Solutions week 2

2.1

Let's split Y and θ into columns

$$\begin{aligned}
Y &= X\theta + \epsilon \\
[Y_1, \ldots, Y_m] &= X[\theta_1, \ldots, \theta_m] + \epsilon \\
[Y_1, \ldots, Y_m] &= [X\theta_1, \ldots, X\theta_m] + \epsilon
\end{aligned}$$

and we see that each column in Y is decided by only one column in θ together with X and one column from the noise. We can then find the optimal theta column for column, and we already have the solution for one column as

$$\begin{aligned}
Y_i &= X\theta_i + \epsilon \\
\hat{\theta}_i &= (X^T X)^{-1} X^T Y_i \\
[\hat{\theta}_1, \ldots, \hat{\theta}_m] &= (X^T X)^{-1} X^T [Y_1, \ldots, Y_m] \\
\hat{\theta} &= (X^T X)^{-1} X^T Y
\end{aligned}$$

2.2

Programming exercise...

2.3

(a)

We want to prove that both ways of writing $\hat{\theta}_{N+1}$ are equivalent, so we set them equal and try to simplify until we can see the equality was true based on the given definitions.

$$
\begin{aligned}
P_{N+1}s_{N+1} &= \hat{\theta}_N + P_{N+1}x_{N+1}(y_{N+1} - x_{N+1}^T\hat{\theta}_N) && (1)\\
s_{N+1} &= P_{N+1}^{-1}\hat{\theta}_N + x_{N+1}y_{N+1} - x_{N+1}x_{N+1}^T\hat{\theta}_N && (2)\\
s_N + x_{N+1}y_{N+1} &= (P_N^{-1} + x_{N+1}x_{N+1}^T)\hat{\theta}_N + x_{N+1}y_{N+1} - x_{N+1}x_{N+1}^T\hat{\theta}_N && (3)\\
s_N &= P_N^{-1}\hat{\theta}_N + x_{N+1}x_{N+1}^T\hat{\theta}_N - x_{N+1}x_{N+1}^T\hat{\theta}_N && (4)\\
P_N s_N &= \hat{\theta}_N && (5)
\end{aligned}
$$

Which we know to be equal from the definition, so the update does indeed do the same thing.

(b)

We want to prove that both ways of writing P_{N+1} are equivalent, so we set them equal and try to simplify until we can see the equality was true.

$$
\begin{aligned}
(P_N^{-1} + x_{N+1}x_{N+1}^T)^{-1} &= P_N - \frac{P_N x_{N+1}x_{N+1}^T P_N}{1 + x_{N+1}^T P_N x_{N+1}}\\
I &= (P_N^{-1} + x_{N+1}x_{N+1}^T)\left(P_N - \frac{P_N x_{N+1}x_{N+1}^T P_N}{1 + x_{N+1}^T P_N x_{N+1}}\right)\\
&= I + x_{N+1}x_{N+1}^T P_N - \frac{x_{N+1}x_{N+1}^T P_N + x_{N+1}x_{N+1}^T P_N x_{N+1}x_{N+1}^T P_N}{1 + x_{N+1}^T P_N x_{N+1}}\\
&= I + x_{N+1}x_{N+1}^T P_N - x_{N+1}\frac{x_{N+1}^T P_N + x_{N+1}^T P_N x_{N+1}x_{N+1}^T P_N}{1 + x_{N+1}^T P_N x_{N+1}}\\
&= I + x_{N+1}x_{N+1}^T P_N - x_{N+1}\frac{1 + x_{N+1}^T P_N x_{N+1}}{1 + x_{N+1}^T P_N x_{N+1}}x_{N+1}^T P_N\\
&= I + x_{N+1}x_{N+1}^T P_N - x_{N+1}x_{N+1}^T P_N\\
&= I
\end{aligned}
$$

2.4

Programming exercise...

2.5

$$
\begin{aligned}
J_{WLS}(\theta) &= \sum_{i=1}^{N}(y_i - \theta^T x_i)^2 w_i^2 \\
&= \sum_{i=1}^{N}(w_i y_i - \theta^T w_i x_i)^2 \\
&= [\hat{y}_i = w_i y_i, \;\; \hat{x}_i = w_i x_i] \\
&= \sum_{i=1}^{N}(\hat{y}_i - \theta^T \hat{x}_i)^2
\end{aligned}
$$

We know the solution for this, it is $\hat{\theta} = (\hat{X}^T \hat{X})^{-1} \hat{X}^T \hat{y}$. Now we insert $\hat{X} = \sqrt{W} X$ and $\hat{Y} = \sqrt{W} Y$ to get

$$
\begin{aligned}
\hat{\theta} &= (X^T \sqrt{W}^T \sqrt{W} X)^{-1} X^T \sqrt{W}^T \sqrt{W} y \\
&= (X^T W X)^{-1} X^T W y
\end{aligned}
$$

2.6

$$
\begin{aligned}
J(\theta) &= \sum_{i=1}^{N} \ln(1 + e^{-y_i \theta^T \mathbf{x}_i}) \\
\frac{\partial J(\theta)}{\partial \theta_j} &= \sum_{i=1}^{N} \frac{1}{1 + e^{-y_i \theta^T \mathbf{x}_i}} e^{-y_i \theta^T \mathbf{x}_i} \cdot (-y_i x_{ij})
\end{aligned}
$$

2.7

Programming exercise...