Andrea Donuro Casarova
309354
Group 1: Tuesday Tutorial

EXERCISE SHEET 3

1. Gradient Descent

$$f(x_1, x_2) = x_1^2 + 3x_2^2 + 2x_1 + 0.5x_2$$

(a) $f'_{x_1}(x_1, x_2) = 2x_1 + 2 = 0$

$$x_1 = \frac{-2}{2} = -1$$

$f''_{x_1}(x_1, x_2) = 2 > 0$ minimum

$f'_{x_2}(x_1, x_2) = 6x_2 + 0.5 = 0$

$$x_2 = \frac{-0.5}{6}$$

$f''_{x_2}(x_1, x_2) = 6 > 0$ minimum

$$x^* = \left(-1, -\frac{0.5}{6}\right)$$

$$p^* = f\left(-1, -\frac{0.5}{6}\right) = (-1)^2 + 3\cdot\left(\frac{-0.5}{6}\right)^2 + 2\cdot(-1) + 0.5\cdot\left(\frac{-0.5}{6}\right) = 1 + \frac{0.75}{36} - 2 - \frac{0.25}{6} =$$

$$= \frac{36 + 0.75 - 72 - 1.5}{36} = \frac{-36.75}{36} \qquad x^* = \left(-1, -\frac{0.5}{6}\right) \text{ and } p^* = \frac{-36.75}{36}$$

(d) Least squares linear regression minimizes the sum of squares of the difference between the observed dependent variables and the predictions by the linear function of the independent variable

The value that minimize this sum is β.

1.1

Closed form solution: $\hat{\beta} = (X^T X)^{-1} X^T y$

(b) initial point $x_0 = (3, -1)$

$\mu = 0.2$

$$\nabla f(x_1, x_2) = \left(\frac{\partial f(x_1, x_2)}{\partial x_1}, \frac{\partial f(x_1, x_2)}{\partial x_2} \right) = (2x_1 + 2, 6x_2 + 0.5)$$

$$x_{(1)} = (3, -1) - 0.2\,(2 \cdot 3 + 2, 6(-1) + 0.5) = (3, -1) - (1.6, -1.1) = (1.4, 0.1)$$

$$x_{(2)} = (1.4, 0.1) - 0.2\,(2 \cdot (1.4) + 2, 6\,(0.1) + 0.5 = (1.4, 0.1) - (0.96, 0.22) = (0.44, -0.12)$$

$$x_{(3)} = (0.44, -0.12) - 0.2\,(2 \cdot 0.44 + 2, 6 \cdot (-0.12) + 0.5) = (0.44, -0.12) - (0.576, -0.044) =$$

$$= (-0.136, -0.076)$$

Evaluating the functions of each iteration

$$f(x_{(1)1}, x_{(1)2}) = (1.4)^2 + 3\,(0.1)^2 + 2\,(1.4) + 0.5 \cdot (0.1) = 4.84$$

$$f(x_{(2)1}, x_{(2)2}) = (0.44)^2 + 3\,(-0.12)^2 + 2 \cdot (0.44) + 0.5 \cdot (-0.12) = 1.057$$

$$f(x_{(3)1}, x_{(3)2}) = (-0.136)^2 + 3 \cdot (-0.076)^2 + 2 \cdot (-0.136) + 0.5 \cdot (-0.076) = -0.27$$

(c) $x_{(1)} = (3, -1) - 0.5\,(2 \cdot 3 + 2, 6(-1) + 0.5) = (3, -1) - (4, -2.75) = (-1, 1.75)$

$$f(x_{(1)1}, x_{(2)2}) = \cancel{2}\,(-1)^2 + 3 \cdot (1.75)^2 + 2 \cdot (-1) + 0.5\,(1.75) = 9.06$$

$$x_{(2)} = (-1, 1.75) - 0.5\,(2 \cdot (-1) + 2, 6 \cdot (1.75) + 0.5) = (-1, 1.75) - (0, 5.5) = (-1, -3.75)$$

