Homework 1

Solutions

1. (a) We start with proving the relation

$$P[X=x|Y=y]=\frac{P[X=x]f_{Y|X}(y|x)}{f_Y(y)}.$$

Since Y is continuous, we must have $P[Y=y]=0$ for any possible y. The standard conditional probability formula $P[X=x|Y=y]=\frac{P[X=x,Y=y]}{P[Y=y]}$ does not work here because the denominator is 0. To avoid this difficulty, we find

$$\begin{aligned}P[X=x|Y=y] &= \lim_{\Delta y\to 0} P[X=x|y\le Y\le y+\Delta y]\\ &= \lim_{\Delta y\to 0}\frac{P[y\le Y\le y+\Delta y|X=x]P[X=x]}{P[y\le Y\le y+\Delta y]}\\ &\approx \lim_{\Delta y\to 0}\frac{f_{Y|X}(y|x)\Delta y P[X=x]}{f_Y(y)\Delta y}\\ &= \frac{f_{Y|X}(y|x)P[X=x]}{f_Y(y)}, \qquad (1)\end{aligned}$$

where the approximation is valid when Δy is very small. It follows that

$$P[X=x|Y=y]f_Y(y)=P[X=x]f_{Y|X}(y|x),$$

which further yields

$$\begin{aligned}\int P[X=x|Y=y]f_Y(y)dy &= P[X=x]\underbrace{\int_y f_{Y|X}(y|x)dy}_{=1}\\ &= P[X=x]. \qquad (2)\end{aligned}$$

Plugging (2) into (1) and rearranging the equation, we obtain

$$f_{Y|X}(y|x)=\frac{P[X=x|Y=y]f_Y(y)}{\int_y P[X=x|Y=y]f_Y(y)dy}.$$

(b) Given that $X=8$ and $f_Y(y)$ is uniform, we have

$$f_{Y|X}(y|8)=\frac{\binom{10}{8}y^8(1-y)^2}{\int_0^1\binom{10}{8}y^8(1-y)^2dy}=495y^8(1-y)^2.$$

We take the derivative of $f_{Y|X}(y|8)$ with respect to y and solve

$$\frac{d}{dy}f_{Y|X}(y|8)=0 \quad\longleftrightarrow\quad 10y^9-18y^8+8y^7=0$$

for y that maximizes $f_{Y|X}(y|8)$. The root for the equation is given by $y=\frac{4}{5}$, $y=1$, and $y=0$, where the latter two roots are clearly not the best choices (because they lead to zero density value). Furthermore, the answer $y=\frac{4}{5}$ is indeed an maximum point, since the 2nd derivative at $y=\frac{4}{5}$ is negative.

The answer $y=\frac{4}{5}$ agrees with our intuition. The data we have collected tells us that 8 out of 10 coin tosses appear to be heads, which reasonably provides the probability of heads to be 8/10.

2. Let X_1, X_2 and X_3 be jointly Gaussian random variables with the following properties: $X_1 \sim \mathcal{N}(0,1)$ and $X_2 \sim \mathcal{N}(0,1)$ are independent, $X_3 \sim \mathcal{N}(0,10)$, $E[X_1X_3] = -1$, and $E[X_2X_3] = 2$.

   (a) Let $\mathbf{x} = [X_1,\ X_2]^T$ and $Y = X_3$. Using equations on the top of page 14 of Topic 2 (which are from equations (2.96),(2.97),(2.79) and (2.80) of the textbook), the covariance matrix of $\mathbf{x} = [X_1,\ X_2]^T$ conditioning on $Y = X_3$ (i.e. $\mathbf{\Lambda}_{aa}^{-1}$ in equation 2.96) can be obtained by

$$\mathbf{K}_{\mathbf{x}|Y} = \mathbf{K}_{\mathbf{x}} - \mathbf{K}_{\mathbf{x}Y}\mathbf{K}_Y^{-1}\mathbf{K}_{Y\mathbf{x}},$$

   where $\mathbf{K}_{\mathbf{x}Y} = E[[X_1,X_2]^TY] = [-1,2]^T$ and $\mathbf{K}_Y = \mathrm{Var}(Y) = 10$. So, we have

$$\mathbf{K}_{\mathbf{x}|Y} = \mathbf{I} - \frac{1}{10}[-1,2]^T \cdot [-1,2],$$

   which can be expressed in terms of $\mathbf{I} + \alpha \cdot \mathbf{u}\mathbf{u}^T$ with $\mathbf{u} = \frac{1}{\sqrt{5}}[-1,2]^T$ and $\alpha = -1/2$.

