Homework 4

Problem 1:

a. $L(w) = \sum_{i=1}^{n}(y^{(i)} - w.x^{(i)})^2$
$\frac{\partial L}{\partial w_j} = -\sum_{i=1}^{n} 2 * \left(y^{(i)} - w.x^{(i)}\right)x_j^{(i)}$

$\frac{\partial 2L}{\partial w_j \partial w_k} = H_{jk} = \sum_{i=1}^{n} 2 * x_j^{(i)} x_k^{(i)}$

If $v_{ji} = \sqrt{2} * x_j^{(i)}$, we can write H_{jk} as a product of 2 vectors.

$H_{jk} = v_j \cdot v_k - (Eq.1)$ Therefore,

$$H = \begin{pmatrix} \cdots & v_1 & \cdots \\ \cdots & \vdots & \cdots \\ \cdots & v_p & \cdots \end{pmatrix}\begin{pmatrix} \vdots & \vdots & \vdots \\ v_1 & \cdots & v_p \\ \vdots & \vdots & \vdots \end{pmatrix} = V^T V - (Eq.2)$$

For convexity, we have to prove that H is positive semidefinite.

$We\ can\ see\ that, H_{jk} = H_{kj}$ from Eq. 1.

From Eq. 2, $H = V^T V$.

Let z be any vector. Multiplying the equation by $z^T\ and\ z\ as\ follows$:

$z^T V^T V z = (z^T V^T)(Vz) = (Vz)^T(Vz) = ||Vz||^2 \geq 0$

Hence proved that H is a PSD and L(w) is a convex function.

b. $w_{t+1} = w_t - \eta_t \frac{\partial L}{\partial w} = w_t + 2\eta_t \sum_{i=1}^{n}\left(y^{(i)} - w.x^{(i)}\right)x_j^{(i)}$

c. $w_{t+1} = w_t - H^{-1} \frac{\partial L}{\partial w} = w_t + 2H^{-1} \sum_{i=1}^{n}\left(y^{(i)} - w.x^{(i)}\right)x_j^{(i)}$

Problem 2:

a. $f(x) = z^T M z \geq 0 = \sum_{i,j} M_{ij} x_i x_j\ (Since\ M\ is\ PSD)$
$\frac{\partial f}{\partial x_i} = 2M_{ii}x_i + \sum_{j\,!=i} M_{ij}x_j + \sum_{j\,!=i} M_{ji}x_j$
Since M is symmetric $M_{ij} = M_{ji}$
$\frac{\partial f}{\partial x_i} = 2 * \sum_{ji} M_{ij}x_j$
$\frac{\partial 2f}{\partial x_i \partial x_j} = H_{ij} = 2 * M_{ij}$
Since M is PSD, H is also PSD. Therefor, f(x) is convex

b. $f(x) = e^{u \cdot x}$
$\frac{\partial f}{\partial x_i} = u_i e^{u \cdot x}$
$\frac{\partial 2f}{\partial x_i \partial x_j} = H_{ij} = u_i u_j e^{u \cdot x} - (Eq.1)$

H *can be written as* $H = VV^T$ *where* $V = U \cdot e^{(u \cdot x)/2}$
$z^T VV^T z = (z^T V)(V^T z) = (V^T z)^T (V^T z) = ||V^T z||^2 \geq 0$

c. For f to be convex, prove that: $f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y)$
$Let\ z = \theta x + (1-\theta) y$
$f(z) = \max(g(z), h(z))$ **– *Eq1***
From the above equations, we can say that
$g(z) \leq \theta f(x) + (1-\theta) f(y)\ and\ h(z) \leq \theta f(x) + (1-\theta) f(y)$
Since both g(x) and h(x) are convex,
$g(z) \leq \theta g(x) + (1-\theta) g(y)$ **– *Eq-2***
From Equation 1 and 2:
$g(z) \leq \theta g(x) + (1-\theta) g(y) \leq f(x) + (1-\theta) f(y)$ **– *Eq-3***

Similarly, we can say this about h(z)
$h(z) \leq \theta h(x) + (1-\theta) h(y) \leq f(x) + (1-\theta) f(y)$ **– *Eq-4***

