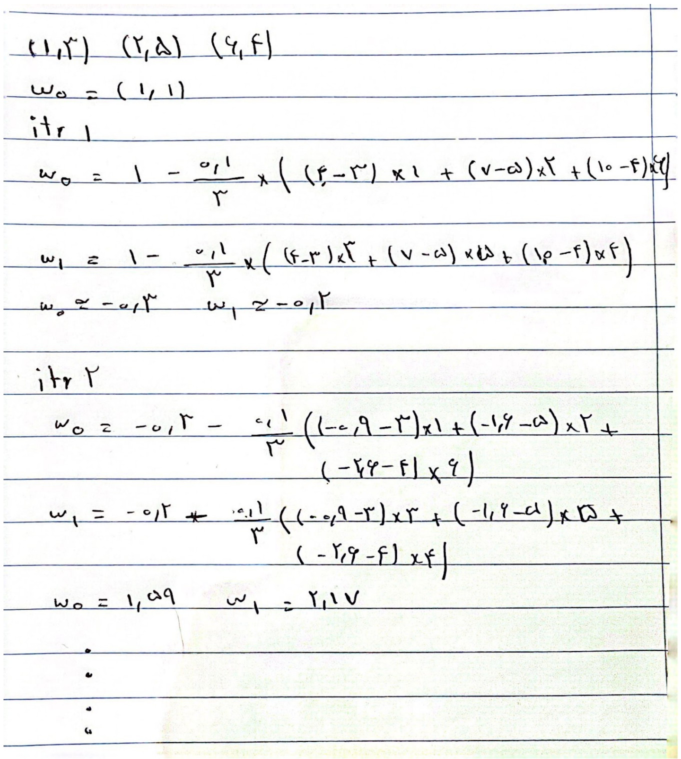
HW1 - Arian Jalalian - 99243026

1. Because it kind of represents the distance between each predicted and actual point (which has squaring in it), and calculating power 4 is computationally more expensive.
2. Because performing matrix operations is slower than gradient descent in a large dataset, so it is not practical.
3. It is better to use polynomial regression and use area (pi * r ^ 2) as a feature with more meaning and relation to pizza price. Also it is possible to have 2 pizzas with the same radius but different width so the thicker one is more expensive and that shows that the relation is not linear.
   So we can use polynomial regression with m = 3 to have 1, r, r^2, r^3 as features that represent radius, area and volume.
4. Not particularly, cost can oscillate over minimum and gets further from it in each step.
5.



(۱,۳) (۲,۵) (۶,۴)

$w_0 = (۱,۱)$

itr ۱

$w_0 = ۱ - \frac{۰,۱}{۳} \times ((۴-۳) \times ۱ + (۷-۵) \times ۲ + (۱۰-۴) \times ۶)$

$w_1 = ۱ - \frac{۰,۱}{۳} \times ((۴-۳) \times ۲ + (۷-۵) \times ۵ + (۱۰-۴) \times ۴)$

$w_0 \simeq -۰,۳ \quad w_1 \simeq -۰,۲$

itr ۲

$w_0 = -۰,۳ - \frac{۰,۱}{۳} ((-۰,۹-۳) \times ۱ + (-۱,۶-۵) \times ۲ + (-۲,۶-۴) \times ۶)$

$w_1 = -۰,۲ - \frac{۰,۱}{۳} ((-۰,۹-۳) \times ۳ + (-۱,۶-۵) \times ۵ + (-۲,۶-۴) \times ۴)$

$w_0 = ۱,۵۹ \quad w_1 = ۲,۱۷$

⋮

Final w is [3.776, 0.071]