Alex Fay
HW 4
Written Portion

(1A) It is given that the distribution is guassian, thus we can use the Marginal equation (4.69 Murphey)

$$p(x_1) = \mathcal{N}(x_1 | \mu_1, \Sigma_{11})$$
$$= \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 6 & 8 \\ 8 & 13 \end{bmatrix}\right)$$

$$\left.\begin{array}{l} \mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \Sigma_{11} = \begin{bmatrix} 6 & 8 \\ 8 & 13 \end{bmatrix} \end{array}\right\} \text{given in problem}$$

(1B) $p(x_2)$ can be fand using the same process.

$$p(x_2) = \mathcal{N}(x_2 | \mu_2, \Sigma_{22}) = \mathcal{N}(\underbrace{5, 14}_{\text{givens}})$$

(1C) From the canditional distribution equation in Murphey 4.69, we know that $p(x_1 | x_2) = \mathcal{N}(x_1 | \mu_{1|2}, \Sigma_{1|2})$

Thus...

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} \quad \text{(given 4.69)}$$

$$= \begin{bmatrix} 6 & 8 \\ 8 & 13 \end{bmatrix} - \frac{1}{14} \begin{bmatrix} 5 \\ 11 \end{bmatrix} [5 \quad 11]$$

$$= \begin{bmatrix} 59/14 & 57/14 \\ 57/14 & 61/14 \end{bmatrix}$$

$$\boxed{p(x_1 | x_2) = \mathcal{N}\left(\frac{1}{14}\begin{bmatrix} 5 \\ 11 \end{bmatrix}(x_2 - 5), \begin{bmatrix} 59/14 & 57/14 \\ 57/14 & 61/14 \end{bmatrix}\right)}$$

(1d) $p(x_2 | x_1)$ can be fand by the same Method above:

$$\Sigma_{2|1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$$

$$\mu_{2|1} = 5 + [5 \quad 11] \begin{bmatrix} 6 & 8 \\ 8 & 13 \end{bmatrix}^{-1} (x_1 - \mu_1)$$

$$= 5 + \begin{bmatrix} \frac{-23}{14} & \frac{13}{7} \end{bmatrix} x_1$$

$$\boxed{p(x_2 | x_1) = \mathcal{N}\left(5 + \begin{bmatrix} \frac{-23}{14} & \frac{13}{7} \end{bmatrix} x_1, \frac{25}{14}\right)}$$

2a) Proof / Math for graphs. Graphs posted on Github.

Regularized logistic model for $P(y=1/x)$:

Use negative distribution log eqn:

$$nll(\theta) = -\sum y_i \log(\sigma(\theta^T x)) + \frac{\lambda}{2}\|\theta\|_2^2 + (1-y_i)\log(1-\sigma(\theta^T x))$$

To find rate of change:

$$\nabla nll(\theta) = \sum y_i(1-\theta^T x|\theta)x - (1-y_i)\sigma(\theta^T x)x + \lambda\theta$$

$$= \sum [y - \sigma(\theta^T x)]x + \lambda\theta$$

$$= X^T(\sigma(X\theta) - y) + \lambda\theta.$$

We can use this to find the local curvature of $nll(\theta)$ by taking the second gradient of $nll(\theta)$

$$\nabla^2 nll(\theta) = \nabla\sum X^T(\sigma(X\theta) - y) + \lambda\theta$$

$$\nabla_\theta^2 = X^T \operatorname{diag}[\sigma(X\theta)(1-\sigma(X\theta))]X + \lambda I$$

2b) Please graphs on github. Math / process for graphs is below:

To find softmax over the entire data set w/ Guassian we use:

$$\text{From softmax: } P(y=c|x,w) = \frac{\exp(w_\ell^T x)}{\sum \exp(w_i x)}$$

$$\nabla_w nll(\theta) = X^T(\mu - y) + \lambda w$$

$$\mu = S(x_i) = \frac{\exp(W^T x)}{[\,\vdots\,]\exp(W^T x)}$$

For each number it is classified as 0 or 1, thus $y=0 \mid y=1$ for $\forall y$.

(2a) The accuracy of λ was best at 0. The values tested were

[0, .1, .5, 1, 2.5, 7, 15, 25, 3] The accuracy was 99.94%

(2b) The optimal parameter is .5 at 92.24%