Assignment 4

~~Samgar Tristan~~, Matthew Chu, Alberto Martin Lionardi

1a. $f(x) = x^2 - 5x + 6$

starting point $x^{(0)} = 4$

step size $= 0{,}3$

$f'(x) = 2x - 5$ ✓ 0.5

$f'(4) = 2.(4) - 5$

$= 3$

$x_1 = 4 - (0{,}3 \cdot (3))$

$= 3{,}1$ ✓

$f'(3{,}1) = 2.(3{,}1) - 5$

$= 1{,}2$ ✓

$x_2 = 3{,}1 - (0{,}3 \cdot (1{,}2))$

$= 2{,}74$ ✓

$f'(2{,}74) = 2.(2{,}74) - 5$

$= 0{,}48$

$x_3 = 2{,}74 - (0{,}3 \cdot (0{,}48))$

$= 2{,}596$

$0{,}144 < 0{,}2$ ✓

$f'(2{,}596) = 2.(2{,}596) - 5$

$= 0{,}192$

$x_4 = 2{,}596 - (0{,}3 \cdot (0{,}192))$

$= 2{,}5384$

→ not necessary

2

b. Yes, it is slowly going towards the point $x = 2{,}5$ as shown in
↳ Analyze the minimum like $f'(x) = 0 \Leftrightarrow x = 2.5 \Rightarrow -0.5$
the x_4 and it will gradually gets lower.

0.5

c. the step size of 1 is too high. The value of x will go up and down countless times. I don't quite understand what does "method" imply as from I know you could only change the step size either make it higher or lower. In the question it's stated that the step size is 1. My answer is to lower the step size as changing the start value doesn't really give any helping information and also with the assumption that the step size in the question is not strictly set and can be changed.
↳ You should have analyzed the behavior of gradient descent (it oscillates between two values) $\Rightarrow -1$

1

d. $x^{(0)} = 1$

convergence point $= 2{,}5$

$$2{,}5 = 1 - a((2 \cdot 1) - 5)$$
$$= 1 - a(2 - 5)$$
$$= 1 + 3a$$
$$1{,}5 = 3a$$
$$\frac{1}{2} = a$$
$$0{,}5 = a$$

step size $= 0{,}5$ ✓ 1

1.2 a. too high ✓

b. about right ✓

c. too low ✓

d. too high ✓

1

↳ reasoning; why is it too high? $\Rightarrow -1$

$\Rightarrow$ Exercise 1: $\frac{5.5}{8}$

2. Cross - Validation

a. ~~No, it tests the performance given various circumstances. Such as in k-fold when it's given different subsets.~~

b. ~~No, in each iteration different subset of data is used. Intuitively I think it then should also gives out different result~~

c. ~~Yes, leave-one-out cross-validation uses the whole the at the same time unlike k-fold. So if you repeat the same process you will also use the same data. The result of that should also be the same as before.~~ → I have to take the worst solution if you don't cross it out.

d. ~~Not really,~~ k-fold divides it's data into k-subsets ~~while leave-one-out uses the data all at the same time~~.

~~Addition~~ (Source : ISLP) → who wants to estimate training error

NO a. CV estimates test error by averaging MSE's obtained in validation set from training data ✓1

No b. Before dividing to subsets (either LOOCV or k-CV) randomize the data first. (they might be correlated) ✓1

Yes c. each data will be the validation set (in this case element) exactly once, with it's corresponding traing data. ✓1

No d. LOOCV needs to train and fit n times.
k-fold CV — — — — — k times
LOOCV more computationally expensive
except k=n ✓1

⇒ Exercise 2: $\frac{4}{4}$

3. KNN Classifier

a. (2,3)

$d([x_1, x_2], [y_1, y_2]) = |x_1 - y_1| + |x_2 - y_2|$

(1,1) = (-)	(3,3) = (+)	$|2-1| + |3-1| = 3$ (-)	$|2-1| + |3-4| = 2$ (-)
(1,4) = (-)	(3,1) = (-)	$|2-3| + |3-1| = 3$ (-)	$|2-2| + |3-4| = 1$ (+)
(2,4) = (+)	(5,1) = (+)	$|2-5| + |3-1| = 5$ (+)	$|2-3| + |3-3| = 1$ (+) ✓ 1

k=1

(3,3) = (+)

↳ (2,3) = (+) ✓

k=3

(2,4) = (+)

(3,3) = (+)

(1,4) = (-)

↳ (2,3) = (+) ✓

2

b. the decision might be tied, for example at the same distance but the values are different.
thus we need to add additional decision (normally type I / type II error) ✓ 1

c. A problem that comes in mind is for data of range 0 and 1. Where the distances become so close that it no longer holds any significant meaning in KNN classifier

euclidean norm only works great in $\mathbb{R}^2, \mathbb{R}^3$
performs worse in higher $\mathbb{R}$ -> computationally more expensive (✓) 1

=> Exercise 3: $\frac{4}{4}$, nice!

