Problem 1

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

$$= \|Xw - t\|^2 + w^T w$$

From assignment 2 Q6, we know

$$\nabla_w J(w) = 2X^T Xw - 2X^T t + 2w$$

Set $\nabla_w J(w) = 0$

$$2X^T Xw - 2X^T t + 2w = 0$$

$$2X^T Xw + 2w = 2X^T t$$

$$(2X^T X + 2)w = 2X^T t$$

$$(X^T X + 1)w = X^T t$$

$$w = (X^T X + 1)^{-1} X^T t$$

Problem 2.

Take the schema of our attendance bonus in this course as a example, student cannot get more than 5 marks even if the student catches infinite errors because the errors we catches later may correspond to less bonus marks than the first few errors we catches due to make sure the total score cannot exceed 5 marks.

For example, student can get

2 marks for the first time, $5-2=3$ marks to get.

x0.75 ↓ 1.5 mark for the second time, $5-2-1.5=1.5$ marks to get.

x0.47 ↓ 0.7 mark for the third time, $5-2-1.5-0.7=0.8$ marks to get.

...

not decrease too fast

In this example, we start from a relatively large, but decrease it gradually. If λ decreased too fast, like from 0.5 to 0.01, the algorithm may not converge to the optimum of a convex function (may not converge to around 0 mark to get or converge to around 5 marks obtained in total).

Thus, λ cannot be decreased too fast.