Thus from equation 1, 3, 4 we can say that f(x) is convex

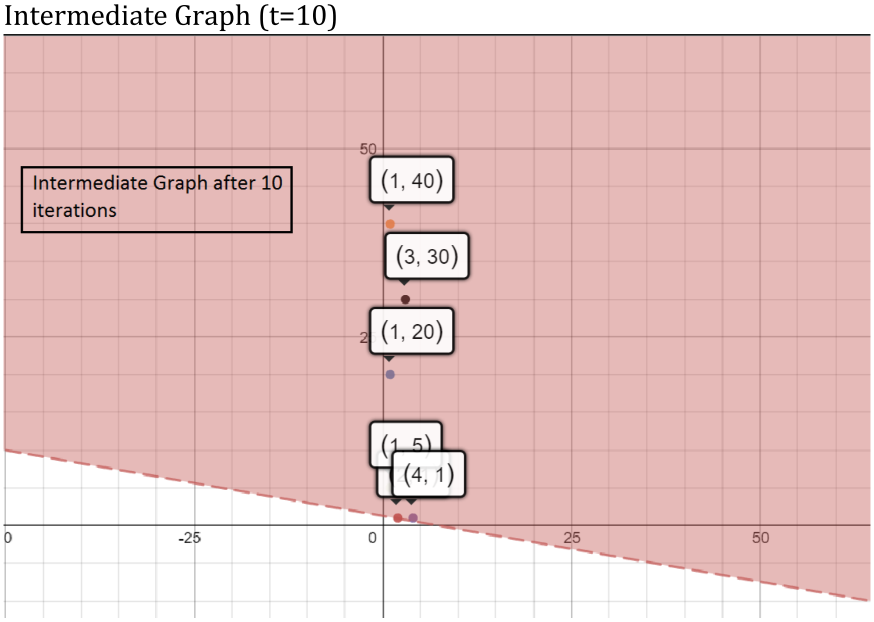
Problem 3:

a. η_t *is obtained by using the backtracking line search algorithm with.*
Limiting condition : $L'(w) < 0.0001$
Update $\eta_t = 0.8\eta_t$ *while the below condition is true*:

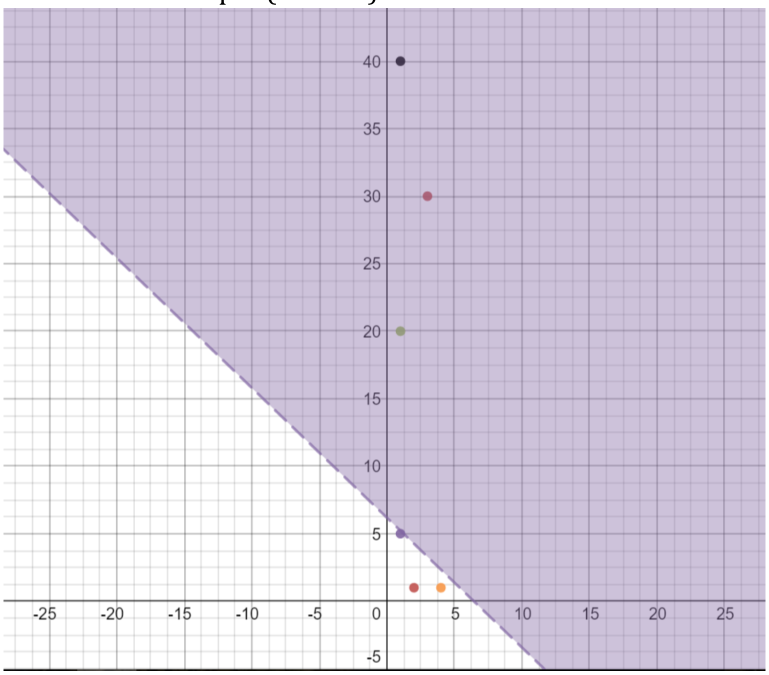
$$\mathrm{L}(\mathrm{w} - \nabla \mathrm{L}(\mathrm{w})) > \mathrm{L}(\mathrm{w}) - \frac{t}{2} ||\nabla \mathrm{L}(\mathrm{w})||^2 ,$$

b. Number of iterations to get a good decision boundary = 22164

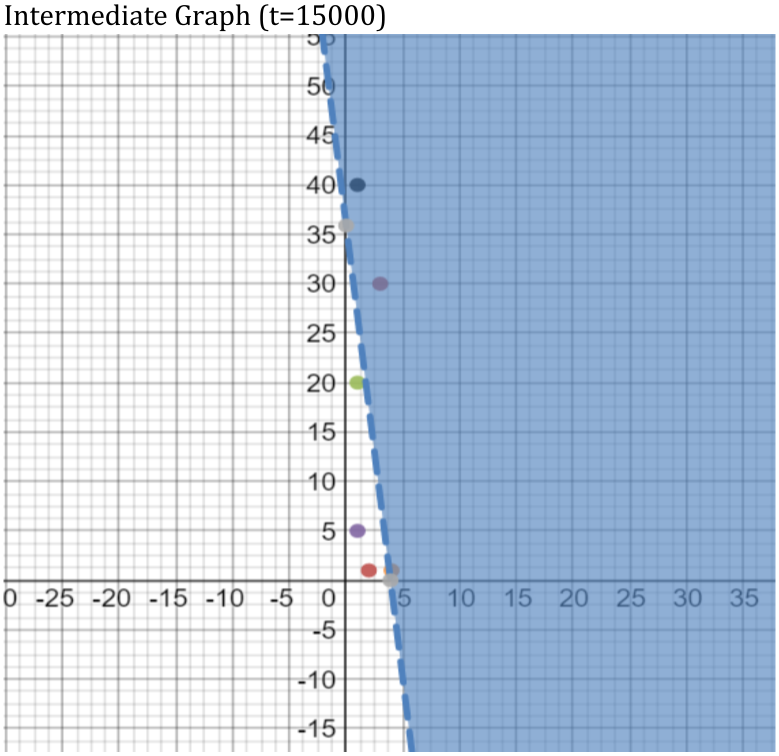
Intermediate Graph (t=10)



