Problem Set 3

1).1 when $x < -1/2$ $\quad f(x) = 1/2\, x^4 + 5x^3 + \frac{27}{2} x^2 + 10x$

$$f'(x) = 2x^3 + 15x^2 + 27x + 10$$
$$f'(-6) = 2(-6)^3 + 15(-6)^2 + 27(-6) + 10$$
$$= -44$$

1.2) when $-1/2 \leq x \leq 5$ $\quad f(x) = -71/16$

$$f(0) = -71/16$$
$$f'(0) = 0$$

1.3 when $x \geq 5$ $\quad f(x) = x^2 - 10x + 329/16$

$$f'(x) = 2x - 10$$
$$f'(8) = 2(8) - 10 = 6$$

1.4 Pacman and Paclady can employ gradient descent and ensure their gradient hits zero but the would reach the bar when their location $x < -2$
For $x = -2$ the gradient is zero and their location is a local maxima.
For $x > -2$ they'll reach gradient $= 0$ at $x = -1/2$ which is local minima.

1.5 The previous gradient descent algorithm would take them to $x = -2$ because gradient would become zero and would terminate. They would be stuck at $x = -2$ instead of going to $x = 3$.

2 x_1 x_2 can take the following 4 values

	x_1	x_2
Case 1	0	0
Case 2	0	1
Cas 3	1	0
Cas 4	1	1

2.1) Case 1

$H_1(n) = \sigma((-5)\cdot 0 + (-1)0 + 0.5) = \sigma(0.5) = 0.622$

$H_2(n) = \sigma((-2)\,0 + (-1)0 + 0.5) = \sigma(0.5) = 0.622$

$y = \sigma((-1)\cdot 0.622 + (-4)0.622 + 0.5) = \sigma(-2.61) = 0.068$

round$(0.068) = 0$

Case 2

$H1(n) = \sigma(0\cdot(-5) + -1(1) + 0.5) = \sigma(-0.5) = 0.377$

$H_2(n) = \sigma((-2)(1) + 0(-1) + 0.5) = \sigma(-1.5) = 0.182$

$y =$ round $\sigma(-1(0.377) + -4(0.182) + 0.5) = \sigma(-0.602) = 0.35$

round$(0.35) = 0$

Case 3

$H1(n) = \sigma(1\times -5 + -1(0) + 0.5) = \sigma(-4.5) = 0.01$

$H2(n) = \sigma(-2(0) + -1(1) + 0.5) = \sigma(-0.5) = 0.377$

$y =$ round $\sigma(-1(0.377) + -4(0.01) + 0.5) = 0.265$

round$(0.265) = 0$

Case 4

$H1(n) = \sigma(-5(1) + -1(1) + 0.5) = \sigma(-5.5) = 0.004$

$H2(n) = \sigma(-2(1) + -1(1) + 0.5) = \sigma(-2.5) = 0.075$

$y(n) = \sigma((-1)(0.004) + -4(0.075) + 0.5) = \sigma(\quad) \; 0.54.$

$\text{round}(0.54) = 1$

This an AND Gate

2.2) Case 1

$H1(n) = \sigma(0(-5) + 0(-4) + 0.5) = \sigma(0.5) = 0.622$

$H2(n) = \sigma(0(3) + 0(4) + 0.5) = \sigma(0.5) = 0.622$

$y(n) = \sigma(0.622 \times 2 + 0.622(-1) + 0.5) = \sigma(1.122) = 0.75$

$\text{round}(0.75) = 1$

Case 2

$H1(n) = \sigma(0(-5) + 1(-4) + 0.5) = \sigma(-3.5) = 0.029$

$H2(n) = \sigma(0(3) + 1(4) + 0.5) = \sigma(4.5) = 0.989$

$y(n) = \sigma(0.029 \times 2 + -1(0.989) + 0.5) = \sigma(-0.431) = 0.39$

$\text{round}(0.39) = 0$

Case 3

$H1(n) = \sigma(1(-5) + 0(-4) + 0.5) = \sigma(-4.5) = 0.010$

$H2(n) = \sigma(1(3) + 0(4) + 0.5) = \sigma(3.5) = 0.970$

$y(n) = \sigma(0.01 \times 2 + 0.97 \times -1 + 0.5) = \sigma(0.450) = 0.38$

$\text{round}(0.38) = 0$

Case 4

$H1(x) = \sigma(1(-5) + 1(-4) + 0.5) = \sigma(-8.5) = 0.0002$

$H2(x) = \sigma(1(3) + 1(4) + 0.5) = \sigma(7.5) = 0.99$

$y(x) = \sigma(2(0.0002) + -1(0.99) + 0.5) = \sigma(-0.48) = 0.38$

$round(0.38) = 0$

This is a NOR gate.