4. Backpropagation

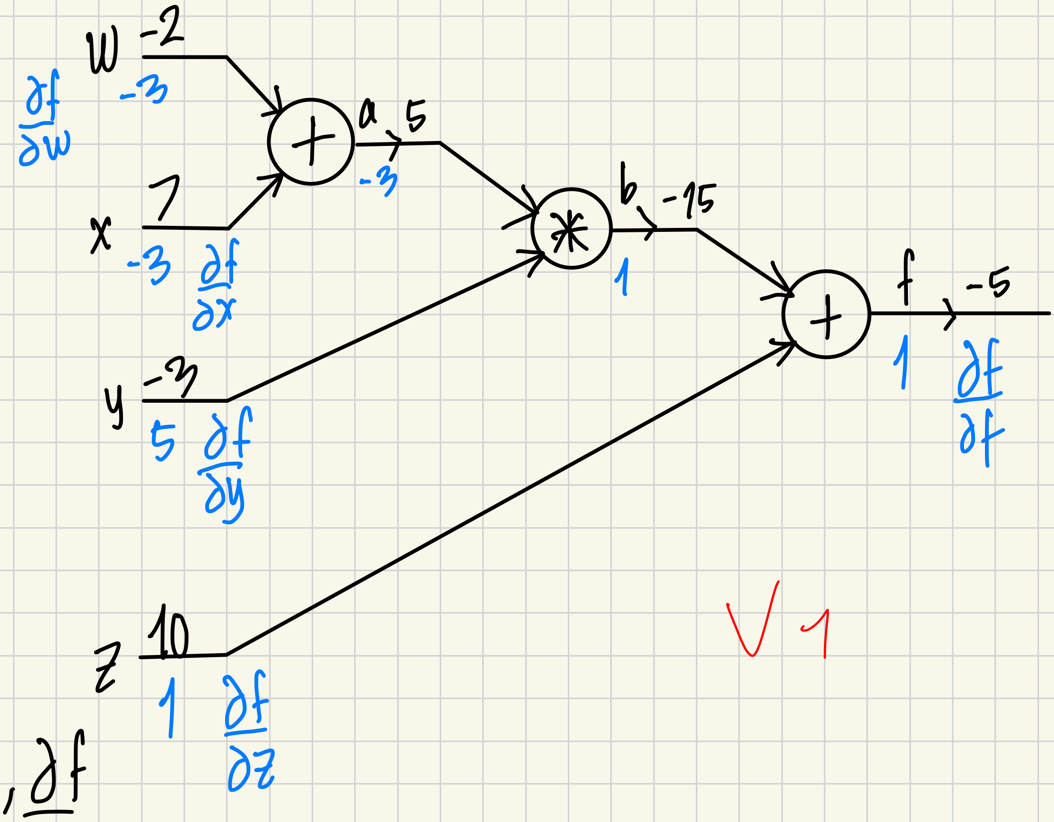
$f = (w + x)y + z$

$w = -2$

$x = 7$

$y = -3$

$z = 10$



✓ 1

$$\frac{\partial f}{\partial w}, \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$$

$a = w + x \qquad \frac{\partial a}{\partial w} = 1 \qquad \frac{\partial a}{\partial x} = 1$

$b = a * y \qquad \frac{\partial b}{\partial a} = y \qquad \frac{\partial b}{\partial y} = a$ ✓

$f = b + z \qquad \frac{\partial f}{\partial b} = 1 \qquad \frac{\partial f}{\partial z} = 1$ ✓

$$\frac{\partial f}{\partial y} = \frac{\partial f}{\partial b}\frac{\partial b}{\partial y} = a = w + x = (-2) + 7 = 5$$ ✓

$$\frac{\partial f}{\partial w} = \frac{\partial f}{\partial b}\frac{\partial b}{\partial a}\frac{\partial a}{\partial w} = 1 . y . 1 = y = -3$$ ✓

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial b}\frac{\partial b}{\partial a}\frac{\partial a}{\partial x} = 1 . y . 1 = y = -3$$ ✓

$$\boxed{\begin{array}{l} f = b + z \\ \frac{\partial f}{\partial b} = 1 \\ f = (a * y) + z \\ \frac{\partial f}{\partial a} = \frac{\partial f}{\partial b}\frac{\partial b}{\partial a} = 1 \cdot y = y \end{array}}$$

⇒ Exercise 4: $\frac{4}{4}$, very good!

↳ $\frac{17.5}{20}$ ☺