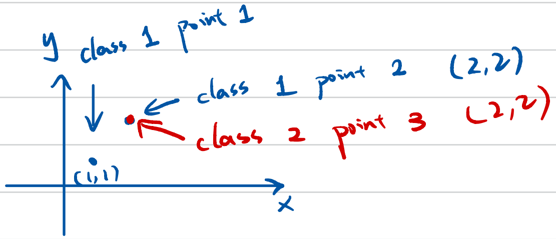
Q1

1.1

Class 2, because the point (14, 4) in class 2 has the smallest L2 distance to the target point (14, 3)

1.2



In the above case, 1-NN is suppose to return itself as its closest neighbour (which for point 1 1-NN is itself) due to distance = 0. However, two points from different class has an overlap in the x, y coordinate, which makes possible for point 2 regard point 3 as its closest neighbour, and vice versa. Hence the training error can be non-zero.

Q2

2.1 $L = \frac{1}{n}\sum_{i}^{n}\left[y^i - (\theta_0 + \theta_1 x^i)\right]^2 = \frac{1}{n}\sum_{i}^{n}(y^i - \theta_0 - \theta_1 x^i)^2$

2.2 $\frac{\partial L}{\partial \theta_0} = -\frac{2}{n}\sum_{i}^{n}(y^i - \theta_0 - \theta_1 x^i)$

2.3 $\frac{\partial L}{\partial \theta_1} = -\frac{2}{n}\sum_{i}^{n} x_i (y^i - \theta_0 - \theta_1 x^i)$

2.4 $\theta_0 \mathrel{-}= \alpha * \left(-\frac{2}{n}\sum_{i}^{n}(y^i - \theta_0 - \theta_1 x^i)\right)$

$\theta_1 \mathrel{-}= \alpha * \left(-\frac{2}{n}\sum_{i}^{n} x_i (y^i - \theta_0 - \theta_1 x^i)\right)$

Q3

3.1 $L = \frac{1}{n}\sum_{i}^{n}\left[y^i - (\theta_0 + \theta_1 x^i)\right]^2 + \lambda|\theta_1|$

3.2 $\frac{\partial L}{\partial \theta_0} = -\frac{2}{n}\sum_{i}^{n}(y^i - \theta_0 - \theta_1 x^i)$

3.3 $\frac{\partial L}{\partial \theta_1} = \begin{cases} -\frac{2}{n}\sum_{i}^{n} x_i(y^i - \theta_0 - \theta_1 x^i) + \lambda, & \theta > 0 \\ -\frac{2}{n}\sum_{i}^{n} x_i(y^i - \theta_0 - \theta_1 x^i), & \theta = 0 \\ -\frac{2}{n}\sum_{i}^{n} x_i(y^i - \theta_0 - \theta_1 x^i) - \lambda, & \theta < 0 \end{cases}$

3.4 $\theta_0 \mathrel{-}= \alpha * \left(-\frac{2}{n}\sum_{i}^{n}(y^i - \theta_0 - \theta_1 x^i)\right)$

if $\theta_1 > 0$:

$\theta_1 \mathrel{-}= \alpha * \left(-\frac{2}{n}\sum_{i}^{n} x_i(y^i - \theta_0 - \theta_1 x^i) + \lambda\right)$

else if $\theta_1 = 0$

$\theta_1 \mathrel{-}= \alpha * \left(-\frac{2}{n}\sum_{i}^{n} x_i(y^i - \theta_0 - \theta_1 x^i)\right)$

else: $[\theta_1 < 0]$

$\theta_1 \mathrel{-}= \alpha * \left(-\frac{2}{n}\sum_{i}^{n} x_i(y^i - \theta_0 - \theta_1 x^i) - \lambda\right)$

Q4

4.1 $$L = \frac{1}{n}\sum_{i}^{n}\left[y^i - (\theta_0 + \theta_1 x^i)\right]^2 + \lambda\theta_1^2$$

4.2 $$\frac{\partial L}{\partial \theta_0} = -\frac{2}{n}\sum_{i}^{n}(y^i - \theta_0 - \theta_1 x^i)$$

4.3 $$\frac{\partial L}{\partial \theta_1} = -\frac{2}{n}\sum_{i}^{n} x_i (y^i - \theta_0 - \theta_1 x^i) + 2\lambda\theta_1$$

4.4 $$\theta_0 -= \alpha * -\frac{2}{n}\sum_{i}^{n}(y^i - \theta_0 - \theta_1 x^i)$$

$$\theta_1 -= \alpha * -\frac{2}{n}\sum_{i}^{n} x_i (y^i - \theta_0 - \theta_1 x^i) + 2\lambda\theta_1$$

Q6

1. $\nabla L(W) = -\frac{2}{n}\sum_{i}^{n} x^{iT} \cdot \left(y^i - (x^i \cdot w)\right) + 2\lambda w$

or alternative form, $\nabla L(w) = \left(-\frac{2}{n} X^T \cdot (y - X \cdot w)\right) + 2\lambda w$

2. $\varepsilon = 0.001 \leftarrow$ error tolerance

$\eta = 0.1 \leftarrow$ step size

$\varepsilon' = w - \eta \cdot \nabla L(w)$

while $\varepsilon' > \varepsilon$:

$\varepsilon' = w - \eta \cdot \nabla L(w)$

$w -= \eta \cdot \nabla L(w)$