Machine Learning Homework 1

Karan Sarkar
sarkak2@rpi.edu

September 10, 2020

Problem 1.1.

(a) We have $|M| = r \cdot 3 \cdot 3 + 6 \cdot r \cdot 4 + 0 \cdot 2 \cdot 7 - r \cdot r \cdot 7 - 6 \cdot 2 \cdot 3 - 0 \cdot 3 \cdot 4$. Therefore, $|M| = 9r + 24r - 7r^2 - 36$. Thus, the determinant is $|M| = -7r^2 + 33r - 36$

(b) The inverse does not exist when $|M| = 0$. We find $|M| = 0$ when $-7r^2 + 33r - 36 = 0$. Thus, $-7r^2 + 12r + 21r - 36 = 0$. We then have $(-7r + 12)(r - 3) = 0$. Thus, we have no inverse when $r = 3, 12/7$. In these two situations, the matrix is singular and the rank is less than 3 i.e. 2.

(c)

$$\left[\begin{array}{ccc|ccc} 4 & 6 & 0 & 1 & 0 & 0 \\ 2 & 3 & 4 & 0 & 1 & 0 \\ 4 & 7 & 3 & 0 & 0 & 1 \end{array}\right]$$
$$\left[\begin{array}{ccc|ccc} 4 & 6 & 0 & 1 & 0 & 0 \\ 4 & 6 & 8 & 0 & 2 & 0 \\ 4 & 7 & 3 & 0 & 0 & 1 \end{array}\right]$$
$$\left[\begin{array}{ccc|ccc} 4 & 6 & 0 & 1 & 0 & 0 \\ 0 & 0 & 8 & -1 & 2 & 0 \\ 0 & 1 & 3 & -1 & 0 & 1 \end{array}\right]$$
$$\left[\begin{array}{ccc|ccc} 4 & 6 & 0 & 1 & 0 & 0 \\ 0 & 1 & 3 & -1 & 0 & 1 \\ 0 & 0 & 8 & -1 & 2 & 0 \end{array}\right]$$
$$\left[\begin{array}{ccc|ccc} 4 & 6 & 0 & 1 & 0 & 0 \\ 0 & 1 & 3 & -1 & 0 & 1 \\ 0 & 0 & 1 & -\frac{1}{8} & \frac{1}{4} & 0 \end{array}\right]$$
$$\left[\begin{array}{ccc|ccc} 4 & 6 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -\frac{5}{8} & -\frac{3}{4} & 1 \\ 0 & 0 & 1 & -\frac{1}{8} & \frac{1}{4} & 0 \end{array}\right]$$
$$\left[\begin{array}{ccc|ccc} 4 & 0 & 0 & \frac{19}{4} & \frac{9}{2} & -6 \\ 0 & 1 & 0 & -\frac{5}{8} & -\frac{3}{4} & 1 \\ 0 & 0 & 1 & -\frac{1}{8} & \frac{1}{4} & 0 \end{array}\right]$$
$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{19}{16} & \frac{9}{8} & -\frac{3}{4} \\ 0 & 1 & 0 & -\frac{5}{8} & -\frac{3}{4} & 1 \\ 0 & 0 & 1 & -\frac{1}{8} & \frac{1}{4} & 0 \end{array}\right]$$

Thus, the inverse is $\begin{bmatrix} \frac{19}{16} & \frac{9}{8} & -\frac{3}{4} \\ -\frac{5}{8} & -\frac{3}{4} & 1 \\ -\frac{1}{8} & \frac{1}{4} & 0 \end{bmatrix}$.

(d) The determinant of M when $r = 4$ is $-7 \cdot 4^2 + 33 * 4 - 36 = -16$. Thus, the determinant of the inverse is $-\frac{1}{16}$.

Problem 1.2. We have $Ax = \lambda x$. Therefore, we have $Ax - \lambda x = 0$. Thus, $Ax - \lambda Ix = 0$. We can now factor to get $(A - \lambda I)x = 0$. Suppose that $A - \lambda I$ is invertible. Therefore, there must exist matrix Q such that $Q(A - \lambda I) = I$. Thus for all x, we have $Q(A - \lambda I)x = Q0$. Thus, for all x we have $Ix = x = 0$. This is a contradiction as this cannot be true for all x. Therefore, our assumption that $A - \lambda I$ is invertible must be false. Thus, $A - \lambda I$ is not invertible and thus $|A - \lambda I| = 0$.

