\infty} \int_0^{z-y} f_X(x) f_Y(y)\, dx\, dy$$

$$f_Z(z) = \frac{d}{dz}(F_Z(z)) = \int_0^{\infty} f_X(z-x) f(y)\, d_y$$

$$= \int_0^{z} (z-y) e^{-(z-y)} y e^{-y}\, dy = \int_0^{z} zye^{-z} - y^2 e^{-z}\, dy$$

$$= ze^{-z} \int_0^{z} y\, dy - e^{-z} \int_0^{z} y^2\, dy = ze^{-z}\left[\frac{1}{2}y^2\right]_0^z - e^{-z}\left[\frac{1}{3}y^3\right]_0^z$$

$$= ze^{-z}\left(\frac{1}{2}z^2\right) - e^{-z}\left(\frac{1}{3}z^3\right) = \frac{1}{2}z^3 e^{-z} - \frac{1}{3}z^3 e^{-z}$$

$$\rightarrow f_Z(z) = \frac{1}{6} z^3 e^{-z} \cdot \mathbb{1}_{[0,\infty)}(z)$$

Exercise 3. (INTERMEDIATE, 15 POINTS)
Two independent random variables X and Y have PDFs

$$f_X(x) = \begin{cases} e^{-x}, & x \geq 0, \\ 0, & x < 0, \end{cases} \qquad f_Y(y) = \begin{cases} 0, & y > 0, \\ e^{y}, & y \leq 0. \end{cases}$$

Find the PDF of $Z = X - Y$.

$$F_Z(z) = \mathbb{P}[Z \leq z] = \mathbb{P}[X - Y \leq z]$$

$$= \int_{-\infty}^{0} \int_{0}^{z+y} f_X(x) f_Y(y)\, dx\, dy$$

$$f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{-\infty}^{0} f_X(y+z) f_Y(y)\, dy$$

$$= \int_{-z}^{0} e^{-(y+z)} e^{y}\, dy = e^{-z} \int_{-z}^{0} e^{(-y+y)}\, dy = e^{-z} \int_{-z}^{0} dy$$

$$= e^{-z} (0 - -z)$$

$$\rightarrow f_Z(z) = z e^{-z} \cdot \mathbb{1}_{[0,\infty)}(z)$$

Exercise 4.
Let X, Y, Z be three independent random variables

$$X \sim \text{Bernoulli}(p), \qquad Y \sim \text{Exponential}(\alpha), \qquad Z \sim \text{Poisson}(\lambda)$$

Find the moment generating function for the following random variables.

(a) $U = 4Y + Z$

(b) $U = 2Z - 1$

(c) $U = 3XY$

(d) $U = XY + 3(1 - X)Z$

a) $M_U(s) = \mathbb{E}\left[e^{sU}\right] = \mathbb{E}\left[e^{s(4Y+Z)}\right] = \mathbb{E}\left[e^{4sY}\right]\mathbb{E}\left[e^{sZ}\right] = M_Y(4s)\,M_Z(s)$

$M_U(s) = \frac{\alpha}{\alpha - 4s}\left(e^{\lambda(e^{s}-1)}\right)$

b) $M_U(s) = \mathbb{E}\left[e^{2sZ}\right](e^{-s}) \rightarrow M_U(s) = e^{-s}\left(e^{\lambda(e^{2s}-1)}\right)$

c) $M_U(s) = \mathbb{E}\left[e^{3sXY}\right] = \sum_X \mathbb{E}\left[e^{3sXY} \mid X = x\right]p_X(x)$

$= \mathbb{E}\left[e^{0}\right](1-p) + \mathbb{E}\left[e^{3sY}\right]p$

$M_U(s) = 1 - p + \frac{\alpha p}{\alpha - 3s}$

d) $M_U(s) = \mathbb{E}\left[e^{s(XY + 3(1-X)Z)}\right]$

$= \mathbb{E}\left[e^{3sZ}\right](1-p) + \mathbb{E}\left[e^{sY}\right]p$

$M_U(s) = (1-p)\left(e^{\lambda(e^{3s}-1)}\right) + \frac{\alpha p}{\alpha - s}$

Exercise 5.
Let $X_0, X_1, \dots$ be a sequence of independent random variables with PDF

$$f_{X_k}(x) = \frac{a_k}{\pi(a_k^2 + x^2)}, \qquad a_k = \frac{1}{3^{k+1}},$$

for $k = 0, 1, \dots$. Let

$$Y = \sum_{k=0}^{\infty} X_k.$$

Find the PDF of Y.

