1. (a)

$h_{(\theta)} = \theta^T x^{(i)}$

(b)

$y^{(i)}$ *is* individual y input for each i

$J_{(\theta)} = 1 / 2m * \sum(\theta^T x^{(i)} - y^{(i)})$

(c)

$\partial J_{(\theta)} / \partial \theta = x / m * \sum x^{(i)}(\theta^T x^{(i)} - y^{(i)})$

(d)

$\theta_j = \theta_j - x / m * \sum x^{(i)}(\theta^T x^{(i)} - y^{(i)})$

(e)

$Y = [y^{(0)}, y^{(1)}, y^{(2)} \ldots y^{(n)}]$

The theta update rule in the gradient descent procedure in complete vector form:

$\theta_j = \theta_j - x / m * (X\theta - Y)^T * x^{(i)}$

2.(a)

$P_{(X=0,Y=0)} = a / (a + b + c + d)$

$P_{(X=0,Y=1)} = c / (a + b + c + d)$

$P_{(X=1,Y=0)} = b / (a + b + c + d)$

$P_{(X=1,Y=1)} = d / (a + b + c + d)$

(b)

$P(X = 0) = (a + c) / (a + b + c + d)$

(c)

$P(x = 1 | y = 0) = b / (a + b)$

(d)

$P(x = 1 \bigcup y = 0) = (a + b + d) / (a + b + c + d)$

(e)

$\mu_x = (b + d) / (a + b + c + d)$

$\mu_y = (c + d) / (a + b + c + d)$

$cov(X,Y) = (0 - \mu_x)(0 - \mu_y)a / (a + b + c + d) + (0 - \mu_x)(1 - \mu_y)c / (a + b + c + d) + (1 - \mu_x)(0 - \mu_y)b / (a + b + c + d) + (1 - \mu_x)(1 - \mu_y)d / (a + b + c + d)$

3.(a)

$\mu = \sum X / N = 55 / 6 \approx 9$

$\sigma^2 = \sum (X - \mu)^2 / N = 1 / 6[(2 - 9)^2 + (5 - 9)^2 + (7 - 9)^2 + (7 - 9)^2 + (25 - 9)^2] = 54.8$

(b)

$f(x;\mu,\sigma^2) = 1 / (\sigma\sqrt{2\pi}) * e^{-1/2 * (x - \mu / \sigma)^2} = 1 / (7.4 * 2.5) * e^{-1.1} = 1 / 56$

(c)

joint probability density function = multiplications of each pdf

$f_{x_1, x_2, \ldots x_6}(2,5,7,7,9,25) = (1 / \sigma\sqrt{2\pi})^6 e^{[-(2-\mu)^2 - (5-\mu)^2 - 2(7-\mu)^2 - (25-\mu)^2] / (2\sigma^2)} = (1 / \sigma\sqrt{2\pi})^6 e^{-3} = 0.0498 * 2.5 * 10^{-8} = 1.2 * 10^{-11}$

(d)

$f_{x_1, \ldots x_6}(2,5,7,7,8,9) > f_{x_1, x_2, \ldots x_6}(2,5,7,7,9,25)$

Because $e^{[-(2-\mu)^2 - (5-\mu)^2 - 2(7-\mu)^2 - (25-\mu)^2] / (2\sigma^2)} \prec e^{[-(2-\mu)^2 - (5-\mu)^2 - 2(7-\mu)^2 - (9-\mu)^2] / (2\sigma^2)}$

(e)

covariance $= E((X - E(X))(Y - E(Y))) = 1 / 6 * [(2 - 9)(4 - 6) + (5 - 9)(4 - 6) + (7 - 9)(5 - 6) + (25 - 9)(10 - 6)] = 14.7$

(f)

MSE is the sum of the variances and the squared bias. In cases of unbiased estimators. MSE and variances are equivalent.

I worked with Julian and Vicky

4.(a)

$P_{X_1 \ldots X_n}(x_{1 \ldots} x_n) = p(x_1)p(x_2)\ldots p(x_n)$

(b)

It depends. The added probability mass funciton can be smaller or bigger than 1; if smaller, threshold should decrease; if bigger, threshold should increase.