

Problem Set 2 Report

Prashanth Kumaran

Student ID: 4549673

Problem 1: Cost Function

Manual Cost Computation:

$(x_1^{(i)}, x_2^{(i)}, y^{(i)}) : (0, 1, 1.5), (1, 1.5, 4), (2, 4, 8.5), (3, 2, 8.5)$

$$(i) \quad \theta = \begin{bmatrix} 0.5 \\ 2 \\ 1 \end{bmatrix}$$

$$\begin{aligned} J(\theta) &= \frac{1}{2(4)} \sum_{i=1}^4 (0.5 + 2(x_1^{(i)}) + x_2^{(i)} - y^{(i)})^2 \\ &= \frac{1}{8} ((0.5 + 2(0) + 1 - 1.5)^2 + (0.5 + 2(1) + 1.5 - 4)^2 + (0.5 + 2(2) + 4 - 8.5)^2 + (0.5 + 2(3) + 2 - 8.5)^2) \\ &= \frac{1}{8} (0 + 0 + 0 + 0) = \boxed{0} \end{aligned}$$

$$(ii) \quad \theta = \begin{bmatrix} 3 \\ -1.5 \\ -4 \end{bmatrix}$$

$$\begin{aligned} J(\theta) &= \frac{1}{2(4)} \sum_{i=1}^4 (3 + -1.5(x_1^{(i)}) + -4(x_2^{(i)}) - y^{(i)})^2 \\ &= \frac{1}{8} ((3 - 1.5(0) - 4(1) - 1.5)^2 + (3 - 1.5(1) - 4(1.5) - 4)^2 + (3 - 1.5(2) - 4(4) - 8.5)^2 + (3 - 1.5(3) - 4(2) - 8.5)^2) \\ &= \frac{1}{8} (6.25 + 72.25 + 600.25 + 324) = \boxed{125.375} \end{aligned}$$

$$(iii) \quad \theta = \begin{bmatrix} 0.5 \\ 1 \\ 2 \end{bmatrix}$$

$$\begin{aligned} J(\theta) &= \frac{1}{2(4)} \sum_{i=1}^4 (0.5 + 1(x_1^{(i)}) + 2(x_2^{(i)}) - y^{(i)})^2 \\ &= \frac{1}{8} ((0.5 + 1(0) + 2(1) - 1.5)^2 + (0.5 + 1(1) + 2(1.5) - 4)^2 + (0.5 + 1(2) + 2(4) - 8.5)^2 + (0.5 + 1(3) + 2(2) - 8.5)^2) \\ &= \frac{1}{8} (1 + 0.25 + 4 + 1) = \boxed{0.78125} \end{aligned}$$

Function Outputs:

(i) $J(\theta)$: 0.0

(ii) $J(\theta)$: 125.34375

(iii) $J(\theta)$: 0.78125

Problem 2: Gradient Descent

Text output: Theta: $\begin{bmatrix} 0.62859468 & 0.85084562 & 0.50626797 \end{bmatrix}$ Final cost: 4.869695081503297

Problem 3: Normal Equation

Text output: Theta: $\begin{bmatrix} 0.5 & 2 & 1 \end{bmatrix}$. There is a significant difference in the thetas between problems 2 and 3 because the gradient descent algorithm requires tuning of alpha and the number of iterations. In order for the gradient descent algorithm to match the normal equation, alpha must be tuned and the number of iterations must be increased to get closer to the true minimum cost.

