$$s = \boldsymbol{w}^T z_n x_n$$

Additionally we know that $\boldsymbol{w}^T z_n x_n$ is convex since it is linear (as proved in HW2).

Now we prove that $f = max(-s, 0)$ is convex for s. For s_1, s_2 we have

$$f(\lambda s_1 + (1-\lambda)s_2) = max(-(\lambda s_1 + (1-\lambda)s_2), 0)$$

$$\leq max(-\lambda s_1, 0) + max(-(1-\lambda)s_2, 0)$$

$$\leq \lambda max(-s_1, 0) + (1-\lambda)max(-s_2, 0)$$

$$\leq \lambda f(s_1) + (1-\lambda)f(s_2)$$

Therefore, by defining $\boxed{J = f}$ and $\boxed{s = \boldsymbol{w}^T z_n x_n}$, we can say that

$$\boxed{J(\lambda \boldsymbol{w_1} + (1-\lambda)\boldsymbol{w_2}) \leq \lambda J(\boldsymbol{w_1}) + (1-\lambda)J(\boldsymbol{w_2})}$$

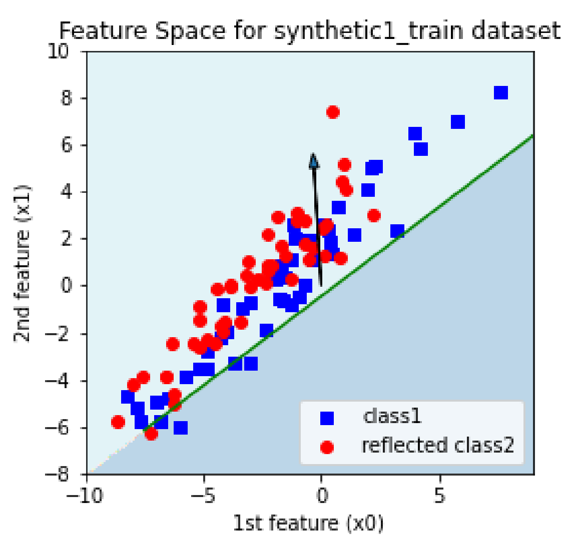
Therefore, Loss Function, $\boxed{J \textbf{ is convex}}$. Hence Proved!

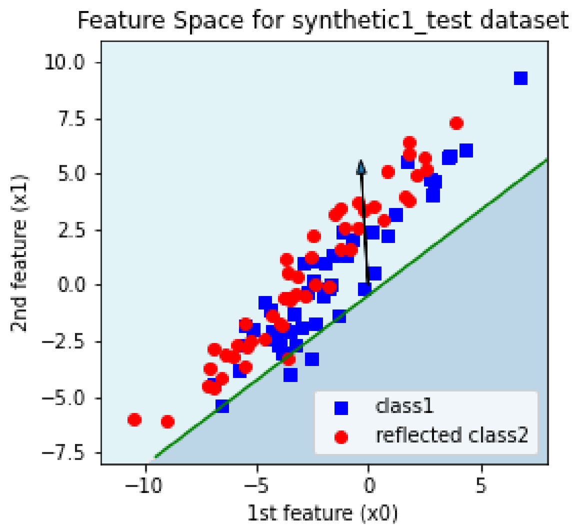
0.2 Problem 2

0.3 2-CLASS PERCEPTRON ALGORITHM

0.3.1 Solution 2(a) - SYNTHETIC1

This dataset has 2 features and the following decision boundary plots are computed for both the train and the test set.





