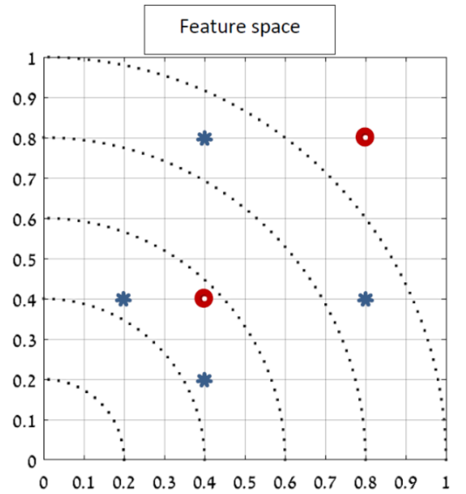
Choose one question, among theory or practical.

Problem 4 – Nonlinear SVM [10 pts bonus]

Before starting, note that sections 4-6 are above the expected level in this course.

We now introduce the normalized linear kernel:

$$K(x, y) = \frac{x^t y}{\|x\| \cdot \|y\|}$$



1. Find the feature vector mapping, i.e., the φ that maps each sample to its feature space.

2. Given the following data, map the points to their new feature representations using the figure as the feature space. Draw the 6 new points (with colors) on the figure.

3. Draw the resulting margin decision boundary in the feature space. Also draw the separating line. You can use the following link.

4. For this section, work **only** with the new feature space.

   Given that the separating hyperplane is defined by $\ell\colon -x - y + 1.378 = 0$, so w=(-1,-1), b=1.378, find the alphas for each sample.

   Guide: start from the rule that $w = \sum_{i=1}^{3} y_i \alpha_i x_i$ to get 2 equations. Obtain the third equation by the rule $\sum_{i=1}^{3} y_i \alpha_i = 0$ and solve system of 3 equations (no need to show, but write the final alphas).

   Note: most of the time, it is not that simple, but since we have only 3 points in 2d, it works.

5. Draw inside the desmos link the decision boundary in the original input space, resulting from the kernel. Recall that the nonlinear hyperplane given by:

$$\sum_{i=1}^{3} y_i \alpha_i K(x, x_i) + b = 0$$

   Where b is given to you, you found the alphas earlier.

6. Consider two distinct points $x_1, x_2 \in \mathbb{R}^d$ with labels $y_1 = 1, y_2 = -1$.

   Compute the hyperplane that Hard SVM will return on this data, i.e., give explicit expressions for w and b as functions of x1, x2.

   Hint: convert the primal problem to the dual one and reduce it to one variable problem.

Problem 4 – SVM [10 pts bonus]

We will change the algorithm for Hard SVM by learning only from the samples with positive labels and ignoring samples with negative labelings. Hence, the optimization problem becomes:

$$\min \frac{1}{2}\|w\|^2$$

$$\text{s.t. } \forall i \in J_+, w^\top x_i + b \geq 1$$

As J_+ is the set of positive labeled samples indexes.

1. Under the settings above, what will be the solution of w? Justify.
2. If we set b=0, meaning we remain only with $w^\top x_i \geq 1$, what is $\min_{i \in J_+} w^\top x_i$?
3. Let w^* be the solution to the problem (b=0). We will classify new samples as:

$$\hat{y} = \begin{cases} 1 & w^\top x \geq \min_{i \in J_+} w^\top x_i - \varepsilon \\ -1 & \text{otherwise} \end{cases}$$

   For some small $\varepsilon > 0$. Will this condition classify correctly all the training samples, both positive and negative? Justify