Solutions

1. **Regularized linear regression.**

   (a) **[3 pts]** The gradient descent update rules for the regularized cost function $\mathcal{E}_{\text{reg}}$ will be of the form:

$$
\begin{aligned}
w_j &\leftarrow w_j - \alpha \frac{\partial \mathcal{E}_{\text{reg}}}{\partial w_j} \\
b &\leftarrow b - \alpha \frac{\partial \mathcal{E}_{\text{reg}}}{\partial b}
\end{aligned}
$$

   Now for the weights w_j we have,

$$
\begin{aligned}
\frac{\partial \mathcal{E}_{\text{reg}}}{\partial w_j} &= \frac{\partial \mathcal{E}}{\partial w_j} + \frac{\partial \mathcal{R}}{\partial w_j} \\
&= \frac{1}{N} \sum_{i=1}^{N} x_j (y^{(i)} - t^{(i)}) + \frac{\partial}{\partial w_j} (\frac{\lambda}{2} \sum_j w_j^2) \\
&= \frac{1}{N} \sum_{i=1}^{N} x_j (y^{(i)} - t^{(i)}) + \lambda w_j.
\end{aligned}
$$

   For the bias term, we have

$$
\begin{aligned}
\frac{\partial \mathcal{E}_{\text{reg}}}{\partial b} &= \frac{\partial \mathcal{E}}{\partial b} + \frac{\partial \mathcal{R}}{\partial b} \\
&= \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - t^{(i)}) + \frac{\partial}{\partial b} (\frac{\lambda}{2} \sum_j w_j^2) \\
&= \frac{1}{N} \sum_{i=1}^{N} (y^{(i)} - t^{(i)}) + 0,
\end{aligned}
$$

which is identicial to the partial derivative without normalization. To put it succinctly:

$$\begin{aligned}
w_j &\leftarrow w_j - \alpha \frac{\partial \mathcal{E}}{\partial w_j} - \alpha \lambda w_j \\
&= (1 - \alpha\lambda) w_j - \alpha \frac{\partial \mathcal{E}}{\partial w_j} \\
b &\leftarrow b - \alpha \frac{\partial \mathcal{E}}{\partial b}
\end{aligned}$$

This form of regularization is sometimes called "weight decay". At each step of gradient descent, weights are pulled towards zero. The further the weights are towards zero, the stronger the force of this "pull". The constant λ also controls the strength of the pull.

(b) **[3 pts]** Recall that we can rewrite

$$\frac{\partial \mathcal{E}}{\partial w_j} = \frac{1}{N} \sum_{j'=1}^{D} \left(\sum_{i=1}^{N} x_j^{(i)} x_{j'}^{(i)} \right) w_{j'} - \frac{1}{N} \sum_{i=1}^{N} x_j^{(i)} t^{(i)}$$

Now,

$$\begin{aligned}
\frac{\partial \mathcal{E}_{\text{reg}}}{\partial w_j} &= \frac{\partial \mathcal{E}}{\partial w_j} + \lambda w_j \\
&= \frac{1}{N} \sum_{j'=1}^{D} \left(\sum_{i=1}^{N} x_j^{(i)} x_{j'}^{(i)} \right) w_{j'} - \frac{1}{N} \sum_{i=1}^{N} x_j^{(i)} t^{(i)} + \lambda w_j \\
&= \sum_{j'=1}^{D} \left[\left(\frac{1}{N} \sum_{i=1}^{N} x_j^{(i)} x_{j'}^{(i)} \right) + \delta_{j,j'} \lambda \right] w_{j'} - \frac{1}{N} \sum_{i=1}^{N} x_j^{(i)} t^{(i)}
\end{aligned}$$

Where

$$\delta_{j,j'} = \begin{cases} 1, & \text{if } j = j' \\ 0, & \text{otherwise} \end{cases}$$