2.3) Case 1

$H1(x) = \sigma(0(3) + 0(-3) + 0.5) = \sigma(0.5) = 0.622$

$H2(x) = \sigma(0.(1) + 0(-4) + 0.5) = \sigma(0.5) = 0.622$

$y(x) = \sigma(-3(0.622) + 5(0.622) + 0.5) = 0.85$

$round(0.85) = 1$

Case 2

$H1(x) = \sigma(0(3) + 1(-3) + 0.5) = \sigma(-2.5) = 0.075$

$H2(x) = \sigma(0(1) + (-4) + 0.5) = \sigma(-3.5) = 0.029$

$y(x) = \sigma(-3(0.075) + 5(0.029) + 0.5) = \sigma(0.42) = 0.603$

$round(0.603) = 1$

Case 3

$H1(x) = \sigma(3 + 0.5) = \sigma(3.5) = 0.97$

$H2(x) = \sigma(1 + 0.5) = \sigma(1.5) = 0.81$

$y(x) = \sigma((0.97)-3 + 5(0.81) + 0.5) = \sigma(1.67) = 0.84$

$round(0.84) = 1$

Case 4

$H1(n) = \sigma(3 - 3 + 0.5) = \sigma(0.5) = 0.622$

$H2(n) = \sigma(1 - 4 + 0.5) = \sigma(-2.5) = 0.075$

$y(n) = \sigma(-3 \times 0.622 + 5 \times 0.075 + 0.5) = \sigma(-0.98) = 0.27$

round $(0.27) = 0$

This is NAND gate.

Q2.4) Case 1

$H1(n) = \sigma(0(2) + 0(-3) + 0.5) = \sigma(0.5) = 0.622$

$H2(n) = \sigma(0(3) + 0(-3) + 0.5) = \sigma(0.5) = 0.622.$

$y(n) = \sigma(0(4) + 0(-4)(0.622) + 0.5) = \sigma(0.5) = 0.622$

round $(0.622) = 1$

Case 2

$H1(n) = \sigma(0(2) + 1(-3) + 0.5) = \sigma(-2.5) = 0.075.$

$H2(n) = \sigma(0(3) + (-3)1 + 0.5) = \sigma(-2.5) = 0.075)$

$y(n) = \sigma(4 \times 0.075 - 4 \times 0.075 + 0.5) = \sigma(0.5) = 0.622.$

round $(0.622) = 1.$

Cas 3.

$H1(n) = \sigma(2 + 0.5) = \sigma(2.5) = 0.92$

$H2(n) = \sigma(3 + 0.5) = \sigma(3.5) = 0.97$

$y(n) = \sigma(4(0.92) + -4(0.97) + 0.5) = \sigma(0.297) = 0.57$

round $(0.57) = 1$

Case 4)

$H1(x) = \sigma(2 + -3 + 0.5) = \sigma(-0.5) = 0.377$

$H2(x) = \sigma(0.5) = 0.622$

$y(x) = \sigma[4(0.377) + -4(0.622) + 0.5] = \sigma(-0.47) = 0.38$

$\text{round}(0.38) = 0$

This is also a NAND gate.

3.1) $w_A\, f(x) = (-1.06 \times -0.82) + (-0.02 \times 0.95) = 0.8502$

$w_B\, f(x) = (-1.06 \times -1.63) + (-0.88 \times 0.95) = 0.8918$

$w_C\, f(x) = (-1.06 \times 0.39) + (0.65 \times 0.95) = 0.2041$

3.2) new $w_{A1} = w_{A1} = -0.82 = -0.82$

new $w_{A2} = w_{A2} = -0.02 = -0.02$

new $w_{B1} = w_{B1} - f(x) = -1.63 + 1.06 = -0.57$

new $w_{B2} = w_{B2} - f(x) = -0.88 - 0.95 = -1.83$

new $w_{C1} = w_{C1} + f(x) = 0.39 - 1.06 = -0.67$

new $w_{C2} = w_{C2} + f(x) = 0.65 + 0.95 = 1.6$

3.3) $w_A\, f_A = (0.82 \times 0.09) + (-0.02 \times 1.48) = 0.0042$

$w_B\, f_B = (-0.57 \times 0.09) + (-1.83 \times 1.48) = -2.7597$

$w_C\, f_C = (-0.67 \times 0.09) + (1.6 \times 1.48) = 2.3077$

The prediction is 'C' but the label given is A.