As seen, the data points are plotted in reflected, non-augmented domain. They are not linearly separable as the **Halting condition 2 was satisfied.**. This can also be justified from the plots above. All the points belonging to both classes are lying together and no line can separate them

The classification error for both the training and the test set are as follows:

Dataset	Classification Error	No. of misclassified points
synthetic1_train	0.06	6
synthetic1_test	0.03	3

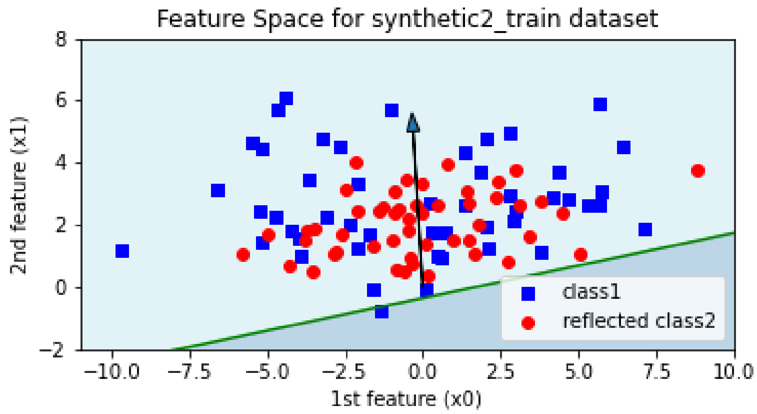
The optimized weights and the optimized criterion function were computed from the last 500 iterations as asked in the question.

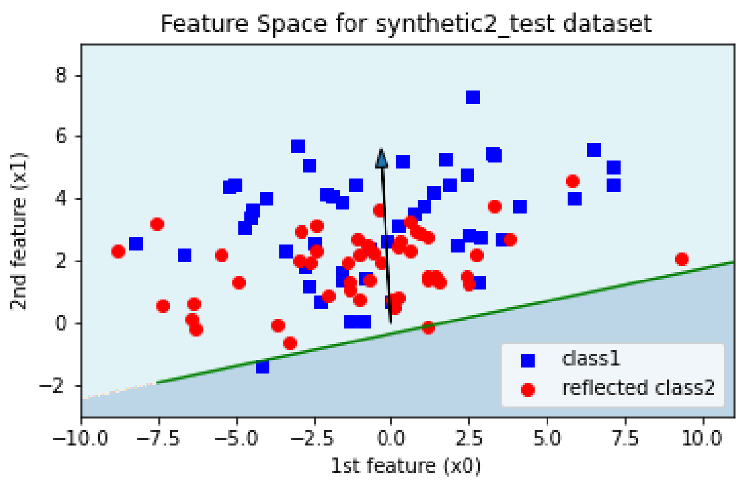
The following table tabulates the optimized values.

Dataset	Optimized Weight	Optimized Criterion
synthetic2	$[17.1, -28.0167, 36.9162]$	-492

0.3.2 Solution 2(b) - SYNTHETIC2

The following are the two plots of the decision boundaries for both the train and the test sets.





The data is not linearly separable as seen in the above plots. There is no line that clearly separates the two classes. And according to the **Halting condition 2 was satisfied.**, the optimized weights and the optimized criterion function are as follows:

Dataset	Optimized Weight	Optimized Criterion
synthetic2	$[6.1, -3.6549, 17.5007]$	-326

For the above optimized values, the classification error is as follows:

Dataset	Classification Error	No. of misclassified points
synthetic2_train	0.01	1
synthetic2_test	0.02	2

0.3.3 Solution 2(c) - WINE

This dataset has a total of 13 Features, of which all were used to train the Perceptron Algorithm.

For this problem only first 2 classes were considered and the resultant total number of data points were 65. The classification error for the training and the test set were calculated as follows:

Dataset	Classification Error	No. of misclassified points
wine_train	0.4654	30
wine_test	0.4462	29

No assertion can be made about the Linear Separability of this dataset and the Algorithm was run for all the 10,000 iterations. So, according to the **halting condition 2** (as this was satified), the optimized weight vector corresponds to the lowest criterion function value $J(\underline{w})$. the following table tabulates the optimized $\underline{\hat{w}}(\hat{w_0}, \hat{w_0}, ..., \hat{w_D})$ - augmented space and the corresponding $J(\underline{\hat{w}})$

As there is no definite way of finding a linear line separating all the data points, the linear separability of the wine dataset cannot be proved.

WINE:
Optimized weight vector: $[-266.9, -2924.62, -481.55, -399.35, -5724.8, -16981.9, -532.67, -214.4, -130.17, -288.86, -452.2, -217.716, -427.85, 1267.1]$

Optimized Criterion: -1014