CM224 HW 3 Solution

Ashish Kumar Singh (UID:105479019)

October 23, 2021

Problem 1. Circumstances for gradient descent.

Answer 1. ii) The first derivative or gradient leads to a large system of equations
iii) There is no closed-form solution to your first derivative or gradient

Solution 1. If there is closed-form solution for the first derivative and gradient, we can directly equate it to zero to get extremum points. Having multiple local minima, maxima is not good as gradient descent/ascent might stuck in local minima/maxima. If function is not differentiable at some points we may use modified gradient descent but not the normal one.

Problem 2. estimate for the frequencies at the next step.

Answer 2. (a) $p(t+1) = (0.1151; 0.3360; 0.4156; 0.1333)$

Solution 2. The equation derived in class (p_3 and p_4 are reversed from class), with $p(t) = 0.11; 0.34; 0.42; 0.13$

$$p_i^{t+1} = p_i^t + \epsilon \left(\frac{\partial g(p)}{\partial p_i} - \lambda \right)$$

$$\frac{\partial g(p)}{\partial p_1} = \frac{1}{p_1} + \frac{1}{p_1 + p_2} = \frac{1}{0.11} + \frac{1}{0.11 + 0.34} = 11.313131$$

$$\frac{\partial g(p)}{\partial p_2} = \frac{1}{p_1 + p_2} = \frac{1}{0.11 + 0.34} = 2.222222$$

$$\frac{\partial g(p)}{\partial p_3} = \frac{1}{p_3 + p_4} = \frac{1}{0.42 + 0.13} = 1.818182$$

$$\frac{\partial g(p)}{\partial p_4} = \frac{1}{p_4} + \frac{1}{p_3 + p_4} = \frac{1}{0.13} + \frac{1}{0.42 + 0.13} = 9.510489$$

$$\lambda = \frac{\sum_{i=1}^{m} \frac{\partial g(p)}{\partial p_i}}{m} = \frac{11.313131 + 2.222222 + 1.818182 + 9.510489}{4} = 6.216007$$

$$p_1^{t+1} = 0.11 + 0.001 * (11.313131 - 6.216007) = 0.1151$$

$$p_2^{t+1} = 0.34 + 0.001 * (2.222222 - 6.216007) = 0.3360$$

$$p_3^{t+1} = 0.42 + 0.001 * (1.818182 - 6.216007) = 0.4156$$

$$p_4^{t+1} = 0.13 + 0.001 * (9.510489 - 6.216007) = 0.1333$$

Problem 3. Gradient descent step size and iteration for lowest value.

Answer 3. (e) $10^{-4}, [10^5, 10^6) iterations$

Solution 3. For eps of 0.001 gradient descent did not converge,
For eps of: 0.001 last iter = 10000000 $f(x, y) = nan$, $x = nan$, $y = nan$
For eps of: 0.0001 last iter = 585111 $f(x, y) = 0.000907394$, $x = 2.96988$, $y = -8.82014$
For eps of: 0.00001 last iter = 3834140, $f(x, y) = 0.00870414$, $x = 2.90672$, $y = -8.44885$