Problem 4: Linear Regression with One Variable

(a) In Code

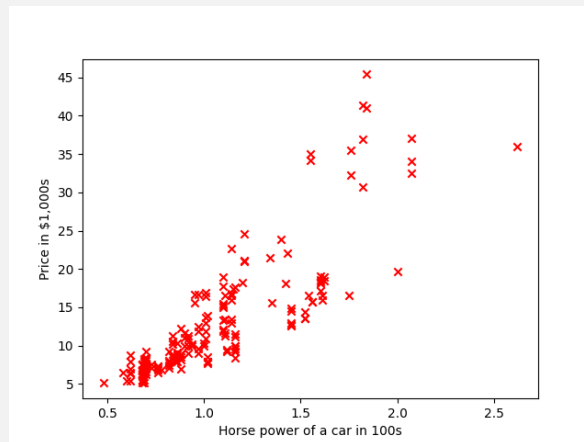


Figure 1: ps2-4-b.png

(b)

(c) Text Output: X size: (179, 2) y size: (179, 1)

(d) In Code

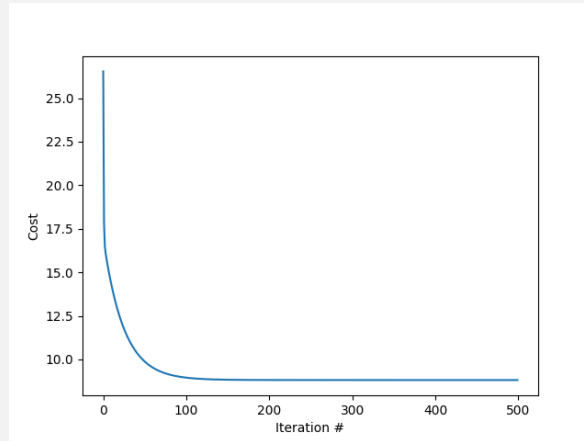


Figure 2: ps2-4-e.png

(e) Theta: $\begin{bmatrix} -5.41558526 \\ 17.39639326 \end{bmatrix}$

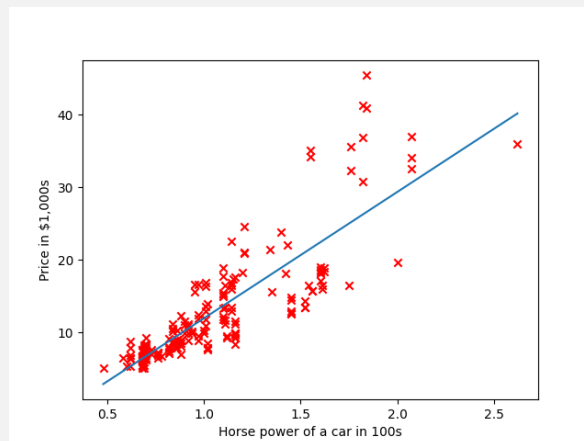


Figure 3: ps2-4-f.png

(f)

(g) Prediction Error: 5.806342949496337

(h) Prediction Error: 5.034370887068474, The Prediction error for both the normal equation and gradient descent match very closely.

(i) The plots for $\alpha = 0.001$ and 0.003 have a very shallow curve, which takes a long time to converge. $\alpha = 0.03$ decreases much quicker and converges around iteration 50. $\alpha = 3$ increases rapidly, which means that the learning rate is too large.

