Problem 1: Write down the gradient descent update rule for the objective function

$$J(\theta) = \theta^4 - 6\theta^3 + 11\theta^2 - 7\theta + 4$$

from class. Suppose the learning rate is α.

The update rule is

$$\theta := \theta - \alpha J'(\theta) = \theta - \alpha(4\theta^3 - 18\theta^2 + 22\theta - 7).$$

The recurrence relation is

$$\theta_{t+1} = \theta_t - \alpha J'(\theta_t) = \theta_t - \alpha(4\theta_t^3 - 18\theta_t^2 + 22\theta_t - 7),$$

for $t \geq 0$.

Problem 2: Consider the affine objective function

$$J(\theta) = m\theta + b,$$

for some parameters $m \neq 0$ and $b \in \mathbb{R}$.

(a) Write down the gradient descent update rule. Suppose the learning rate is α.

The update rule is

$$\theta := \theta - \alpha m.$$

The recurrence relation is

$$\theta_{t+1} = \theta_t - \alpha m,$$

for $t \geq 0$.

(b) Find a closed form expression for θ_t.

We have

$$\theta_t = \theta_0 - \alpha m t$$

for all $t \geq 1$.

(c) Using your answer to (b), discuss convergence of the gradient descent algorithm.

We have $\theta_t \to -\infty$ as $t \to \infty$.

Problem 3: Consider the quadratic objective function

$$J(\theta) = \theta^2.$$

(a) Write down the gradient descent update rule. Suppose the learning rate is α.

The update rule is

$$\theta := \theta - 2\alpha\theta.$$

The recurrence relation is

$$\theta_{t+1} = \theta_t - 2\alpha\theta_t,$$

for $t \geq 0$.

(b) Find a closed form expression for θ_t.

We have

$$\theta_t = (1 - 2\alpha)^t \theta_0.$$

(c) Using your answer to (b), discuss convergence of the gradient descent algorithm.

We have $\theta_t \to 0$ exponentially fast provided $|1-2\alpha| < 1$, which occurs if and only if $\alpha < 1$; the value θ_t orbits back and forth between $-\theta_0$ and $+\theta_0$ if $\alpha = 1$; the algorithm diverges to ∞ if $\alpha > 1$.

Problem 4: Consider again the affine objective function

$$J(\theta) = m\theta + b,$$

for some parameters $m \neq 0$ and $b \in \mathbb{R}$.

(a) Write down the gradient descent update rule with learning rate α and decay rate β.

The t-th update rule is

$$\theta := \theta - \alpha m(1-\beta)^{t+1}.$$

The recurrence relation is

$$\theta_{t+1} = \theta_t - \alpha m(1-\beta)^{t+1},$$

for $t \geq 0$.

(b) Find a closed form expression for θ_t.

Setting $\gamma = 1-\beta$ for convenience, we have

$$\theta_t = \theta_0 - \alpha m \sum_{k=1}^{t} \gamma^k,$$

for $t \geq 1$. But

$$\sum_{k=1}^{t} \gamma^k = \frac{\gamma - \gamma^{t+1}}{1-\gamma},$$

and so

$$\theta_t = \theta_0 - \alpha m \left(\frac{\gamma - \gamma^{t+1}}{1-\gamma} \right),$$

for $t \geq 1$.

(c) Using your answer to (b), discuss convergence of the gradient descent algorithm.

As we saw in Problem 2, the algorithm diverges if $\beta = 0$. But if $\beta > 0$, then $\gamma < 1$ and

$$\lim_{t\to\infty} \theta_t = \theta_0 - \alpha m \left(\frac{\gamma}{1-\gamma} \right) = \theta_0 - \alpha m \left(\frac{1-\beta}{\beta} \right).$$

Thus, the algorithm converges if the decay rate β is positive.