ECE 403 June 23, 2020 Ashlynn Steeves
Assignment 5 V00850631

All MATLAB code used to generate answers for this assignment can be found at
https://github.com/ashlynns/ECE403/tree/master/A5

Q2.2

*(c) Apply BFGS to minimize ELR with input **w**=**0** , u = 0.03, and e = 10^-6. Report the numerical result of the minimizer **w** obtained. Use the optimized model to classify new input samples.*

$$\hat{\boldsymbol{w}}^* = [-1.3764, 1.5014, 1.0892, 1.3155, 0.0711, -0.1547]'$$

x_tr_1 is classified as -1 (Class N)

x_tr_1 is classified as 1 (Class P)

*(d) Apply memoryless BFGS to minimize ELR with input **w**= **0**, u = 0.03, and e = 10^-6. Report the numerical result of the minimizer **w** obtained. Use the optimized model to classify new input samples.*

$$\hat{\boldsymbol{w}}^* = [-1.3764, 1.5014, 1.0892, 1.3155, 0.0711, -0.1547]'$$

x_tr_1 is classified as -1 (Class N)

x_tr_1 is classified as 1 (Class P)

Q2.3

(a) *Use the minimizer $\hat{\boldsymbol{w}}^*$ obtained from Prob. 2.2(a) to produce a decision boundary defined in (P2.9). This can be accomplished by doing the following.*
   i. *Assuming the value of x1 is given and the minimizer $\hat{\boldsymbol{w}}^*$ is known, express the equation in (P2.9) as:*

$$Ax_2^2 + Bx_2 + C = 0$$

$$w_1^* x_1 + w_2^* x_2 + w_3^* x_1^2 + w_4^* x_1 x_2 + w_5^* x_2^2 + b^* = 0$$
$$(w_5^*)x_2^2 + (w_2^* + w_4^* x_1)x_2 + (w_1^* x_1 + w_3^* x_1^2 + b^*) = 0$$
$$Ax_2^2 + Bx_2 + C = 0$$

   where:

$$A = w_5^* = 0.0711$$
$$B = w_2^* + w_4^* x_1 = 1.5014 + 1.3155x_1$$
$$C = w_1^* x_1 + w_3^* x_1^2 + b^* = -1.3765x_1 + 1.0891x_1^2 - 0.1547$$

   * w values are taken from the posted solutions to Assignment 4

   ii. *Solve the equation in for x2 and take one of the solutions with plus sign for the square. Now place 200 evenly spaced grid points on interval [–1.5, 2.5]. Let x1 be the i-th grid point. Use that x1 together with $\hat{\boldsymbol{w}}^*$ to* *compute x2. If x2 turns out to be a real number use it to form a point pi = [x1, x2]'. If x2 turns out to be a complex number, simply ignore it. Repeat the above steps until every grid point x1 on [–1.5, 2.5] has been tried. In this way, you will get enough number of points to construct a matrix.*

$$x_2 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$= \frac{-(1.5014 + 1.3155x_1) \pm \sqrt{(1.5014 + 1.3155x_1)^2 - 4(0.0711)(-1.3765x_1 + 1.0891x_1^2 - 0.1547}}{2(0.0711)}$$

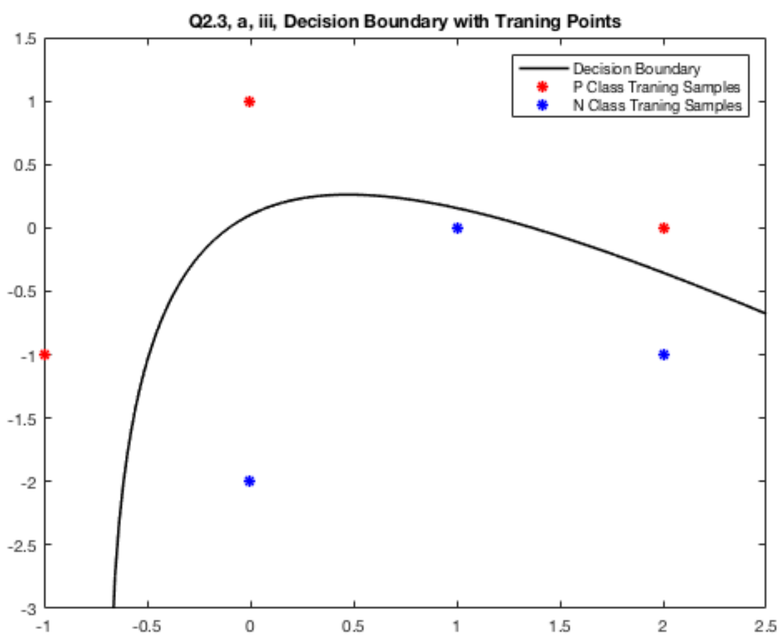
$$x_2 = \frac{-(1.5014 + 1.3155x_1) + \sqrt{1.4207x_1^2 + 4.3417x_1 + 2.298}}{0.1422}$$