Problem 1.3.

(a) We want to find the eigenvalues. The eigenvalues λ satisfy:

$$\begin{vmatrix} x-\lambda & 3 \\ 1 & x-\lambda \end{vmatrix} = 0$$

This means $(x-\lambda)^2 - 3 = 0$. Consequently, $(x-\lambda)^2 = 3$. Therefore, $x - \lambda = \pm\sqrt{3}$. Thus, we have the eigenvalues are $\lambda = x \pm \sqrt{3}$

(b) The eigenvectors satisfy the following property.

$$\begin{bmatrix} x & 3 \\ 1 & x \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} ax \pm a\sqrt{3} \\ bx \pm b\sqrt{3} \end{bmatrix}$$

We now get two equations. The first is $ax + 3b = ax \pm a\sqrt{3}$. We yields $a = \pm\sqrt{3}b$. The second equation is $a + bx = bx \pm b\sqrt{3}$ which gives the same result. Our eigenvectors look like $(\pm\sqrt{3}b, b)$. THe norm of this is $\sqrt{3b^2 + b^2} = 2b$. Thus the normalized eigenvectors are $(\sqrt{3}/2, 1/2)$ and $(-\sqrt{3}/2, 1/2)$.

Problem 2.

(a) We want to compute $p(z)$. We can express the random variable Z as $P(Z = z) = P(Z = z|X = c)P(X = c) + P(Z = z|X = -c)P(X = -c)$. Therefore, we have $P(Z = z) = 0.5P\left(Y = \frac{z}{c}\right) + 0.5P\left(Y = -\frac{z}{c}\right)$. Now note that Y is symmetrical, therefore, we have $P(Z = z) = 0.5P\left(Y = \frac{z}{c}\right) + 0.5P\left(Y = \frac{z}{c}\right) = P\left(Y = \frac{z}{c}\right)$. The distribution of Z is the distribution of Y scaled out by c. Therefore, Z is just a Gaussian centered at 0 but scaled out by c. Thus $Z = \mathcal{N}(0, c)$. Thus the mean is 0 and the standard deviation is c.

(b) We want to compute the covariance of Z and Y. We will first compute $E[YZ] = E[XY^2]$. Because X and Y are independent, the covariance of X and Y^2 is zero. Thus $E[XY^2] = E[X]E[Y^2]$. This equals zero because $E[X] = 0$. THus, $E[YZ] = 0$. Moreover because $E[Y] = 0$ and $E[Z] = 0$. We have $E[Y]E[Z] = 0$. Thus the covariance $\text{Cov}[X, Y] = E[XY] - E[X][Y] = 0$. Thus the covariance is always zero.

(c) We assert that Y and Z are not independent. Consider $P[Z \in (2c, 3c)]$. From using a Z-Table and the result $P(Z = z) = P\left(Y = \frac{z}{c}\right)$, the probability evaluates to $0.99865 - 0.97725 = 0.0214$. Now consider $P[Z \in (2c, 3c)|Y \in (-1, 0)]$. The probability is now 0 as $|Z| < 1$ if Y is bounded like thus. Because the probability of Z changes, the variables are not independent.

Problem 3.

(a) We get that given $\lambda_1, \lambda_2 > 0$. the Lagrange function is:

$$\mathcal{L}(x, y) = 2x^2 + 3xy - \lambda_1 \left(\frac{1}{2}x^2 + y - 4\right) - \lambda_2(-y + 2)$$

(b) . The KKT conditions are all follows.

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial x} &= 4x + 3y - \lambda_1 x = 0 \\ \frac{\partial \mathcal{L}}{\partial y} &= 3x - \lambda_1 + \lambda_2 = 0 \\ \lambda_1 \left(\frac{1}{2}x^2 + y - 4\right) &= 0 \\ \lambda_2(-y + 2) &= 0\end{aligned}$$

(c) We will first consider the case where both constraints are binding. Thus, we have $\frac{1}{2}x^2 + y = 4$ and $\lambda_1 > 0$. Similarly, we get $y = 2$ and $\lambda_2 > 0$. Thus, we get $x = \pm 2$. First consider $x = 2$. Therefore, we have $4 \cdot 2 + 3 \cdot 2 - \lambda_1 \cdot 2 = 0$. This yields $\lambda_1 = 7$. We also get $6 \cdot 2 - 7 + \lambda_2 = 0$. Thus, we have $\lambda_2 = 1$. This gives us the candidate $(2, 2)$. From $x = -2$, we get $4 \cdot -2 + 3 \cdot 2 - \lambda_1 \cdot -2 = 0$. This gives $\lambda_1 = 1$. We also get $3(-2) - 1 + \lambda_2 = 0$. Thus $\lambda_2 = 7$. Thus, $(-2, 2)$ is a candidate.

