Q1)

a)

i: No because the new decision boundary becomes $\theta \cdot x = 0$. And if we were to take a point that was 0 for one of the xs, such as $(1,1,0)$, then that translates to $\theta \cdot (1,1,0) = \theta_1 + \theta_2$. We want $\theta \cdot (1,1,1)$ to be positive, but if there are any xs that are 0, it would create a contradiction if we try to make it ⊖. You end up with $\theta_1 + \theta_2 + \theta_3 > 0$ and $\theta_1 + \theta_2 < 0$, $\theta_1 + \theta_3 < 0$, and $\theta_2 + \theta_3 < 0$.

ii: It is possible. An example would be $\theta = (1,1,1)$, so we set $\theta \cdot (1,1,1) + b > 0$, and $\theta \cdot (x_1, x_2, x_3) + b < 0$ for all $(x_1, x_2, x_3) \neq (1,1,1)$. So $\theta \cdot (1,1,1) = 3$, $\theta \cdot (1,1,0) = 2$, $\theta \cdot (1,0,0) = 1$. The largest dot-product among the negatives is 2, so $b < -2$ for all negatives, but $3 + b > 0$, so $b > -3$, $\Rightarrow$ $-3 < b < -2$.

b)

i: $\|(1,1)\| = \sqrt{2}$, $\|(-1,1)\| = \sqrt{2}$, $\|(0,1)\| = 1$, $\|(-1,0)\| = 1$

We can make the inside of a circle positive, so r would have to be at least $\sqrt{2}$ to include $[1,1]$ and $[-1,-1]$, but then $[0,1]$ and $[-1,0]$ are also inside, and this contradicts their negative values. We can also make the outside of the circle be ⊕, and we can pick an r between 1 and $\sqrt{2}$, so $[1,1]$ and $[-1,-1]$ are outside, while $[0,1]$ and $[-1,0]$ are inside. So $1 \leq r \leq \sqrt{2}$ allows us to classify all 4 points with "outside" = positive and "inside" = negative

ii: $\{x : a \cdot x = 0\}$, $a = (a_1, a_2)$, we need $a \cdot (1,1) > 0$, $a \cdot (-1,-1) > 0$, $a \cdot (0,1) < 0$, and $a \cdot (-1,0) < 0$. There's a contradiction b/c $a \cdot (1,1) > 0$ means $a_1 + a_2 > 0$, but $a \cdot (-1,-1) > 0$ means $-(a_1 + a_2) > 0$. So no classifiers exist.

iii: Corners $[a \pm s/2, b \pm s/2]$. Points are inside if $\max(|x-a|, |y-b|) \leq s/2$ and outside otherwise. We need $[1,1]$ and $[-1,-1]$ on the same side, and $[0,1]$ and $[0,-1]$ on the other side. We can make "inside" negative so that $[0,1]$ and $[-1,0]$ are inside, and $[1,1]$ and $[-1,-1]$ are outside).
Example $\Rightarrow$ $(a,b) = (-0.5, 0.5)$, $s = 1$, so $\max(|0-(-0.5)|, |0-0.5|) = 0.5$, so $[0,1]$ is inside, $\max(|-1+0.5|, |0-0.5|) = 0.5$, so $[-1,0]$ is inside, $\max(|1+0.5|, |1-0.5|) = 1.5$, which makes it outside, and finally, $\max(|-1+0.5|, |-1-0.5|) = 1.5$, which makes it outside

iv: angle $\alpha = 45°$, $\max(|x'|, |y'|) \le \frac{s}{2}$ $\Rightarrow \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$
$[1,1] \Rightarrow (\sqrt{2}, 0)$, $[-1,-1] \Rightarrow (-\sqrt{2}, 0)$, $[0,1] \Rightarrow (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, $[-1,0] \Rightarrow (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$
Inside is negative, and we pick an s s.t. $||x'||$ or $||y'|| \le \frac{s}{2}$, for negatives but $> \frac{s}{2}$ for positives. Example $\Rightarrow$ $s = 2$, $\max(|\sqrt{2}|, 0) = \sqrt{2}$, greater than $\frac{s}{2} (=1)$, so $[1,1]$ and $[-1,-1]$ are outside / positive while $[0,1]$ and $[-1,0]$ are inside / negative