$\text{new } w_{A1} = w_{A1} + f(x) = -0.82 + 0.09) = -0.73$

$\text{new } w_{A2} = w_{A2} + f(x) = -0.02 + 0.48 = -1.50$

$\text{new } w_{B1} = w_{B1} - f(x) = -0.57 = -0.57$

$\text{new } w_{B2} = w_{B2} - f(x) = -1.83 = -1.83$

$\text{new } w_{C1} = w_{C1} - f(x) = -0.67 - 0.09 = -0.76$

$\text{new } w_{C2} = w_{C2} - f(x) = 1.06 - 1.48 = 0.12$

3.4) $w_A \; f(x) = (3.12 \times -1.35) + (0.96 \times 0.42) = -3.8088$

$w_B \; f(x) = (3.11 \times -1.35) + (-0.97 \times 0.42) = -4.6059$

$w_C \; f(x) = (-8.29 \times -1.35) + (-0.24 \times 0.42) = 11.0907$

3.5) Predicted label is "C"

4).1 This may or may not increase the performance. The perceptron is guaranteed reach zero error provided α is small enough. (α - learning rate). If learning rate is large, training the weights for a longer time will keep the algorithm from converging and will keep bouncing around.

4.2) By adding higher order features such as pairwise products, increases the dimentionality which increases the chance of finding a hyperplane that can seperate the data points. This can be useful if original features failed to converge

4.3) If some features were irrelavant or noisy then the training data might not reduce. So removing features ~~will definitely~~ might improve on training error. Sometimes more feature can fit noise and inconsistancies in data too

4.4) Collecting larger data set is likely to improve performance. A larger data will help the perceptron fit better ~~of~~ over the distributed data.

5.1) False	5.2) True.	5.3) True	5.4) True
5.5) False	5.6) False	5.7) False	5.8) True
5.9) ~~True~~ False	5.10) True	5.11) True	5.12) ~~True~~ False.
5.13) False	5.14) True	5.15) True	5.16) True.
5.17) True	5.18) False	5.19) True.	

6.1) $h_1 = \sigma(i_1 w_1 + i_2 w_2 + b_1) = \sigma(0.05 \times 0.15 + 0.1 \times 0.3 + 0.35)$

$= \sigma(0.3875) = 0.5956$

$h_2 = \sigma(i_1 w_3 + i_2 w_4 + b_1) = \sigma(0.05 \times 0.12 + 0.1 \times 0.28 + 0.35)$

$= \sigma(0.384) = 0.5948$

$$O_1 = \sigma(h_1 w_5 + h_2 w_6 + b_2) = \sigma(0.5956 \times 0.5 + 0.5948 \times 0.55 + 0.6)$$
$$= \sigma(1.224) = 0.77$$

$$O_2 = \sigma(h_1 w_7 + h_2 w_8 + b_2) = \sigma(0.5956 \times 0.45 + 0.5948 \times 0.4 + 0.6)$$
$$= \sigma(1.105) = 0.75$$

6.2) $$\text{Error} = \sum \tfrac{1}{2}(\text{target} - \text{output})^2$$
$$= \tfrac{1}{2}\left\{(0.1 - 0.77)^2 + (0.75 - 0.99)^2\right]$$
$$= 0.2532$$

6.3) $$\frac{\partial E}{\partial w} = \frac{\partial E}{\partial out} \times \frac{\partial out}{\partial net} \times \frac{\partial net}{\partial w}$$

$$= -(\text{target} - \text{out}) \ \sigma(\text{net} \times 1 - \text{net}) \times h$$

$$\frac{dE}{dw_5} = -(0.1 - 0.77) \times 0.77(1-0.77) \times 0.5956 = 0.0706$$

$$w_5^{update} = 0.5 - 0.5(0.0706) = 0.4647$$

$$\frac{dE}{dw_6} = -(0.1 - 0.77 \times 0.77(1-0.77) \times 0.5948 = 0.0704$$

$$w_6^{update} = 0.55 - (0.5 \times 0.0704) = 0.5147$$

$$\frac{dE}{dw_7} = -(0.99 - 0.75) \times 0.75(1-0.75) \times 0.5956 = -0.0268$$

$$w_7^{update} = 0.45 - 0.5(-0.0268) = 0.4634$$

$$\frac{dE}{dw_8} = -(0.99 - 0.75) \times 0.75(1-0.75) \times 0.5948 = -0.0267$$

$$w_8^{update} = 0.4 - 0.5(-0.0267) = 0.4135$$

Hence,
$w_5 = 0.4647$
$w_6 = 0.0706$
$w_7 = 0.4634$
$w_8 = 0.4135$