We will now consider the case where the first constraint is binding but not the second. Thus, we have $\frac{1}{2}x^2 + y = 4$ and $\lambda_1 > 0$. Now we, get $y > 2$ and $\lambda_2 = 0$. We get that $y = 4 - \frac{1}{2}x^2$. Therefore, we can substitute to get $4x + 3\left(4 - \frac{1}{2}x^2\right) - \lambda_1 x = 0$. Therefore, we have $\lambda_1 = 4 + \frac{12}{x} - \frac{3}{2}x$. More substitution gets us

$$\begin{aligned}3x - 4 - \frac{12}{x} + \frac{3}{2}x &= 0 \\ -4 - \frac{12}{x} + \frac{9}{2}x &= 0 \\ 9x^2 - 8x - 24 &= 0 \\ x &= \frac{4 \pm 2\sqrt{58}}{9}\end{aligned}$$

This gives $x = -1.248$ and $x = 2.139$. We will check the negative solution first. This gives $\lambda_1 = -3.743$. Thus that is not a candidate. In the other case, we get $\lambda_1 = 6.401$. Solving for y gets us $y = 1.712$. This violates the condition for y so it is also not a candidate.

We will now consider the case where the second constraint is binding but not the first. We have $\frac{1}{2}x^2 + y < 4$ and $\lambda_1 = 0$. Similarly, we get $y = 2$ and $\lambda_2 > 0$. Thus, we have $4x + 3(2) = 0$. Thus $x = -\frac{3}{2}$. We then get $\lambda_2 = \frac{9}{2}$. Thus $\left(-\frac{3}{2}, 2\right)$ is a candidate.

We will now consider the case when both constraints are non binding. We have $\frac{1}{2}x^2 + y < 4$ and $\lambda_1 = 0$. Now we, get $y > 2$ and $\lambda_2 = 0$. We have that $x = 0$ and $y = 0$. This is not a candidate because it does not satisfy the constraints.

(d) The candidates are $(-2, 2), (2, 2), \left(-\frac{3}{2}, 2\right)$

(e) We can substitute each of our candidates. For $(-2, 2)$, we get $2(-2)^2 + 3(-2)(2) = -4$. For $(2, 2)$, we get $2(2)^2 + 3(2)(2) = 20$. For $\left(-\frac{3}{2}, 2\right)$, we get $2(-1.5)^2 + 3(-1.5)(2) = -4.5$. Thus $(2, 2)$ getting a value of 20 is the maximal value.

Problem 4.1.

(a) The likelihood is $(1-\theta)^5\theta^2(2\theta)^3 = 8\theta^5(1-\theta)^5$.

(b) The maximal value either occurs at the end of the interval or at a critical point. The likelihood at the end of the interval when $\theta = 0, 1$ is 0. The derivative can be calculated as follows:

$$\begin{aligned} p &= 8\theta^5(1-\theta)^5 \\ p &= 8(\theta-\theta^2)^5 \\ p' &= 8(5)(\theta-\theta^2)^4(1-2\theta) \end{aligned}$$

The roots of the derivative are at $\theta = 0, 1, 0.5$. We previously saw the likelihood at $0, 1$ was 0. At 0.5, the likelihood is $8(0.5)^10 = 0.0078125$. Thus, the maximal likelihood estimation is $\theta = 0.5$.

Problem 4.2.

