HW 1

Huaiyu Wang ID: 3075870

2.1 (a) 1st derivative: $g(w) = \frac{1}{2}qw^2 + rw + d$

$$g'(w) = qw + r$$

2nd derivative: $g''(w) = q$

(b) $g(w) = -\cos(2\pi w^2) + w^2$

1st derivative: $g'(w) = \sin(2\pi w^2) \cdot 4\pi w + 2w$

$$= 4\pi w \sin(2\pi w^2) + 2w$$

2nd derivative: $g''(w) = 4\pi \sin(2\pi w^2) + 4\pi w \cos(2\pi w^2) \cdot 4\pi w + 2$

$$= 4\pi \sin(2\pi w^2) + 16\pi^2 w^2 \cos(2\pi w^2) + 2$$

(c) $g(w) = \sum_{p=1}^{P} \log(1 + e^{-a_p w})$

$$g'(w) = \sum_{p=1}^{P} \frac{-a_p e^{-a_p w}}{1 + e^{-a_p w}} = \sum_{p=1}^{P} \frac{-a_p}{1 + e^{a_p w}}$$

$$g''(w) = \sum_{p=1}^{P} \frac{a_p e^{a_p w} \cdot a_p}{(1 + e^{a_p w})^2} = \sum_{p=1}^{P} \frac{a_p^2 e^{a_p w}}{(1 + e^{a_p w})^2}$$

2.2 (a) $g(\bar{w}) = \frac{1}{2}\bar{w}^T Q \bar{w} + \bar{r}^T \bar{w} + d$

Gradient $\nabla g(\bar{w}) = \frac{1}{2}(Q + Q^T)\bar{w} + \bar{r}$ ($Q = Q^T$ for Q is symemetric)

$$= Q\bar{w} + \bar{r}$$

Hessian: $\nabla^2 g(\bar{w}) = \frac{1}{2}(Q + Q^T) = Q$

(b) $g(\bar{w}) = -\cos(2\lambda \bar{w}^T\bar{w}) + \bar{w}^T\bar{w}$

$$= -\cos\left(2\lambda \sum_{n=1}^{N} w_n^2\right) + \sum_{n=1}^{N} w_n^2$$

$$\frac{\partial g(\bar{w})}{\partial w_j} = \sin\left(2\lambda \sum_{n=1}^{N} W_n^2\right) \cdot 4\lambda w_j + 2w_j$$

$$\frac{\partial g(\bar{w})}{\partial(w_j w_i)} = \begin{cases} \cos\left(2\lambda \sum_{n=1}^{N} W_n^2\right)(4\lambda)^2 W_i w_j & i \neq j \\ \cos\left(2\lambda \sum_{n=1}^{N} W_n^2\right)(4\lambda)^2 W_i w_j + \sin\left(2\lambda \sum_{n=1}^{N} W_n^2\right) 4\lambda + 2 & i = j \end{cases}$$

$$\therefore \nabla g(\bar{w}) = \sin(2\lambda \bar{w}^T w) \cdot 4\lambda \bar{w} + 2\bar{w}$$

$$\nabla^2 g(\bar{w}) = \cos(2\lambda \bar{w}^T w)(4\lambda^2)\bar{w}\bar{w}^T + \left(4\lambda \sin(2\lambda \bar{w}^T\bar{w}) + 2\right) I_{N\times N}$$

$I_{N\times N}$ is $N\times N$ identity matrix $\begin{pmatrix} 1 & & \\ & 1 \ldots & \\ & & 1 \end{pmatrix}$

c) $g(\bar{w}) = \sum_{p=1}^{P} \log\left(1 + e^{-\bar{a}_p^T\bar{w}}\right)$

$$\frac{\partial g(\bar{w})}{\partial w_j} = \frac{\partial\left(\sum_{p=1}^{P} \log\left(1 + e^{-\sum_{n=1}^{N} a_{pn} w_n}\right)\right)}{\partial w_j} = \sum_{p=1}^{P} \frac{-e^{-\sum_{n=1}^{N} a_{pn} w_n} a_{pj}}{1 + e^{-\sum_{n=1}^{N} a_{pn} w_n}}$$

$$= -\sum_{p=1}^{P} \frac{a_{pj}}{1 + e^{\bar{a}_p^T\bar{w}}}$$

$$\therefore \nabla g(\bar{w}) = -\sum_{p=1}^{P} \frac{\bar{a}_p}{1 + e^{\bar{a}_p^T\bar{w}}}$$

$$\frac{\partial g(\bar{w})}{\partial w_i \partial w_j} = \sum_{p=1}^{P} \frac{a_{pi} e^{\bar{a}_i^T\bar{w}} a_{pi}}{\left(1 + e^{\bar{a}_p^T\bar{w}}\right)^2}$$

$$\therefore \nabla^2 g(\bar{w}) = \sum_{p=1}^{P} \frac{e^{\bar{a}_p^T\bar{w}} \bar{a}_p^T\bar{a}_p}{\left(1 + e^{\bar{a}_p^T\bar{w}}\right)^2}$$