$$Y = \frac{1}{3} + \frac{1}{9} + \frac{1}{27} + \cdots \frac{1}{3^{n+1}}$$

$$= \frac{a_1(1-r^n)}{1-r} = \frac{\frac{1}{3}(1-\cancel{(\frac{1}{3})^n})}{1-\frac{1}{3}} = \frac{\frac{1}{3}}{\frac{2}{3}} \rightarrow Y = \frac{1}{2}$$

$$\Phi_Y(j\omega) = e^{-|\omega|Y} = e^{-\frac{|\omega|}{2}} \xleftrightarrow{F} f_Y(y)$$

$$a = \frac{1}{2} \rightarrow f_Y(y) = \frac{a^2}{(a^2+y^2)a\pi}$$

$$\rightarrow f_Y(y) = \frac{\frac{1}{4}}{\frac{1}{2}\pi(\frac{1}{4}+y^2)}$$

$$\rightarrow f_Y(y) = \frac{\frac{1}{2}}{\pi(\frac{1}{4}+y^2)}$$

Exercise 6.

This is part 2 of the programming exercise we started last homework. In `data.zip`, you will also find `training_cat.txt` and `training_grass.txt` which is the training data for this project. Load the data using `train_cat = np.matrix(np.loadtxt('train_cat.txt', delimiter = ','))`. The data need to be stored as matrix for later use.

The sizes of the arrays are $64 \times K$, where K corresponds to the number of training samples and 64 corresponds to the size of the block 8×8. In this project, you will need to compute the mean vectors and the covariance matrices of these data arrays. What are mean vectors and covariance matrices? For any collection of vectors, e.g., $\{\boldsymbol{x}_1^{(\text{cat})}, \ldots, \boldsymbol{x}_K^{(\text{cat})}\}$, the mean vector and the covariance matrix are

$$\boldsymbol{\mu}_{(\text{cat})} = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{x}_k^{(\text{cat})}, \quad \text{and} \quad \boldsymbol{\Sigma}_{(\text{cat})} = \frac{1}{K}\sum_{k=1}^{K} (\boldsymbol{x}_k^{(\text{cat})} - \boldsymbol{\mu}_{(\text{cat})})(\boldsymbol{x}_k^{(\text{cat})} - \boldsymbol{\mu}_{(\text{cat})})^T. \tag{1}$$

Compute these in MATLAB / Python. The size of $\boldsymbol{\mu}_{(\text{cat})}$ should be 64×1 and the size of $\boldsymbol{\Sigma}_{(\text{cat})}$ should be 64×64.

Let us discuss how to perform the classification using the training data. Our data is a collection of 8×8 patches (or equivalently 64×1 vectors). Therefore, we have to use a multivariate Gaussian instead of the single variable Gaussian. Multi-variate Gaussian is nothing but some generalization of the single variable Gaussian. If you like to know more about the multi-variate Gaussian, you can read textbook chapter 5.

Here are more discussions about a multivariate Gaussian. Let $\boldsymbol{Z}$ be a vector random variable representing a patch taken from the image "Cat and Grass". We assume that the conditional distribution of $\boldsymbol{Z}$ given its class label follows the distribution:

$$f_{\boldsymbol{Z}|\text{class}}(\boldsymbol{z}\,|\,\text{cat}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}_{(\text{cat})}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\mu}_{(\text{cat})})^T \boldsymbol{\Sigma}_{(\text{cat})}^{-1} (\boldsymbol{z} - \boldsymbol{\mu}_{(\text{cat})})\right\}, \tag{2}$$

$$f_{\boldsymbol{Z}|\text{class}}(\boldsymbol{z}\,|\,\text{grass}) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}_{(\text{grass})}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\mu}_{(\text{grass})})^T \boldsymbol{\Sigma}_{(\text{grass})}^{-1} (\boldsymbol{z} - \boldsymbol{\mu}_{(\text{grass})})\right\}. \tag{3}$$

These equations may look complicated to you in a first glance, but they are just the high-dimensional extension of the single variable Gaussian we learned in class — if $\boldsymbol{\mu}$ is a scalar μ, and $\boldsymbol{\Sigma}$ is also a scalar σ^2, then we will obtain the one-dimensional Gaussian.

