Problem 1

Give a closed-form solution to the loss

$$J = \sum_{m=1}^{M} \left(\sum_{i=0}^{d} w_i x_i^{(m)} - t^{(m)} \right)^2 + \sum_{i=0}^{d} w_i^2$$

Problem 2

In the gradient descent algorithm, $\alpha > 0$ is the learning rate. If α is small enough, then the function value guarantees to decrease. In practice, we may anneal α, meaning that we start from a relatively large α, but decrease it gradually.

Show that α cannot be decreased too fast. If α is decreased too fast, even if it is strictly positive, the gradient descent algorithm may not converge to the optimum of a convex function.

Hint: Show a concrete loss and an annealing scheduler such that the gradient descent algorithm fails to converge to the optimum.

Another Hint: Think of the schema of our attendance bonus in this course. Why can't a student get more than five marks even if the student catches infinite errors?

END of W3