17. $g(\bar{w}) = \log(1+e^{\bar{w}^T\bar{w}})$

$\nabla g(\bar{w}) = \frac{1}{1+e^{\bar{w}^T\bar{w}}} \cdot e^{\bar{w}^T\bar{w}} \cdot 2\bar{w} = \bar{0}.$

$\bar{w} = \bar{0}$

since $\bar{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$, $w_1 = 0$, $w_2 = 0$.

b) See figure $g(\bar{w})$, we could conclude. $g(\bar{w})$ is convex and the stationary point. is a global minimum.

c) d) see figure.

d). Compared with the global minimum of $g(\bar{w})$, $\bar{w}^0 = 4 \cdot 1_{N\times1}$ is a large number

so, $\log(1+e^t) \approx t$. the second order Taylor series approximation is:

$h(\bar{w}) = w_0 + \bar{w} - \bar{w}_0 = \bar{w}$

For Newton method, $\nabla h(\bar{w}) = 0. \rightarrow$ w_0' is the stationary point of $h(\bar{w})$

is also the stationary point of $g(\bar{w})$, So the situation makes sense.

3.3

a) $g(\tilde{w}) = \sum_{p=1}^{P} (\tilde{w}^T\tilde{x}_p\tilde{x}_p^T\tilde{w} - 2\tilde{x}_p^T\tilde{w}y_p + y_p^2) = \frac{1}{2}\tilde{w}^T\left(2\sum_{p=1}^{P}\tilde{x}_p^T\tilde{x}_p\right)\tilde{w} - 2\sum_{p=1}^{P}\tilde{x}_p^T y_p\tilde{w} + \sum_{p=1}^{P} y_p^2$

$\bar{Q} = 2\sum_{p=1}^{P}\tilde{x}_p^T\tilde{x}_p$ $\quad \bar{r} = \left[-2\sum_{p=1}^{P}\tilde{x}_p^T y_p\right]^T = -2\sum_{p=1}^{P}\tilde{x}_p y_p$ $\quad d = \sum_{p=1}^{P} y_p^2$

b) $\nabla g(\tilde{w}) = \sum_{p=1}^{P} 2\tilde{x}_p(\tilde{x}_p^T\tilde{w} - y_p)$ $\quad \nabla^2 g(\tilde{w}) = 2\sum_{p=1}^{P}\tilde{x}_p^T\tilde{x}_p = \bar{Q}$

c) $\nabla g(\tilde{w}) = \bar{Q}\tilde{w} + \bar{r}$ $\quad \nabla^2 g(\tilde{w}) = \bar{Q}$

$\nabla^2 g(\bar{w}^0)\bar{w}^1 = \nabla^2 g(\bar{w}^0)\bar{w}^0 - \nabla g(\bar{w}^0)$

$Q\bar{w}^1 = Q\bar{w}^0 - Q\bar{w}^0 - \bar{r}$

$Q\bar{w}^1 = -r.$

So $\left(\sum_{p=1}^{P}\tilde{x}_p\tilde{x}_p^T\right)\tilde{w} = \sum_{p=1}^{P}\tilde{x}_p y_p.$

3.10.

a) $\sigma(t) = \frac{1}{1+e^{-t}}$ $\quad \sigma^{-1}(\sigma(t)) = \log\left(\frac{\frac{1}{1+e^{-t}}}{1-\frac{1}{1+e^{-t}}}\right) = \log(e^{t}) = t.$

b) $\sigma(b+x_p w) \approx y_p \quad p = 1 \cdots P \quad (3.24)$

$\sigma^{-1}(\sigma(b+x_p w)) = b + x_p w \approx \log\left(\frac{y_p}{1-y_p}\right) \quad p = 1 \cdots P$

c)