CSC 321

HW 1

1. (a) $y^{(i)} = W^T X^{(i)} + b = \sum_{j=1}^{D} w_j x_j^{(i)} + b$

$$\frac{\partial \mathcal{E}_{reg}}{\partial w_j} = \frac{\partial}{\partial w_j}\left(\frac{1}{2N}\sum_{i=1}^{N}\left(\sum_{j'=1}^{D} w_{j'} x_{j'}^{(i)} + b - t^{(i)}\right)^2 + \frac{\lambda}{2}\sum_{j'=1}^{D} w_{j'}^2\right)$$

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

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

$$w_j \leftarrow w_j - \alpha\left(\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j'=1}^{D} w_{j'} x_{j'}^{(i)} + b - t^{(i)}\right) x_j^{(i)} + \lambda w_j\right)$$
$$b \leftarrow b - \alpha\left(\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j'=1}^{D} w_{j'} x_{j'}^{(i)} + b - t^{(i)}\right)\right)$$

because the regularization penalize large weights using the regularizer $\frac{\lambda}{2}\sum_{j=1}^{D} w_j^2$. This prevent our generalization from overfitting. It is called weight decay because it subtracts a term that is proportional to w_j and cause w_j to decay.

(b) $$\frac{\partial \mathcal{E}_{reg}}{\partial w_j} = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j'=1}^{D} w_{j'} x_{j'}^{(i)} - t^{(i)}\right) x_j^{(i)} + \lambda w_j = 0$$

$$\Rightarrow \frac{1}{N}\sum_{i=1}^{N}\sum_{j'=1}^{D} w_{j'} x_{j'}^{(i)} x_j^{(i)} - \frac{1}{N}\sum_{i=1}^{N} t^{(i)} x_j^{(i)} + \lambda w_j = 0$$

$$\Rightarrow \frac{1}{N}\sum_{i=1}^{N}\sum_{j'=1}^{D} x_j^{(i)} x_{j'}^{(i)} w_{j'} - \left(\frac{1}{N}\sum_{i=1}^{N} t^{(i)} x_j^{(i)} - \lambda w_j\right) = 0$$

when $j = j'$ $\quad A_{jj'} = \frac{1}{N}\sum_{i=1}^{N} x_j^{(i)} x_{j'}^{(i)} + \lambda \qquad C_j = \frac{1}{N}\sum_{i=1}^{N} x_j^{(i)} t^{(i)}$

when $j \neq j'$ $\quad A_{jj'} = \frac{1}{N}\sum_{i=1}^{N} x_j^{(i)} x_{j'}^{(i)}$

2. (a) $$\mathcal{E}(w_1, w_2) = \frac{1}{2N}\sum_{i=1}^{N}\left(y^{(i)} - t^{(i)}\right)^2$$

$$= \frac{1}{6}\left[(2w_1 - 1)^2 + (w_2 - 2)^2 + (w_2)^2\right]$$
$$= \frac{2}{3}\left(w_1 - \frac{1}{2}\right)^2 + \frac{1}{6}\left(w_2^2 + 4 - 4w_2 + w_2^2\right)$$
$$= \frac{2}{3}\left(w_1 - \frac{1}{2}\right)^2 + \frac{1}{3}\left(w_2^2 - 2w_2 + 2\right)$$
$$= \frac{2}{3}\left(w_1 - \frac{1}{2}\right)^2 + \frac{1}{3}(w_2 - 1)^2 + \frac{1}{3}$$

$$\mathcal{E}(w_1, w_2) = 1 \Rightarrow \frac{2}{3}\left(w_1 - \frac{1}{2}\right)^2 + \frac{1}{3}(w_2 - 1)^2 + \frac{1}{3} = 1$$
$$\Rightarrow \left(w_1 - \frac{1}{2}\right)^2 + \frac{1}{2}(w_2 - 1)^2 = 1$$

the center of the ellipse is $(\frac{1}{2}, 1)$
the major radii is $\sqrt{2}$
the minor radii is 1