Given a patch $\boldsymbol{Z}$, how would you decide whether it belongs to "cat" or "grass"? From the discussion above, it is tempting to compare $f_{\boldsymbol{Z}|\text{class}}(\boldsymbol{z}\,|\,\text{cat}) > f_{\boldsymbol{Z}|\text{class}}(\boldsymbol{z}\,|\,\text{grass})$. However, comparing these conditional probabilities is not what we want: it is the probability that we see a patch $\boldsymbol{Z}$ given the class label, not the class label given the patch $\boldsymbol{Z}$. What we should really compare is the posterior probability

$$f_{\text{class}|\boldsymbol{Z}}(\text{cat}\,|\,\boldsymbol{z}) \quad \overset{\text{cat}}{\underset{\text{grass}}{\gtrless}} \quad f_{\text{class}|\boldsymbol{Z}}(\text{grass}\,|\,\boldsymbol{z}). \tag{4}$$

The posterior probabilities are related to the conditional probabilities via the Bayes rule as

$$f_{\text{class}|\boldsymbol{Z}}(\text{cat}\,|\,\boldsymbol{z}) = \frac{f_{\boldsymbol{Z}|\text{class}}(\boldsymbol{z}\,|\,\text{cat}) f_{\text{class}}(\text{cat})}{f_{\boldsymbol{Z}}(\boldsymbol{z})} \tag{5}$$

$$f_{\text{class}|\boldsymbol{Z}}(\text{grass}\,|\,\boldsymbol{z}) = \frac{f_{\boldsymbol{Z}|\text{class}}(\boldsymbol{z}\,|\,\text{grass}) f_{\text{class}}(\text{grass})}{f_{\boldsymbol{Z}}(\boldsymbol{z})}. \tag{6}$$

The distributions $f_{\text{class}}(\text{cat})$ and $f_{\text{class}}(\text{grass})$ are called the prior distributions. Since we only have two classes, the prior distribution is a discrete probability mass function. For simplicity we assume that $f_{\text{class}}(\text{cat})$ and $f_{\text{class}}(\text{grass})$ are the ratios between the number of samples in the training set:

$$f_{\text{class}}(\text{cat}) = \frac{K_{(\text{cat})}}{K_{(\text{cat})} + K_{(\text{grass})}}, \tag{7}$$

$$f_{\text{class}}(\text{grass}) = \frac{K_{(\text{grass})}}{K_{(\text{cat})} + K_{(\text{grass})}}, \tag{8}$$

where $K_{(\text{cat})}$ is the number of training samples in train_cat, and $K_{(\text{grass})}$ is the number of training samples in train_grass. Substituting these equations into (4) yields

$$f_{\boldsymbol{Z}|\text{class}}(\boldsymbol{z} \mid \text{cat}) f_{\text{class}}(\text{cat}) \underset{\text{grass}}{\overset{\text{cat}}{\gtrless}} f_{\boldsymbol{Z}|\text{class}}(\boldsymbol{z} \mid \text{grass}) f_{\text{class}}(\text{grass}), \tag{9}$$

where we canceled out the common factor $f_{\boldsymbol{Z}}(\boldsymbol{z})$. The decision rule based on (9) is called the Maximum-a-Posteriori (MAP) decision.

(a) Compute mu_cat, mu_grass, Sigma_cat, Sigma_grass, K_cat, K_grass from the training data. Nothing to submit here.

(b) Implement the decision making for the image we tested in Homework 6. The main routine of your program should look like the following (this is a Python version):

```
Output = np.zeros((M-8,N-8))
for i in range(M-8):
    for j in range(N-8):
        z = Y(i:i+8, j:j+8)
        % compute f(z | cat)
        % compute f(z | grass)
        % if f(cat | z) > f(grass | z), then set Output(i,j) = 1
```

Your task is to fill in these missing blanks. The output of your decision should be a binary mask, indicating which pixel is foreground and which pixel is background. Submit your code (MATLAB or Python) and your figure.