   (b)

$$\begin{aligned}\mathbf{K}\cdot\mathbf{u} &= \left(\mathbf{I} + \alpha\cdot\mathbf{u}\mathbf{u}^T\right)\cdot\mathbf{u} \\ &= \mathbf{u} + \alpha\mathbf{u}\cdot\ \underbrace{(\mathbf{u}^T\mathbf{u})}_{\text{a scalar, =1 here}} \\ &= (1+\alpha)\cdot\mathbf{u},\end{aligned}$$

   where we see that $\mathbf{u}$ is an eigenvector of $\mathbf{K}$ with eigenvalue $1+\alpha$.

3. (a) Since $\mathbf{\Sigma}$ is just a 2×2 matrix here, i.e.

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix},$$

   its inverse has the following closed form representation

$$\mathbf{\Sigma}^{-1} = \frac{1}{\det(\mathbf{\Sigma})}\begin{bmatrix} \sigma_Y^2 & -\rho\sigma_X\sigma_Y \\ -\rho\sigma_X\sigma_Y & \sigma_X^2 \end{bmatrix},$$

   Direct expansion of the matrix-vector multiplication can show the result.

   (b) Consider the following pdf for X and Y:

$$f_{XY}(x,y) = \frac{1}{\pi\sqrt{3/4}} \cdot e^{-\frac{1}{2}\left(\frac{4}{3}x^2 + \frac{16}{3}y^2 + \frac{8}{3}xy - 8x - 16y + 16\right)}. \tag{3}$$

   We can write the joint pdf of any bivariate Gaussian X and Y as

$$f_{XY}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \cdot e^{-\left\{a(x-\mu_X)^2 + b(y-\mu_Y)^2 + c(x-\mu_X)(y-\mu_Y)\right\}},$$

   where

$$a = \frac{1}{2(1-\rho^2)\sigma_X^2}, b = \frac{1}{2(1-\rho^2)\sigma_Y^2}, c = \frac{-2\rho}{2(1-\rho^2)\sigma_X\sigma_Y}.$$

   1) By inspection of the given $f_{XY}(x,y)$, we find

$$a = \frac{2}{3}, b = \frac{8}{3}, c = \frac{4}{3},$$

which can be used to solve the three unknowns

$$\rho = \frac{c}{2\sqrt{ab}} = -\frac{1}{2},$$
$$\sigma_X^2 = \frac{1}{2(1-\rho^2)a} = 1,$$
$$\sigma_Y^2 = \frac{1}{2(1-\rho^2)b} = \frac{1}{4}.$$

To find μ_X and μ_Y, we solve a system of two linear equations:

$$2a\mu_X x + c\mu_Y x = 4x$$
$$2b\mu_Y y + c\mu_X y = 8y,$$

which yields $\mu_X = 2$ and $\mu_Y = 1$.
Then we can obtain $\mathrm{Cov}(X,Y) = \rho\sigma_X\sigma_Y = -\frac{1}{4}$.

2) See Fig. 1.