$$\cancel{f_{(2)1}}\ f(x_{(2)1}, x_{(2)2}) = (-1)^2 + 3 \cdot (-3.75)^2 + 2 \cdot (-1) + 0.5 \cdot (-3.75) = 39.31$$

$$x_{(3)} = (-1, -3.75) - 0.5\,(2 \cdot (-1) + 2, 6 \cdot (-3.75) + 0.5) = (-1, -3.75) - (0, -11) = (-1, 7.25)$$

$$f(x_{(3)1}, x_{(3)2}) = (-1)^2 + 3 \cdot (7.25)^2 + 2 \cdot (-1) + 0.5\,(7.25) = \cancel{\ldots}\ 160.31$$

2. Backtracking Line Search

(a) $f(x+\mu\Delta x) > f(x) + \alpha\mu \nabla f(x)^T \Delta x$ $\qquad \Delta x = -\nabla f(x) = \begin{bmatrix} -2x_1 \\ -2x_2 \end{bmatrix}$

$$f((x_1, x_2) + \mu(-2x_1, -2x_2) > x_1^2 + x_2^2 + a\mu(-2x_1, -2x_2)\begin{pmatrix} -2x_1 \\ -2x_2 \end{pmatrix}$$

$$f((x_1, x_2) + (2\mu x_1, 2\mu x_2) > x_1^2 + x_2^2 + a\mu(4x_1^2 + 4x_2^2)$$

$$\underbrace{f((x_1 - 2\mu x_1, x_2 - 2\mu x_2)}) > x_1^2 + x_2^2 + 4a\mu x_1^2 + 4a\mu x_2^2$$

$$f(x_1, x_2) = (x_1 - 2\mu x_1)^2 + (x_2 - 2\mu x_2)^2 = x_1^2 - 4\mu x_1^2 + 4\mu^2 x^2 + x_2^2 - 4\mu x_2^2 + 4\mu^2 x_2^2$$

$$\cancel{x_1^2} - 4\mu x_1^2 + 4\mu^2 x^2 + \cancel{x_2^2} - 4\mu x_2^2 + 4\mu^2 x_2^2 > \cancel{x_1^2} + \cancel{x_2^2} + 4a\mu x_1^2 - 4a\mu x_2^2$$

$$\left[-4\mu x_1^2 + 4\mu^2 x_1^2 - 4\mu x_2^2 + 4\mu^2 x_2^2 > 4a\mu x_1^2 + 4a\mu x_2^2\right]$$ Backtracking condition

3.1

(b) $a = 0.5$, $b = 0.1$, $\mu = 10$

$$-4 \cdot 10 \cdot (0.5)^2 + 4 \cdot 10^2 \cdot (0.5)^2 - 4 \cdot 10 \cdot 1^2 + 4 \cdot 10^2 \cdot 1^2 > 4 \cdot 0.5 \cdot 10 \cdot 0.5^2 + 4 \cdot 0.5 \cdot 10 \cdot 1^2$$

$450 > 25$ it is true so we ~~increase~~ update μ

$$\mu = b\mu = 0.1 \cdot 10 = 1$$

$$-4 \cdot 1 \cdot 0.5^2 + 4 \cdot 1^2 (0.5)^2 - 4 \cdot 1 \cdot 1^2 + 4 \cdot 1^2 \cdot 1^2 \overset{?}{>} 4 \cdot 0.5 \cdot 1 \cdot 0.5^2 + 4 \cdot 0.5 \cdot 1 \cdot 1^2$$

$$0 \not> 2.5$$

The condition will be false when $\mu = 1$ and just by updating once μ (after 1st iteration), the condition will be false.

3.2

Index der Kommentare

1.1 here you should have derived the formula starting from the least square loss function

3.1 there is math mistake that caused the formula in a) to differ

3.2 the correct answer should be stopping after 2 iterations with learning rate of 0.1