(a) Consider $X_n = \theta t_n^3 + Z_n$. This is essentially shifting the Normal distribution by a constant. Adding a constant to a normal distribution just shifts the mean. This means X_n can be described by $\mathcal{N}(\theta t_n^3, \sigma^2)$ where σ^2 describes the variance. Thus the pdf of X_n is:

$$f_{X_n} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x_n - \theta t_n^3}{\sigma}\right)^2}$$

We can substitute to find the log likelihood.

$$\begin{aligned} l(\theta; x_1, x_2, \dots x_N) &= \sum_{n=1}^{N} \log \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x_n - \theta t_n^3}{\sigma}\right)^2} \\ &= \sum_{n=1}^{N} -\left(\frac{x_n - \theta t_n^3}{\sigma}\right)^2 - \log(\sigma\sqrt{2\pi}) \\ &= -N\log(\sigma\sqrt{2\pi}) - \frac{1}{\sigma^2}\left[(x_1 - \theta t_1^3)^2 + (x_2 - \theta t_2^3)^2 + \cdots (x_N - \theta t_N^3)^2\right] \end{aligned}$$

(b) This is equivalent to minizing the following quantity:

$$(x_1 - \theta t_1^3)^2 + (x_2 - \theta t_2^3)^2 + \cdots (x_N - \theta t_N^3)^2$$

We will differentiate by θ and try to find what value of θ sets the derivative to 0.

$$\begin{aligned} &(x_1 - \theta t_1^3)^2 + (x_2 - \theta t_2^3)^2 + \cdots (x_N - \theta t_N^3)^2 \\ 0 &= 2(x_1 - \theta t_1^3)(-t_1^3) + 2(x_2 - \theta t_2^3)(-t_2^3) + \cdots + 2(x_N - \theta t_N^3)(-t_N^3) \\ 0 &= \theta t_1^6 - x_1 t_1^3 + \theta t_2^6 - x_2 t_2^3 + \cdots \theta t_N^6 - x_N t_N^3 \\ \theta\left(t_1^6 + t_2^6 + \cdots + t_N^6\right) &= x_1 t_1^3 + x_2 t_2^3 + \cdots + x_N t_N^3 \\ \theta &= \frac{x_1 t_1^3 + x_2 t_2^3 + \cdots + x_N t_N^3}{t_1^6 + t_2^6 + \cdots + t_N^6} \end{aligned}$$

This expression provides the maximum likelihood estimate of θ.

Problem 4.3.

1. We can compute the pdf simply by differentiating the cdf. Thus, $p(x) = 0$ when $x < 0$. Thus, when $0 \le x \le \beta$, we have $p(x) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}$. Lastly, when $x > \beta$, $p(x) = 0$.

2. Suppose we have observations $x_1, x_2, \dots x_N$. We can compute the log-likelihood as follows:

$$
\begin{aligned}
&\frac{\alpha}{\beta}\left(\frac{x_1}{\beta}\right)^{\alpha-1}\frac{\alpha}{\beta}\left(\frac{x_2}{\beta}\right)^{\alpha-1}\cdots\frac{\alpha}{\beta}\left(\frac{x_N}{\beta}\right)^{\alpha-1}\\
&\frac{\alpha^N}{\beta^N}(x_1x_2\cdots x_N)^{\alpha-1}\\
&N\log\alpha - N\log\beta + (\alpha-1)(\log x_1 + \log x_2 + \cdots + \log x_N)
\end{aligned}
$$

We will now differentiate α find when the partial derivative equals 0.

$$
\begin{aligned}
&N\log\alpha - N\log\beta + (\alpha-1)(\log x_1 + \log x_2 + \cdots + \log x_N)\\
0 &= \frac{N}{\alpha} + \log x_1 + \log x_2 + \cdots + \log x_N\\
\frac{1}{\alpha} &= -\frac{\log x_1 + \log x_2 + \cdots + \log x_N}{N}\\
\alpha &= -\frac{N}{\log x_1 + \log x_2 + \cdots + \log x_N}
\end{aligned}
$$

We will now differentiate β find when the partial derivative equals 0. The partial derivative is $-\frac{N}{\beta}$. There seems to be no dependency on the data values x_n. What we see is that to maximize the likelihood we want β to be as small as possible. Note that the likelihood will equal 0 if any x_n exceeds β. Therefore, $\beta = \max(x_1, x_2, \dots, x_n)$. and $\alpha = -\frac{N}{\log x_1 + \log x_2 + \cdots + \log x_N}$.

Problem 5.1.

(a) The marginal distribution of X is $P(X = 0) = 2/3$ and $P(X = 1) = 1/3$. The marginal distribution of Y is $P(Y = 1) = 1/3$ and $P(Y = 2) = 2/3$.