$(h(\bar{w}), \bar{w})$
2.5 all ~~$(\bar{w}, h(\bar{w}))$~~ is on the the hyperplane ~~$(\bar{v}-\bar{w})$~~ $(h(\bar{w})-g(\bar{v}), \bar{w}-\bar{v})$

n should be tangent to all ~~$(\bar{w}-\bar{v}, h(\bar{w})-g(\bar{v}))$~~, denote $\bar{m}$,

~~$\bar{m} \cdot \bar{n} = (\bar{w}-\bar{v}, h(\bar{w})-g(\bar{v})) \begin{pmatrix} 1 \\ -\nabla g(\bar{v}) \end{pmatrix}$~~ $\bar{m} \cdot \bar{n} = (h\bar{w}-g(\bar{v}), \bar{w}-\bar{v}) \begin{pmatrix} 1 \\ -\nabla g(\bar{v}) \end{pmatrix}$

~~$= \bar{w}-\bar{v} - \nabla g(\bar{v})^T (\bar{w}-\bar{v}) \nabla g(\bar{v})$~~ $= \nabla g(\bar{v})^T (\bar{w}-\bar{v}) - (\bar{w}-\bar{v}) \nabla g(\bar{v})$

$= 0$

~~$= \bar{0}$~~

$\therefore$ The normal vector is $\bar{n} = \begin{pmatrix} 1 \\ -\nabla g(\bar{v}) \end{pmatrix}$

2.7 (a) $g(w) = w^2$ $g'(w) = 2w$ $g''(w) = 2 > 0$ $g(w)$ is convex

(b) $g(w) = e^{w^2}$ $g'(w) = e^{w^2} \cdot 2w$ $g''(w) = 4e^{w^2}w^2 + 2e^{w^2} > 0$

$g(w)$ is convex

(c) $g(w) = \log(1+e^w)$

$g'(w) = \dfrac{e^w}{1+e^w}$

$g''(w) = \dfrac{e^w}{(1+e^w)^2} > 0$ $g(w)$ is convex

(d) $g(w) = -\log(w)$ $g'(w) = -\frac{1}{w}$ $g''(w) = \frac{1}{w^2} > 0$

$\therefore g(w)$ is convex

Exercises 2.8 A non-convex function whose only stationary point is a global minimum

a) Use the first order condition to determine the stationary point of $g(w) = w\tanh(w)$ where $\tanh(w)$ is the hyperbolic tangent function. To do this you might find it helpful to graph the first derivative $\frac{\partial}{\partial w} g(w)$ and see where it crosses the w axis. Plot the function to verify that the stationary point you find is the global minimum of the function.

b) Use the second order definition of convexity to show that g is non-convex. *Hint: you can plot the second derivative* $\frac{\partial^2}{\partial w^2} g(w)$.