MATLAB is used to create the x1 grid and evaluate x2 at each point using the code below

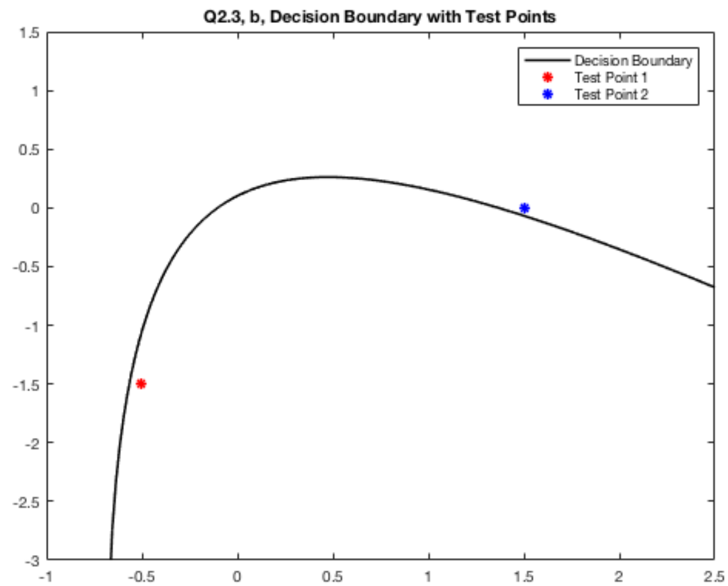
```
x1 = linspace(-1.5, 2.5, 200)
x2 = ((-(1.5014+1.3155.*x1))+sqrt((1.4207.*x1.^2)+(4.3417.*x1)+2.298))/0.1422
p = [];

for i = 1:length(x2)
    if isreal(x2(i))
        p = [p [x1(i); x2(i)]];
    end
end
```

iii. *The decision boundary can now be produced in MATLAB as plot(P(1,:),P(2,:),'k-','linew',1.5). For the decision boundary to be useful, plot it in a figure where the train samples given in (P2.1) have been marked.*



(b) Mark the two test samples used in Q2.2 in the figure generated in part (a), and classify them by using the decision boundary. Are the classification results consistent with those obtained in Prob. 2.2(a)?



Each testing point falls on a different side of the boundary, indicating that they are from different classes. Test point 1 is classified as class N and test point 2 is classified as class P. These results are consistent with the results obtained in problem 2.2.

Q2.5

Apply the memoryless BFGS algorithm to the function below with initial point ***x0*** *= [0 0]'.*

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \boldsymbol{x} + \boldsymbol{x}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

From example 2.4 we know that

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$
$$\nabla^2 f(\boldsymbol{x}) = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$$

Iteration 1:

$$\boldsymbol{x_0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{g_0} = \nabla f(\boldsymbol{x_0}) = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \boldsymbol{S_0} = \boldsymbol{I}$$
$$\boldsymbol{d_0} = -\boldsymbol{S_0}\nabla f(\boldsymbol{x_0}) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

$$\alpha_0 = \frac{\boldsymbol{g_0^T S_0 g_0}}{\boldsymbol{g_0^T S_0 H S_0 g_0}} = \frac{[1\ 0]\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}}{[1\ 0]\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}2 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\begin{bmatrix}1\\0\end{bmatrix}} = 0.5$$

$$\boldsymbol{x_1} = \boldsymbol{x_0} + \alpha_0 \boldsymbol{d_0} = \begin{bmatrix}0\\0\end{bmatrix} + 0.5\begin{bmatrix}-1\\0\end{bmatrix} = \begin{bmatrix}-0.5\\0\end{bmatrix}$$

$$\boldsymbol{\delta_0} = \boldsymbol{x_1} - \boldsymbol{x_0} = \begin{bmatrix}-0.5\\0\end{bmatrix}$$

$$\boldsymbol{g_1} = \nabla f(\boldsymbol{x_1}) = \begin{bmatrix}0\\-0.5\end{bmatrix}$$