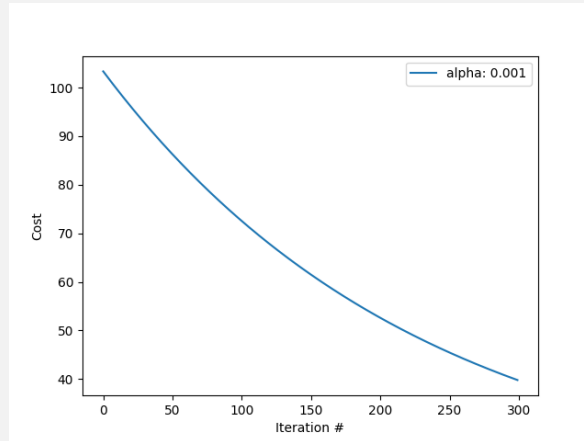


Figure 4: ps2-4-i-1.png

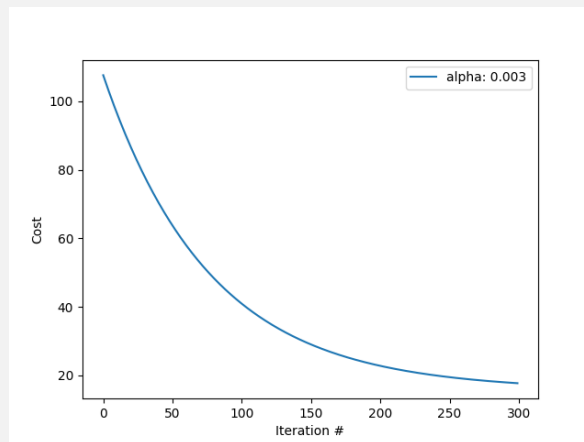


Figure 5: ps2-4-i-2.png

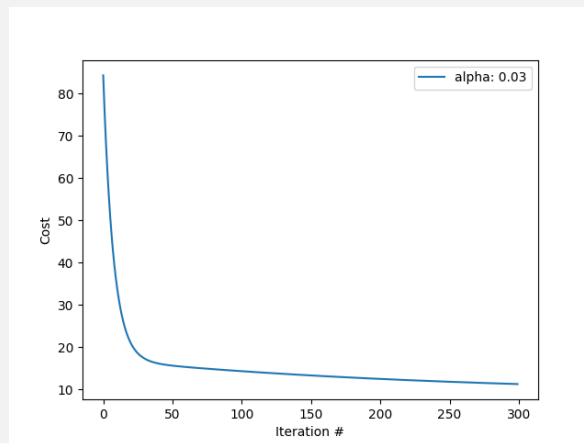


Figure 6: ps2-4-i-3.png

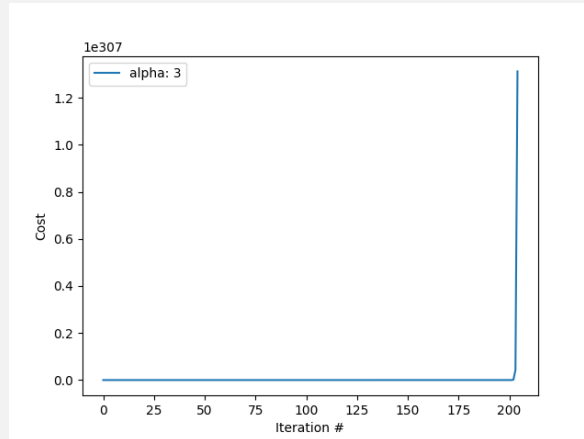


Figure 7: ps2-4-i-4.png

Problem 5: Linear Regression with Multiple Variables

- (a) Text Output: x1 mean: 1611.1111111111111 x2 mean: 1292.2777777777778 y mean: 102.02777777777777
 x1 standard deviation: 383.53456875762674 x2 standard deviation: 238.73737443185826
 y standard deviation: 7.350306535977905
 X size: (36, 3) y size: (36, 1)

- (b) Text Output: Theta = $\begin{bmatrix} 2.50484977e-04 & 3.76591635e-01 & 2.75935447e-01 \end{bmatrix}$

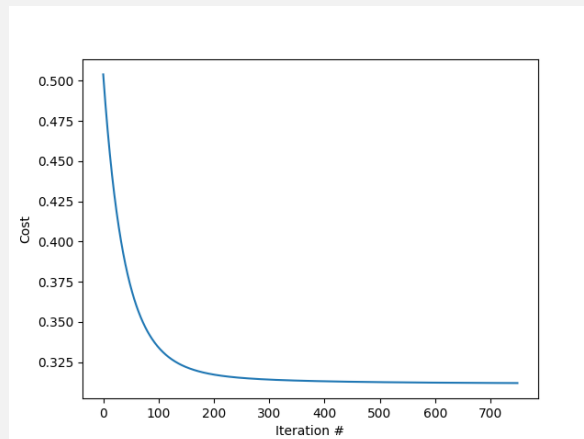


Figure 8: ps2-5-b.png

- (c) Text Output: Prediction: 104.77409875163221