2.1

(a)

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

(b)

$$\begin{aligned}
g'(w) &= \sin(2\pi w^2) \cdot 4\pi w + 2w \\
&= 4\pi w \sin(2\pi w^2) + 2w \\
g''(w) &= 4\pi \sin(2\pi w^2) + 4\pi w \cos(2\pi w^2) \cdot 4\pi w + 2 \\
&= 16\pi^2 w^2 \cos(2\pi w^2) + 4\pi \sin(2\pi w^2) + 2
\end{aligned}$$

(c)

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

2.2

(a)

$$\nabla g(\mathbf{w}) = \mathbf{Q}\mathbf{w} + \mathbf{r}.$$

$$\begin{aligned}
\nabla^2 g(\mathbf{w}) &= \frac{1}{2}\left(\mathbf{Q} + \mathbf{Q}^T\right) \\
&= \mathbf{Q}
\end{aligned}$$

(b)

$$\nabla g(\mathbf{w}) = \sin\left(2\pi \mathbf{w}^T \mathbf{w}\right) 4\pi \mathbf{w} + 2\mathbf{w}.$$

$$\nabla^2 g(\mathbf{w}) = \cos\left(2\pi\mathbf{w}^T\mathbf{w}\right)(4\pi)^2\mathbf{w}\mathbf{w}^T + \left[\sin\left(2\pi\mathbf{w}^T\mathbf{w}\right) + 2\right]\mathbf{I}_{N\times N},$$

where $\mathbf{I}_{N\times N}$ is the $N \times N$ identity matrix.

(c)

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

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

2.5

The tangent hyperplane generated by the first order Taylor series approximation centered at a point $\mathbf{v}$ is given as

$$h(\mathbf{w}) = g(\mathbf{v}) + \nabla g(\mathbf{v})^T(\mathbf{w} - \mathbf{v}).$$

Any point $\mathbf{Q} = [h(\mathbf{w})\ \mathbf{w}]^T$ that satisfies the above equation is in the plane. It is easy to see from the equation that point $\mathbf{P} = [g(\mathbf{v})\ \mathbf{v}]^T$ is in the plane. To verify that $\mathbf{n} = [1 - \nabla g(\mathbf{v})]^T$ is the normal vector of the tangent hyperplane, we can prove that for any point $\mathbf{Q}$ in the plane, the inner product of $\mathbf{n}$ and $\mathbf{PQ}$ is always zero. We have

$$\begin{aligned}\mathbf{PQ} &= \mathbf{Q} - \mathbf{P}\\ &= [h(\mathbf{w})\ \mathbf{w}^T]^T - [g(\mathbf{v})\ \mathbf{v}^T]^T\\ &= [h(\mathbf{w}) - g(\mathbf{v})\ (\mathbf{w} - \mathbf{v})^T]^T.\end{aligned}$$

The inner product of $\mathbf{n}$ and $\mathbf{PQ}$ is

$$\mathbf{n}^T\mathbf{PQ} = h(\mathbf{w}) - g(\mathbf{v}) - \nabla g(\mathbf{v})(\mathbf{w} - \mathbf{v}) = 0$$

We see that $\mathbf{n}$ is perpendicular to any vector within the plane, thus $\mathbf{n}$ is the normal vector of the tangent hyperplane.

2.7

(a)

$$g'(w) = 2w,$$
$$g''(w) = 2 > 0.$$

(b)

$$g'(w) = 2we^{w^2},$$
$$g''(w) = 2e^{w^2} + (2w)^2 e^{w^2} > 0.$$

(c)

$$g'(w) = \frac{e^w}{1+e^w},$$
$$g''(w) = \frac{e^w(1+e^w) - (e^w)^2}{(1+e^w)^2}$$
$$= \frac{e^w}{(1+e^w)^2} > 0.$$

(d)

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

2.17

(a)

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

Since the scalar weight $\frac{2e^{\mathbf{w}^T\mathbf{w}}}{1+e^{\mathbf{w}^T\mathbf{w}}} \geq 1$ the only way the equality can occur is when $\mathbf{w} = \mathbf{0}_{N\times 1}$.

(b)

$$\nabla^2 g(\mathbf{w}) = \frac{4e^{\mathbf{w}^T\mathbf{w}}}{\left(1+e^{\mathbf{w}^T\mathbf{w}}\right)^2}\mathbf{w}\mathbf{w}^T + \frac{2e^{\mathbf{w}^T\mathbf{w}}}{1+e^{\mathbf{w}^T\mathbf{w}}}\mathbf{I}_{N\times N}.$$

(d) why faster convergence with initialization point [4,4] ?

Because $\mathbf{w}$ is large, $g(\mathbf{w}) \approx \mathbf{w}^T\mathbf{w}$

And

$$\begin{aligned}
h(\mathbf{w}) &= g(\mathbf{w}^0) + \nabla g(\mathbf{w}^0)^T(\mathbf{w}-\mathbf{w}^0) + \frac{1}{2}(\mathbf{w}-\mathbf{w}^0)^T\nabla^2 g(\mathbf{w}^0)(\mathbf{w}-\mathbf{w}^0) \\
&= \mathbf{w}^{0^T}\mathbf{w}^0 + 2\mathbf{w}^{0^T}(\mathbf{w}-\mathbf{w}^0) + \frac{1}{2}(\mathbf{w}-\mathbf{w}^0)^T \cdot 2(\mathbf{w}-\mathbf{w}^0) \\
&= \mathbf{w}^T\mathbf{w}
\end{aligned}$$

Newton's method optimizes the quadratic function $h(\mathbf{w})$, which is the same as g(w), when w is large. So the optimal solution of Newton's method is the same as optimal point as g(w).