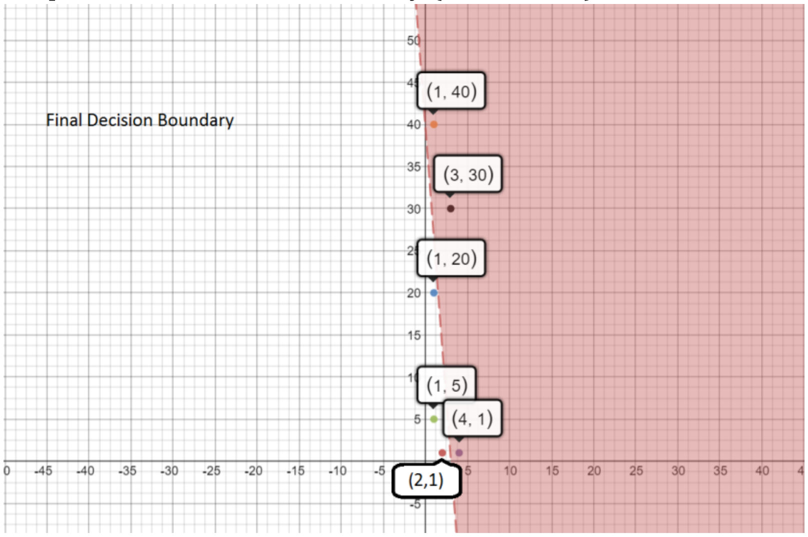
Intermediate Graph: (t=1000)



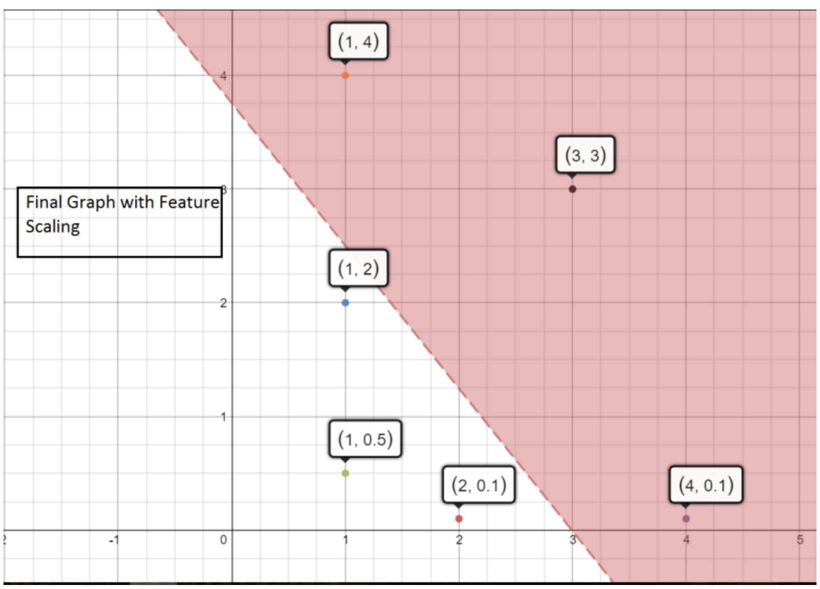
Intermediate Graph (t=15000)



Graph of Final Decision Boundary (iter=22164)



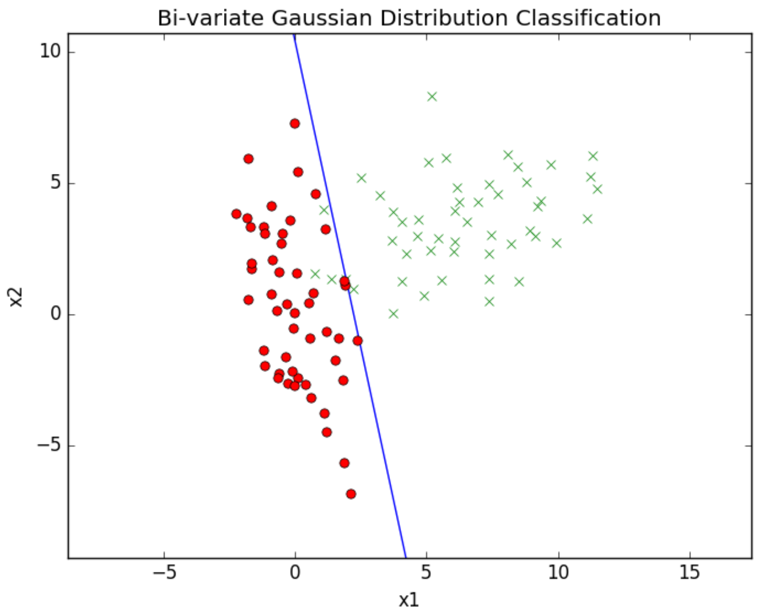
Final Graph by feature scaling:



c. Yes, the number of iterations to get a good decision boundary reduces to 4632

d. Graph with random bivariate gaussians:

Overlapping boundaries: **(iter = 82)**



Non-overlapping boundaries: **(iter = 78)**