Q2

1):

let $\sqrt{A} = \begin{bmatrix} \sqrt{a^{(1)}} & & 0 \\ & \ddots & \\ 0 & & \sqrt{a^{(N)}} \end{bmatrix}$

Then $L(w) = \frac{1}{2}\sum_{i=1}^{N} a^{(i)} (y^{(i)} - w^T x^{(i)})^2 + \frac{\lambda}{2}\|w\|^2$

$= \frac{1}{2}[\sqrt{A}(y - Xw)]^T [\sqrt{A}(y - Xw)] + \frac{\lambda I}{2} w^T w$

$= \frac{1}{2}\left[(\sqrt{A}y)^T \sqrt{A}y - 2(\sqrt{A}Xw)^T \sqrt{A}y + (\sqrt{A}Xw)^T \sqrt{A}Xw + \frac{\lambda I}{2} w^T w\right]$

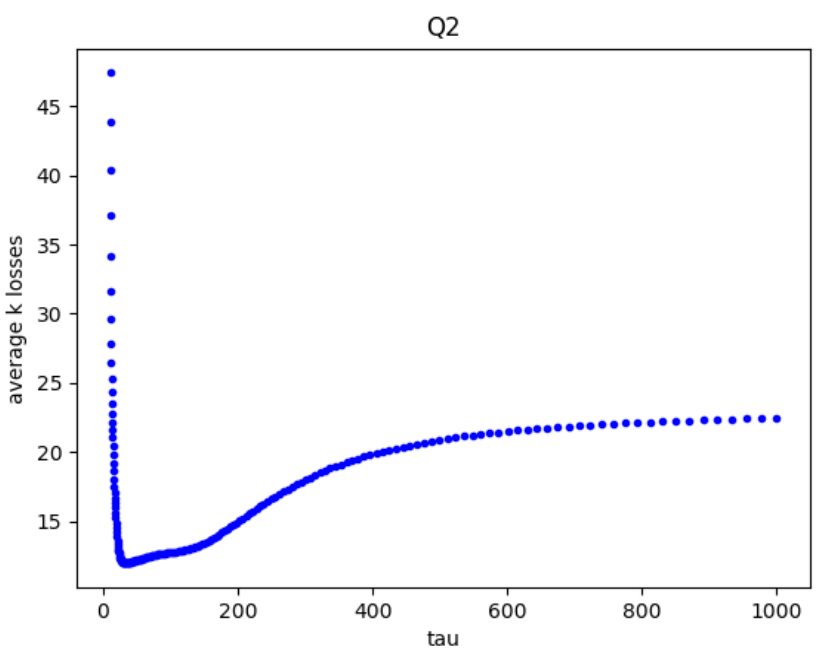
$= \frac{1}{2}\left[y^T A y - 2(Xw)^T A y + (Xw)^T A Xw + \frac{\lambda I}{2} w^T w\right]$

Then $\nabla L(w) = -X^T A y + X^T A X w + \lambda I w$

Then $\nabla L(w^*) = \vec{0}$ $\Rightarrow (X^T A X + \lambda I) w^* = X^T A y$

$\Rightarrow w^* = (X^T A X + \lambda I)^{-1} X^T A y$

2): plot



4): when $\tau \to 0$

losses $\to \infty$; Algorithm performs badly

when $\tau \to \infty$

losses converges around 21; Algorithm performs ok.