P1

(a) The Hessian at w is

$$H = \sum_{x \in S} \lambda + \frac{2}{|S|} x^T x$$

and it is positive semidefinite at any point $w \in \mathbb{R}^d$. So the the objective function is convex by definition. Since, there's no other contraints, the optimization problem is convex.

(b)
```
w = initial(w);
x = data;
y = labels;
gradient(w) = lambda.*w + 2/abs(S) * sum((x(n,:)*w - y(n,:)).*x(n,:));
f(w) = lambda/2*norm(w)^2 + 1/abs(S) * sum((x(n,:)*b - y(n,:))^2);

for i = 1: number of iterations
g = grad(b);
b = b-eta.*g';
fx = f(b);
end
```

(c) The optimization problem is still convex, because $w_i^2 <= 1$ is a convex function

(d) We can rewrite the constraints as follows:

$$w_{2i-1} <= 1$$
$$w_{2i} <= 1$$

Because these two functions are convex, so the intersect of these function is also convex. So the optimization problem is still convex.

(e) The optimization problem is no longer convex, because $w_i^2 = 1$ is not convex. The line segment between thses points is not part the set.

P2

(a)
```
data = [data 1];
b = [initial(b) initial(b0)];
x = data;
y = labels;
gradient = sum((exp(x(n,:)*b))/(1+exp(x(n,:)*b)).*x(n,:) - y(n,:).*x(n,:););
f = sum(ln(1+exp(x(n,:)*b)) - y(n,:).*x(n,:)*b;)
```

```
for i = 1: number of iterations
g = grad(b);
b = b-eta.*g';
fx = f(b);
end
```

(b) Around 4500 iterations are needed to achieve an objective value.

(c) I first seperated the datas based on their labels, I then printed out the 3-D scatter diagrams for both groups of datas. After inspecting the X-Z, Y-Z and X-Y diagrams, I noticed that the label of the data highly depends on the y value of the data. Therefore, I chose a

$$\begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 0.1 \end{bmatrix}$$

as my linear transformation matrix. And now it only requires 209 iterations to achieve the objective value.

(d) Original Data:
Number of iterations: 512
Final objective value: 0.655
Final hold-out error rate: 0.383

Modified Data:
Number of iterations: 32
Final objective value: 0.653
Final hold-out error rate: 0.376