Now, the values of $A_{jj'}$ and c_j are:

$$A_{jj'} = \frac{1}{N}\left(\sum_{i=1}^{N} x_j^{(i)} x_{j'}^{(i)}\right) + \delta_{j,j'}\lambda$$

$$c_j = \frac{1}{N}\sum_{i=1}^{N} x_j^{(i)} t^{(i)}$$

2. **Visualizing the cost function.**

   (a) **[2pts]** We begin with the cost function

$$\begin{aligned}
\mathcal{E} &= \frac{1}{6}\left((2w_1-1)^2 + (w_2-2)^2 + (w_2-0)^2\right) \\
&= \frac{1}{6}\left(4(w_1-\frac{1}{2})^2 + (w_2^2 - 4w_2 + 4 + w_2^2)\right) \\
&= \frac{2}{3}(w_1-\frac{1}{2})^2 + \frac{2}{6}(w_2^2 - 2w_2 + 2) \\
&= \frac{2}{3}(w_1-\frac{1}{2})^2 + \frac{1}{3}(w_2^2 - 2w_2 + 1 + 1) \\
&= \frac{2}{3}(w_1-\frac{1}{2})^2 + \frac{1}{3}(w_2-1)^2 + \frac{1}{3}
\end{aligned}$$

   (b) **[2pts]** For $\mathcal{E} = 1$, the equation of the ellipse is:

$$\frac{2}{3}(w_1-\frac{1}{2})^2 + \frac{1}{3}(w_2-1)^2 + \frac{1}{3} = 1$$

   or simplifying:

$$2(w_1-\frac{1}{2})^2 + (w_2-1)^2 = 2$$

   Figure 1 shows the plotted ellipse. The center is at (0.5, 1), and the major and minor radius are $\sqrt{2}$ and 1, respectively.

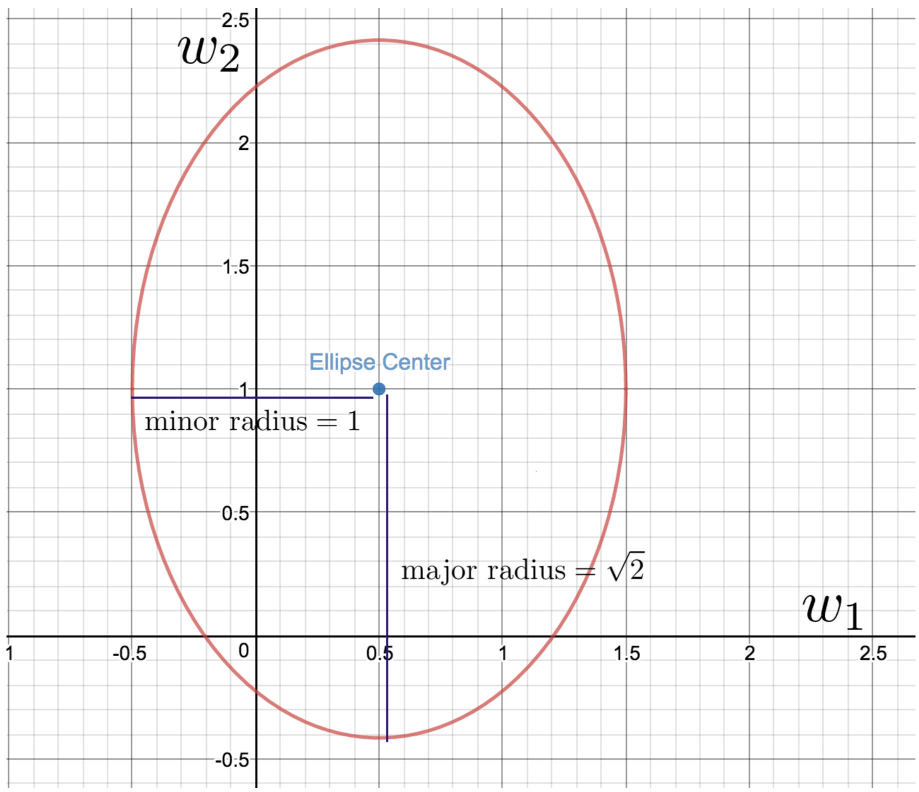

Figure 1: Ellipse plot.