Q2)

a) $\theta = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}$, $b = \sum_{i=1}^{n} \alpha_i y^{(i)}$

b) Decision boundary; $\theta \cdot x + b = 0$ $\rightarrow$ unsigned distance; $\frac{|\theta \cdot x_0 + b|}{||\theta||} = \frac{|\theta \cdot 0 + b|}{||\theta||} = \frac{|b|}{||\theta||}$

d) $\theta = [4, 0]$, $b = -6.0$

They are consistent with the results found in part A

α

i	# of misclassifications
0	1
1	0
2	11
3	2
4	0
5	4

1F

e) The order does not matter. Perceptron will converge if the points are linearly separable. The order does affect the total # of mistakes.

Q3.1)

d)

Algorithm	η	θ_0	θ_1	iter	runtime (s)
GD	10^{-4}	0.3509	0.0322	353622	2.1320
GD	10^{-3}	0.3559	.0233	52628	0.3206
GD	10^{-2}	0.3575	0.0205	6988	0.0433
GD	10^{-1}	0.3580	0.0196	870	0.0053
SGD	10^{-4}	0.3509	0.0322	353621	15.0201
SGD	10^{-3}	0.3559	0.0233	52620	2.2318
SGD	10^{-2}	0.3573	0.0203	6960	0.2966
SGD	10^{-1}	0.3564	0.0185	775	0.0338
Closed form	-	0.3587	0.0192	-	0.000123

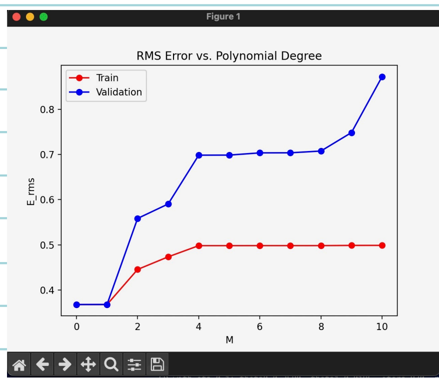
e) For both SGD and GD, a learning rate of 10^{-1} led to the least amount of iterations, and the smallest runtime. The θ_0 for both were generally in the same range, as well as their θ_1. But the runtime for SGD when $\eta = 10^{-4}$ is way higher than that of GD.

f) For the closed form solution the runtime is extremely small compared to the SGD runtimes. A learning rate of 10^{-1} for SGD produces a runtime that is the closest to closed form's.

G) With my new proposed learning rate, it takes 0.1452 seconds to converge with 3399 iterations. My coefficients are extremely close to that of the closed form solution. It performs about the same.

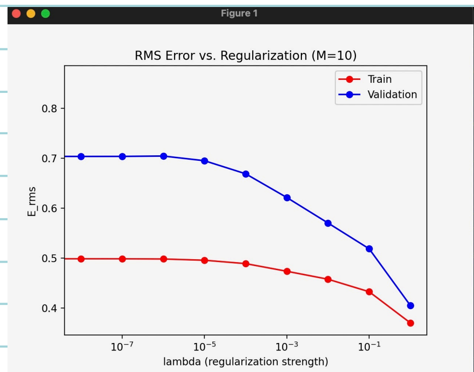
Q 3.2)

b)



C) We can see evidence for under-fitting when M is very low, but it yields the smallest validation error. We can see overfitting when M=2 or 3 because the validation error rises significantly, and the gap between that and the train error shows overfitting. So the best degree polynomial would be 1.

e)



f) $\lambda = 10^{-1}$ performed the best because it had the lowest validation RMS

G) I believe the value of M can vary for this dataset, but for λ it should be 10^{-1}.