(b) We have $I(X; Y) = H(X) + H(Y) - H(X, Y)$. We have $H(X) = -\frac{1}{3}\log_2(1/3) - \frac{2}{3}\log_2(2/3) = 0.918$. We have $H(Y) = -\frac{1}{3}\log_2(1/3) - \frac{2}{3}\log_2(2/3) = 0.918$. We have $H(X, Y) = -\frac{1}{3}\log_2(1/3) - \frac{1}{3}\log_2(1/3) - \frac{1}{3}\log_2(1/3) = 1.58$. Thus the mutual information is $I(X; Y) = 0.2516$.

Problem 5.2.

(a) We have $H(Y) = -\frac{3}{7}\log_2(3/7) - \frac{4}{7}\log_2(4/7) = 0.985$.

(b) We have $H(Y|x_1) = H(Y, x_1) - H(x_1)$. We get $H(Y, x_1) = -\frac{2}{7}\log_2(2/7) - \frac{2}{7}\log_2(2/7) - \frac{1}{14}\log_2(1/14) - \frac{3}{14}\log_2(3/14) - \frac{1}{7}\log_2(1/7) = 2.182$. We also have $H(x_1) = -\frac{5}{14}\log_2\left(\frac{5}{14}\right) - \frac{2}{7}\log_2\left(\frac{2}{7}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right) = 1.577$. Thus, $H(Y|x_1) = 0.604$.

We have $H(Y|x_4) = H(Y, x_4) - H(x_4)$. We get $H(Y, x_4) = -\frac{3}{14}\log_2\left(\frac{3}{14}\right) - \frac{3}{14}\log_2\left(\frac{3}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right) - \frac{3}{14}\log_2\left(\frac{3}{14}\right) = 1.959$. We have $H(x_4) = -\frac{3}{7}\log_2(3/7) - \frac{4}{7}\log_2(4/7) = 0.985$. Thus, $H(Y|x_4) = 0.974$.

(c) We have $I(x_1; Y) = H(Y) - H(Y|x_1) = 0.985 - 0.604 = 0.38$. Similarly, we have $I(x_4; Y) = H(Y) - H(Y|x_4) = 0.985 - 0.974 = 0.011$. THus x_1 or Age is a more informative predictor or whether one will quarantine.

(d) We have $H(Y, x_3) = -\frac{4}{14}\log_2\left(\frac{4}{14}\right) - \frac{3}{14}\log_2\left(\frac{3}{14}\right) - \frac{5}{14}\log_2\left(\frac{5}{14}\right) - \frac{2}{14}\log_2\left(\frac{2}{14}\right) = 1.924$.

Problem 5.3.

(a) Note that because X and Y are independent we have $P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$. Also note that $\sum_i P(X = x_i) = 1$ and $\sum_j P(Y = y_j) = 1$.

$$\begin{aligned}
H(X|Y) &= -\sum_i \sum_j P(X = x_i, Y = y_j) \log \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} \\
&= -\sum_i \sum_j P(X = x_i)P(Y = y_j) \left(\log P(X = x_i)P(Y = y_j) - \log P(Y = y_j)\right) \\
&= -\sum_i \sum_j P(X = x_i)P(Y = y_j) \log P(X = x_i)P(Y = y_j) + \sum_i \sum_j P(X = x_i)P(Y = y_j) \log P(Y = y_j) \\
&= -\sum_i \sum_j P(X = x_i)P(Y = y_j) \log P(X = x_i) - \sum_i \sum_j P(X = x_i)P(Y = y_j)P(Y = y_j) \\
&\quad + \sum_i \sum_j P(X = x_i)P(Y = y_j) \log P(Y = y_j) \\
&= -\left(\sum_j P(Y = y_j)\right) \sum_i P(X = x_i) \log P(X = x_i) \\
&= -\sum_i P(X = x_i) \log P(X = x_i) \\
H(X|Y) &= H(X)
\end{aligned}$$

(b) Note that because X and Y are independent we have $P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$. Also note that $\sum_i P(X = x_i) = 1$ and $\sum_j P(Y = y_j) = 1$.