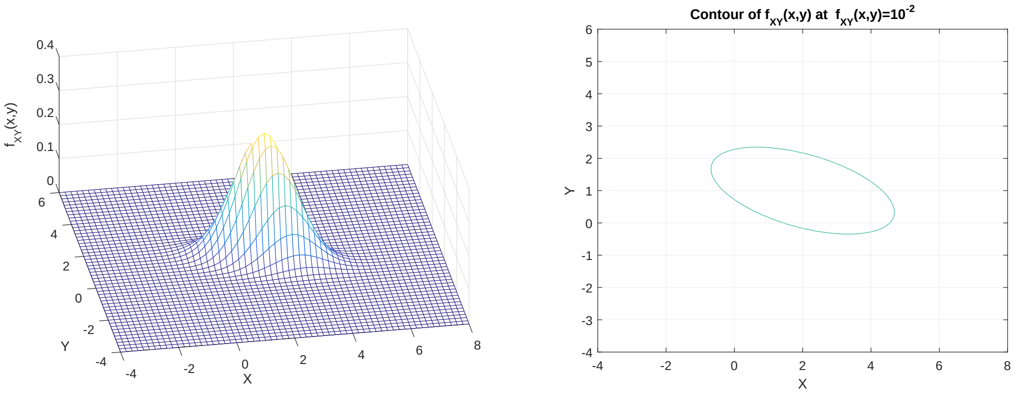

Figure 1: (Left) The meshed surface of $f_{XY}(x,y)$. (Right) The contour of $f_{XY}(x,y)$ at $f_{XY}(x,y) = 10^{-2}$.

This is how I plot the figures.

```
x_axis_sample = [ -4:.1:8 ];
y_axis_sample = [ -4:.1:6 ];
% Create x and y position matrices
[ x y ] = meshgrid( x_axis_sample, y_axis_sample );
f_xy = 1/( pi*sqrt( 3/4 ) )*exp( -.5*( 4*x.^2/3 + 16*y.^2/3 ...
+ 8*x.*y/3 - 8*x - 16*y + 16 ) );
c = 1e-2;
clf;
contour( x_axis_sample, y_axis_sample, f_xy, [ c c ] );
figure(2)
mesh(x_axis_sample, y_axis_sample, f_xy);
view(-10, 50)
```

4. (a) We need to evaluate the posterior probability:

$$\begin{aligned}
P[H=h|S=s,C=c,X=x] &\overset{(a)}{=} \frac{f_{X|S,C,H}(x|s,c,h)P[H=h|S=s,C=c]}{f_{X|S,C}(x|s,c)} \\
&\overset{(b)}{=} \frac{f_{X|S,C,H}(x|s,c,h)P[S=s,C=c|H=h]P_H(h)}{f_{X|S,C}(x|s,c)P[S=s,C=c]} \\
&\overset{(c)}{\propto} f_{X|S,C,H}(x|s,c,h)P[S=s,C=c|H=h]P_H(h) \\
&\overset{(d)}{=} f_{X|H}(x|h)P_{S|H}[s|h]P_{C|H}[c|h]P_H(h),
\end{aligned}$$

where (a) essentially bears the same form as equation (1) except with the conditioning on $\{S=s, C=c\}$, (b) holds from the Bayes' rule applied to $P[S=s,C=c|H=h]P_H(h)$, (c) holds from removing the denominator which is irrelevant when deciding H, and (d) follows from the assumption that S, C, and X are conditionally independent given H. Note that the likelihood functions $P_{S|H}[s|h$, $P_{C|H}[c|h]$ and the prior $P_H(h)$ can empirically obtained from the training data and have been readily written in the code `heartDisease.m`. The likelihood function $f_{X|H}(x|h)$ is assumed Gaussian, whose mean and variance are given and can also be learned from the training data.

So, the MAP reduces to

$$\hat{H} = \begin{cases} 0, & \text{if } f_{X|H}(x|0)P_{S|H}[s|0]P_{C|H}[c|0]P_H(0) > f_{X|H}(x|1)P_{S|H}[s|1]P_{C|H}[c|1]P_H(1) \\ 1, & \text{otherwise.} \end{cases}$$

The Matlab code for the estimate is as follows.

```
MAP_estimates = P_H1.*P_S_H1(sex_test).*P_C_H1(chest_pain_test).* f_X_H1(cholesterol_test)...
> P_H0.* P_S_H0(sex_test).* P_C_H0(chest_pain_test).* f_X_H0(cholesterol_test)};
```

(b) The error rate is 0.14.

```
error_rate= sum(abs( MAP_estimates- heart_disease_test ))./ length(heart_disease_test);
```