$$\boldsymbol{\gamma_0} = \boldsymbol{g_1} - \boldsymbol{g_0} = \begin{bmatrix}0\\-0.5\end{bmatrix} - \begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}-1\\-0.5\end{bmatrix}$$

$$\rho_0 = \frac{1}{\boldsymbol{\gamma_0^T \delta_0}} = \frac{1}{[-1, 0.5]\begin{bmatrix}-0.5\\0\end{bmatrix}} = 2$$

$$t_0 = \boldsymbol{\delta_0^T g_1} = [-0.5, 0]\begin{bmatrix}0\\-0.5\end{bmatrix} = 0$$

$$\boldsymbol{q_0} = \boldsymbol{g_1} - \rho_0 t_0 \boldsymbol{\gamma_0} = \begin{bmatrix}0\\-0.5\end{bmatrix} - 2 * 0 * \begin{bmatrix}-1\\-0.5\end{bmatrix} = \begin{bmatrix}0\\-0.5\end{bmatrix}$$

Iteration 2:

$$\boldsymbol{x_1} = \begin{bmatrix}-0.5\\0\end{bmatrix},\ \boldsymbol{g_1} = \begin{bmatrix}0\\-0.5\end{bmatrix},\ \rho_0 = 2,\ \boldsymbol{\gamma_0} = \begin{bmatrix}-1\\-0.5\end{bmatrix},\ \boldsymbol{q_0} = \begin{bmatrix}0\\-0.5\end{bmatrix},\ t_0 = 0,\ \boldsymbol{\delta_0} = \begin{bmatrix}-0.5\\0\end{bmatrix}$$

$$\boldsymbol{d_1} = \rho_0(\ \boldsymbol{\gamma_0^T}\ \boldsymbol{q_0} -\ t_0)\boldsymbol{\delta_0} - \boldsymbol{q_0} = 2\left([-1, -0.5]\begin{bmatrix}0\\-0.5\end{bmatrix} - 0\right)\begin{bmatrix}-0.5\\0\end{bmatrix} - \begin{bmatrix}0\\-0.5\end{bmatrix}$$

$$\boldsymbol{d_1} = 0.5\begin{bmatrix}-0.5\\0\end{bmatrix} - \begin{bmatrix}0\\-0.5\end{bmatrix} = \begin{bmatrix}-0.25\\0.5\end{bmatrix}$$

Calculating S1 and in order to find α_1:

$$\boldsymbol{S_1} = \left(\boldsymbol{I} - \frac{\boldsymbol{\delta_0 \gamma_0^T}}{\boldsymbol{\gamma_0^T \delta_0}}\right)\left(\boldsymbol{I} - \frac{\boldsymbol{\gamma_0 \delta_0^T}}{\boldsymbol{\gamma_0^T \delta_0}}\right) + \frac{\boldsymbol{\delta_0 \delta_0^T}}{\boldsymbol{\gamma_0^T \delta_0}} = \begin{bmatrix}0 & -0.5\\0 & 1\end{bmatrix}\begin{bmatrix}0 & 0\\-0.5 & 1\end{bmatrix} + \begin{bmatrix}0.5 & 0\\0 & 0\end{bmatrix} = \begin{bmatrix}0.75 & -0.5\\-0.5 & 1.0\end{bmatrix}$$

$$\alpha_1 = \frac{\boldsymbol{g_1^T S_1 g_1}}{\boldsymbol{g_1^T S_1 H S_1 g_1}} = \frac{[0 - 0.5]\begin{bmatrix}0.75 & -0.5\\-0.5 & 1.0\end{bmatrix}\begin{bmatrix}0\\-0.5\end{bmatrix}}{[0 - 0.5]\begin{bmatrix}0.75 & 0.5\\0.5 & 2.0\end{bmatrix}\begin{bmatrix}2 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}0.75 & -0.5\\-0.5 & 1.0\end{bmatrix}\begin{bmatrix}0\\-0.5\end{bmatrix}} = 0.5$$

$$\boldsymbol{x_2} = \boldsymbol{x_1} + \alpha_1 \boldsymbol{d_1} = \begin{bmatrix} -0.5 \\ 0 \end{bmatrix} + 0.5 \begin{bmatrix} -0.25 \\ 0.5 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$\boldsymbol{g_2} = \nabla f(\boldsymbol{x_2}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Since the gradient at x2 evaluates to zero we know x2 = [-1, 1] is a stationary point. This is the same stationary point that is found in example 2.4.