a) In Matlab:

```
w=-5:0.01:5;

y=w.*tanh(w)

plot(w,y)
```

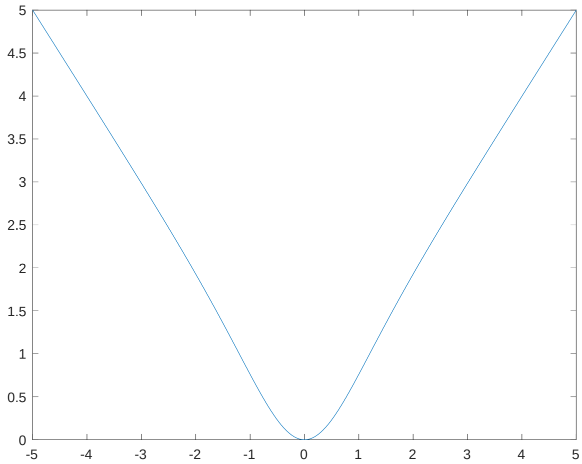

Figure1 $g(w)$

$$g'(w) = \tanh(w) + w(1 - \tanh^2(w))$$

In Matlab:

```
w=-5:0.01:5;

y=tanh(w))+w.*(1-tanh(w).^2)

plot(w,y)
```

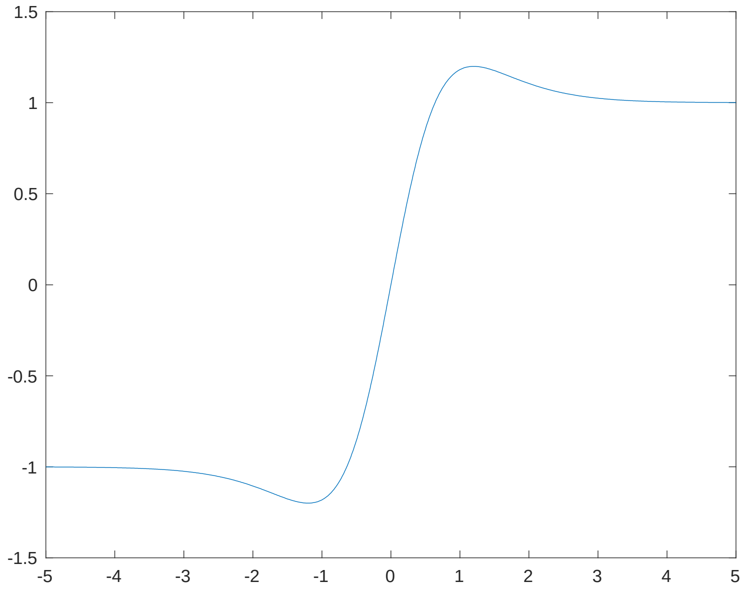

Figure2 $g'(w)$

So we can know $w=0$ is the stationary point, and from Figure1 it is true that the stationary point is the global minimum of the function.

b)

$$g''(w) = 2(1-\tanh^2(w))(1-w\tanh(w))$$

In Matlab:

```
w=-5:0.01:5;

y=2*(1-w.*tanh(w)).*(1-tanh(w).^2)

plot(w,y)
```

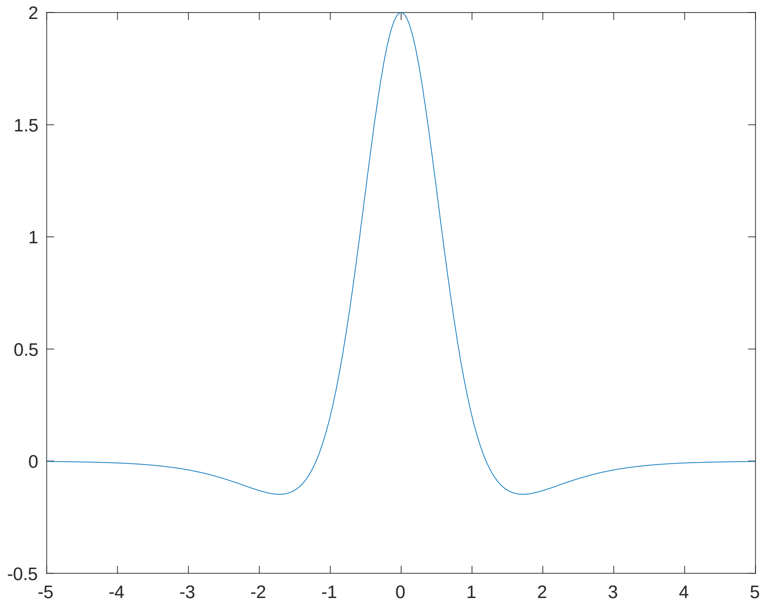

Figure3 Graph of $g''(w)$

From Figure3, $g''(w)$ is sometimes less than 0, so it is non-convex.

`

Exercises 2.13 Code up gradient descent

Only change the following code:

```
grad =4*pi*w*sin(2*pi*(w'*w))+4*w;   %%% PLACE GRADIENT HERE
```

Thus, the result of gradient descent should be:

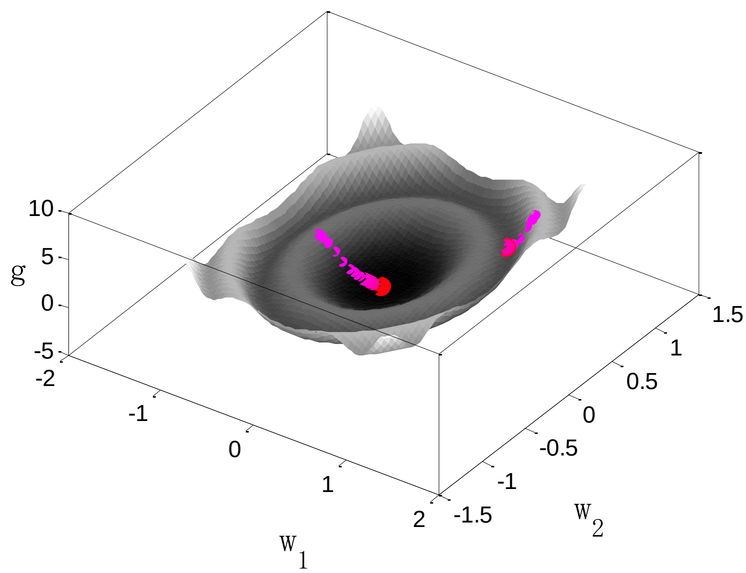

Fig4. Gradient descent

Excercises 2.17 Code up Newton's method

a) The first order condition is:

$$\frac{\alpha g(w)}{\alpha w_i} = \frac{2w_i e^{\sum_{n=1}^{N} w_n^2}}{1+e^{\sum_{n=1}^{N} w_n^2}}$$

So, $\nabla g(w) = \frac{2we^{w^T w}}{1+e^{w^T w}}$

Let $\nabla g(w) = 0$

So, $w=\left[0,0\right]^T$, which is the unique stationary point of the function.

b) The surface plot of the function $g(w)$ is:

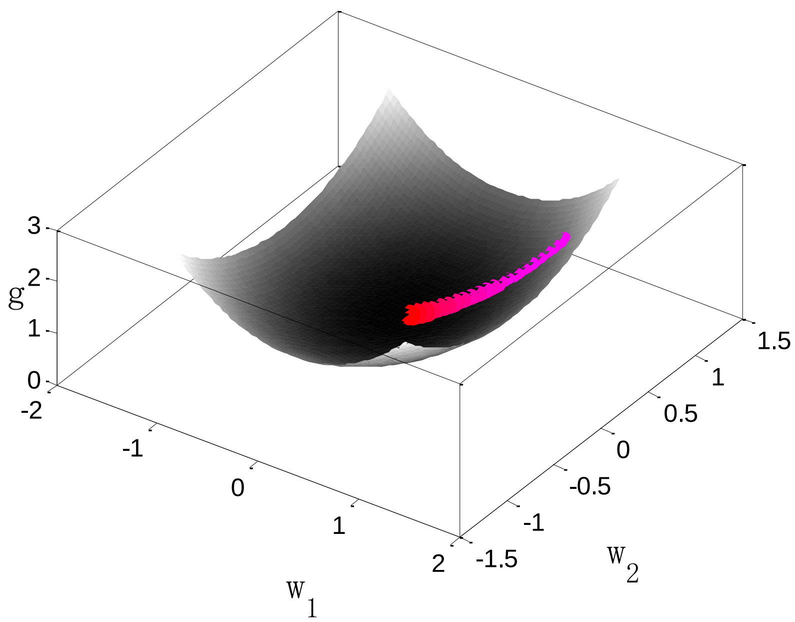

Figure5 the surface plot of $g(w)$

From Figure5, we can see that g (w) is convex, and the stationary point found in part a) is a global minimum.

c) $w^0 = 1_{N\times 1}$

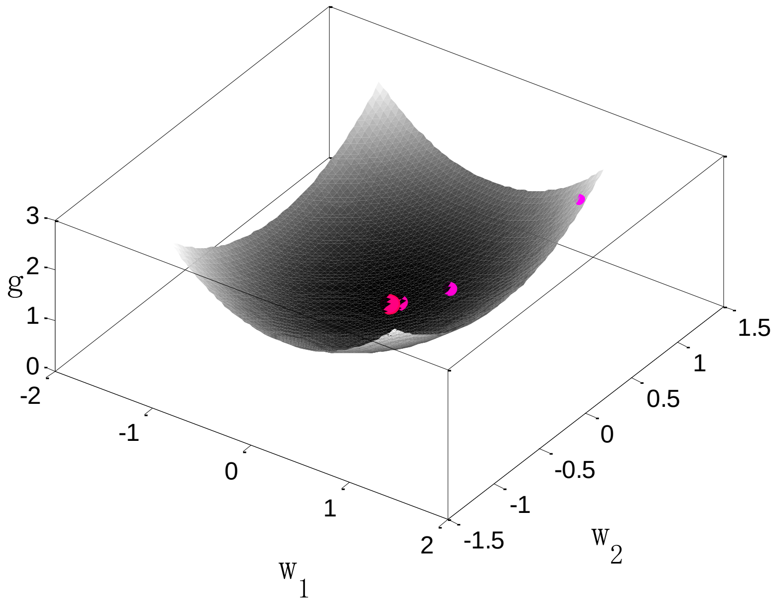

Fig. 6 Using Gradient descent

Process:

$$\frac{\alpha^2 g(w)}{\alpha w_i \alpha w_j} = \begin{cases} \dfrac{4w_i w_j e^{\sum_{n=1}^{N} w_n^2}}{(1+e^{\sum_{n=1}^{N} w_n^2})^2} & i \neq j \\ \dfrac{2e^{\sum_{n=1}^{N} w_n^2}(1+e^{\sum_{n=1}^{N} w_n^2}+2w_i^2)}{(1+e^{\sum_{n=1}^{N} w_n^2})^2} & i = j \end{cases}$$

$$\therefore \quad \nabla^2 g(w) = \frac{4ww^T e^{w^T w} + 2e^{w^T w}(1+e^{w^T w}) \cdot I_{N\times N}}{(1+e^{w^T w})^2}$$

($\mathrm{I}_{N\times N}$ is the $N \times N$ identity matrix)

$\therefore$

$$\frac{\nabla g(w)}{\nabla^2 g(w)} = \frac{\dfrac{2we^{w^T w}}{1+e^{w^T w}}}{\dfrac{4ww^T e^{w^T w} + 2e^{w^T w}(1+e^{w^T w}) \cdot \mathrm{I}_{N\times N}}{(1+e^{w^T w})^2}}$$

$$= \frac{w(1+e^{w^T w})}{2ww^T + (1+e^{w^T w}) \cdot I_{N\times N}}$$

For code, only do the following changes:

```
grad =w*(1+exp((w'*w)))/(1+exp((w'*w))+2*w'*w);   %%% PLACE GRADIENT
HERE
w0=[4 4];
```

d)

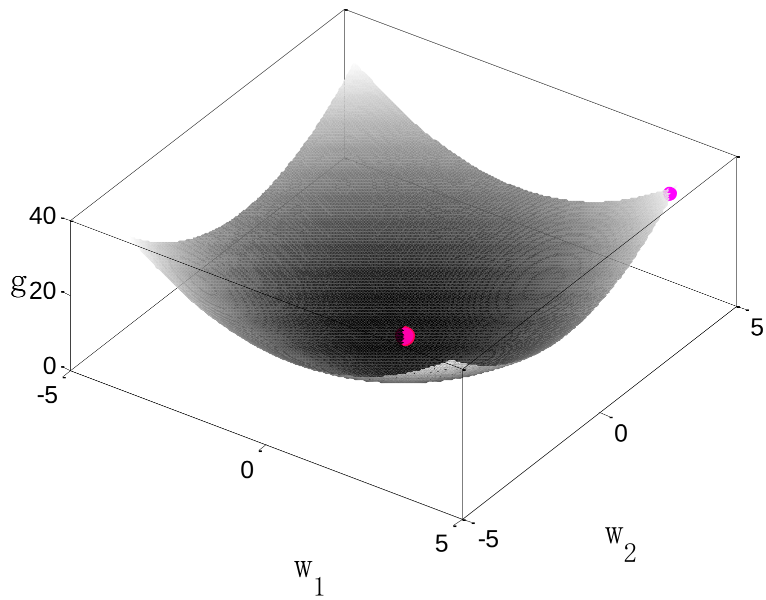

Figure7 Using Gradient descent

Only change the following code:

W0=[4 4];

Explanation:

$$g(w) = \log(1 + e^{w^T w}) \approx w^T w$$

When $w^T w$ is large.

So

$$h(w) = g(w^0) + \nabla g(w^0)^T (w - w^0) + \frac{1}{2}(w - w^0)^T \nabla^2 g(w^0)(w - w^0)$$

$$= w^{0T} w^0 + 2w^{0T}(w - w^0) + \frac{1}{2}(w - w^0)^T \cdot 2(w - w^0)$$

$$= w^T w$$

If $h(w)$=0, w=[0, 0]^T, which is stationary point. Thus, the minimum of the second order Taylor series is the minimum of $g(w)$