$$\begin{aligned}
H(X,Y) &= -\sum_i \sum_j P(X = x_i, Y = y_j) \log P(X = x_i, Y = y_j) \\
&= -\sum_i \sum_j P(X = x_i, Y = y_j)(\log P(X = x_i) + \log(Y = y_j)) \\
&= -\sum_i \sum_j P(X = x_i)P(Y = y_j) \log P(X = x_i) - \sum_i \sum_j P(X = x_i)P(Y = y_j) \log(Y = y_j) \\
&= -\left(\sum_i P(X = x_i) \log P(X = x_i)\right) \sum_j P(Y = y_j) - \left(\sum_i P(X = x_i)\right) \sum_j P(Y = y_j) \log(Y = y_j) \\
&= -\sum_i P(X = x_i) \log P(X = x_i) - \sum_j P(Y = y_j) \log P(Y = y_j) \\
H(X,Y) &= H(X) + H(Y)
\end{aligned}$$

(c) Note that $\sum_i P(X = x_i) = 1$ and $\sum_j P(Y = y_j) = 1$. We will simplify $I(X;Y)$ in both ways and show that both ways lead to the same answer.

$$
\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) \\
&= H(X) + \sum_i \sum_j P(X = x_i, Y = y_j) \log \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} \\
&= H(X) + \sum_i \sum_j P(X = x_i, Y = y_j) \left(\log P(X = x_i, Y = y_j) - \log P(Y = y_j)\right) \\
&= H(X) + \sum_i \sum_j P(X = x_i, Y = y_j) \log P(X = x_i, Y = y_j) - \sum_i \sum_j P(X = x_i, Y = y_j) \log P(Y = y_j) \\
&= H(X) + \sum_i \sum_j P(X = x_i, Y = y_j) \log P(X = x_i, Y = y_j) - \sum_j P(Y = y_j) \log P(Y = y_j) \\
I(X;Y) &= H(X) + H(Y) - H(X,Y)
\end{aligned}
$$

Now will will do it the other way.

$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= H(Y) + \sum_i \sum_j P(X = x_i, Y = y_j) \log \frac{P(X = x_i, Y = y_j)}{P(X = x_i)} \\
&= H(Y) + \sum_i \sum_j P(X = x_i, Y = y_j) \left(\log P(X = x_i, Y = y_j) - \log P(X = x_i)\right) \\
&= H(Y) + \sum_i \sum_j P(X = x_i, Y = y_j) \log P(X = x_i, Y = y_j) - \sum_i \sum_j P(X = x_i, Y = y_j) \log P(X = x_i) \\
&= H(Y) + \sum_i \sum_j P(X = x_i, Y = y_j) \log P(X = x_i, Y = y_j) - \sum_i P(X = x_i) \log P(X = x_i) \\
I(X;Y) &= H(X) + H(Y) - H(X,Y)
\end{aligned}
$$

Thus, we see that mutual information is symmetric.

Problem 6.

(a) We know that $E[X] = 8$. Moreover, because we have a Poisson Distribution, $\text{VAR}[X] = 8$.

$$\begin{aligned} E[(X+2)^2] &= E[X^2 + 4X + 4] \\ &= E[X^2] + 4E[X] + 4 \\ &= \text{VAR}[X] + E[X]^2 + 4(8) + 4 \\ &= 8 + 8^2 + 32 + 4 \\ &= 108 \end{aligned}$$

(b) We will list the possibilities with their probability. $P(H) = 1/2$. $P(TH) = 1/4$. $P(TTH) = 1/8$. $P(TTT) = 1/8$. Consider random variable Y. We have $P(Y = 1) = 7/8$ and $P(Y = 0) = 1/8$. Thus, $E[Y] = 7/8$. $E[Y^2] = 7/8$. Thus, $\text{VAR}[Y] = 56/64 - 49/64 = 7/64$. Thus, the variance is $7/64$.

(c) We can write this as an integral.

$$\begin{aligned} P(X + Y \le 1) &= \int_0^1 \int_0^{1-y} (x+y)dxdy \\ &= \int_0^1 \frac{1}{2}x^2 + xy\Big|_0^{1-y} dy \\ &= \int_0^1 \frac{1}{2}(1-y)^2 + (1-y)ydy \\ &= \int_0^1 \frac{1}{2} - y + \frac{1}{2}y^2 + y - y^2 dy \\ &= \int_0^1 \frac{1}{2} + \frac{1}{2}y^2 dy \\ &= \frac{1}{2}y + \frac{1}{6}y^3\Big|_0^1 \\ &= \frac{2}{3} \end{aligned}$$

Thus, the probability is $2/3$.