Henry Trinh

HW4

1. Marginals and conditionals of an MVN:
Suppose $x = (x_1, x_2)$ is jointly gaussian with parameters:

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad \Lambda = \Sigma^{-1} = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}.$$

Then the marginals are given by $p(x_1) = N(x_1 | \mu_1, \Sigma_{11})$; $p(x_2) = N(x_2 | \mu_2, \Sigma_{22})$
and the posterior conditional is given by:

$p(x_1 | x_2) = N(x_1 | \mu_{1|2}, \Sigma_{1|2})$, where

$$\mu_{1|2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) = \mu_1 - \Lambda_{11}^{-1}\Lambda_{12}(x_2 - \mu_2) = \Sigma_{1|2}(\Lambda_{11}\mu_1 - \Lambda_{12}(x_2 - \mu_2))$$

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

Now, suppose a distribution where $\mu_1 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\mu_2 = 5$, $\Sigma_{11} = \begin{bmatrix} 6 & 8 \\ 8 & 13 \end{bmatrix}$, $\Sigma_{21} = \Sigma_{12} = \begin{bmatrix} 5 \\ 11 \end{bmatrix}$,
and $\Sigma_{22} = [14]$.

a. To compute the marginal distribution $p(x_1)$, we substitute:
$p(x_1) = N(x_1 | \mu_1, \Sigma_{11}) = \boxed{N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 6 & 8 \\ 8 & 13 \end{bmatrix}\right)}$ //

b. $p(x_2) = N(x_2 | \mu_2, \Sigma_{22}) = \boxed{N(5, 14)}$ //

c. To compute the conditional distribution $p(x_1 | x_2)$, we substitute:

$$\mu_{1|2} = \cancel{\mu_1} + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) = \frac{1}{14}\begin{bmatrix} 5 \\ 11 \end{bmatrix}(x_2 - 5)$$

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \begin{bmatrix} 6 & 8 \\ 8 & 13 \end{bmatrix} - \frac{1}{14}\begin{bmatrix} 5 \\ 11 \end{bmatrix}\begin{bmatrix} 5 & 11 \end{bmatrix} = \begin{bmatrix} 59/14 & 57/14 \\ 57/14 & 61/14 \end{bmatrix}$$

$$\therefore p(x_1 | x_2) = N(\mu_{1|2}, \Sigma_{1|2}) = \boxed{N\left(\frac{1}{14}\begin{bmatrix} 5 \\ 11 \end{bmatrix}(x_2 - 5), \begin{bmatrix} 59/14 & 57/14 \\ 57/14 & 61/14 \end{bmatrix}\right)} //$$

d. $p(x_2 | x_1) = \boxed{N\left(5 + \begin{bmatrix} -\frac{23}{14} & \frac{13}{7} \end{bmatrix} x_1, \frac{25}{14}\right)}$ //, based on the following calculations

$$\mu_{2|1} = \mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1) = 5 + \begin{bmatrix} 5 & 11 \end{bmatrix}\begin{bmatrix} 6 & 8 \\ 8 & 13 \end{bmatrix}^{-1}\begin{bmatrix} 5 \\ 11 \end{bmatrix}(x_1 - \mu_1)$$

$$\begin{bmatrix} 6 & 8 \\ 8 & 13 \end{bmatrix}^{-1} = \left[\begin{array}{cc|cc} 6 & 8 & 1 & 0 \\ 8 & 13 & 0 & 1 \end{array}\right] = \begin{bmatrix} 1 & 4/3 & 1/6 & 0 \\ 8 & 13 & 0 & 1 \end{bmatrix} \Rightarrow \left[\begin{array}{cc|cc} 1 & 0 & 13/14 & -4/7 \\ 0 & 1 & -4/7 & 3/7 \end{array}\right]$$

$$\therefore \mu_{2|1} = 5 + \begin{bmatrix} -23/14 & 13/7 \end{bmatrix} x_1$$

$$\Sigma_{2|1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} = 14 - \begin{bmatrix} 5 & 11 \end{bmatrix}\begin{bmatrix} 6 & 8 \\ 8 & 13 \end{bmatrix}^{-1}\begin{bmatrix} 5 \\ 11 \end{bmatrix} \Rightarrow \frac{25}{14}$$

2. MNIST dataset and regressions. We have handwritten digits with 28x28 pixels in each image, as well as the label of which digit $0 \leq label \leq 9$ the written digit corresponds to. Given a new image of a handwritten digit, we wish to predict what digit it is. The format of the data is:

label, pix-11, pix-12, pix-13, ... ; where pix-ij is the pixel in the i-th row and the j-th column.

a. Let $0 \leq label \leq 1$ for simplicity. Implement L2 regularized logistic regression to then compute $P(y=1|x)$ for a different regularization parameter λ. Then, plot the learning curve using Newton's Method v. Gradient Descent.

See hw4pr2a.py. Model + iteration results in hw4pr2a.txt.
We thus are able to see that the accuracy of our model is
This value is above 90%, as desired. We also note how Newton's Method is significantly more efficient than Gradient Descent, based on our plots.

b. Now, we use the whole dataset and predict the label of each digit using L2 regularized softmax regression {aka multinomial logistic regression}. Implement using gradient descent, and plot the accuracy on the test set for different values of λ.

See hw4pr2b.py. Results found in hw4pr2b.txt.
From our generated hw4pr2b_lva.png file, we see that our maximum test accuracy was roughly 92% at $\lambda = 0.01$.