H/W 4

(a) $x_1' = \cos(x_n^{(1)} - x_n^{(2)})$ $x_2' = \sin(x_n^{(1)} - x_n^{(2)})$

The resulting plot after the above transformation will be a circle. Essentially the part " $x_n^{(1)} - x_n^{(2)}$ " reduces the problem to 1D and the second part $[\cos(x_n^{(1)} - x_n^{(2)}), \sin(x_n^{(1)} - x_n^{(2)})]$ transforms it back to 2D and maps it into a point (if we speak about a single point) in the unit circle. This is natural as cosine and sine functions describe points on a circle based on the angle, which in this case is determined by the difference between $x_n^{(1)}$ and $x_n^{(2)}$. (We can further prove this by getting a random point and calculating its Euclidean distance, which will always be 1 due to the identity $\cos^2 x + \sin^2 x = 1$).

(b) • $\boxed{y(x)=0}$

This will be the decision boundary and it will be a single line ~~[illegible]~~ that will divide the circle into 2 parts, where one will be the part belonging to the blue class and the other will belong to the red.

• $\boxed{w}$

The vector w will control the orientation of the decision boundary, so essentially the slope of the line set by $y(x)=0$.

• $\boxed{b}$

b affects the position of the decision boundary in the new space, so essentially it will be the intercept of the line set by $y(x)=0$

(c) $L(\mu,\lambda,w,b,\xi) = \frac{1}{2}\|w\|_2^2 + C\sum_{n=1}^{N}\xi_n - \sum_{n=1}^{N}\lambda_n\left(t_n\left(\Phi^T(x_n)w - b\right) - 1 + \xi_n\right) - \sum_{n=1}^{N}\mu_n \cdot \xi_n$

(d) $\boxed{KKT}$ s.t. $\lambda_u \geq 0$ $\mu_u \geq 0$ (dual), $\forall x \in [1, N]$ 6N conditions

$t_u(\phi^T(x_u)w - b) - 1 + \xi_u \geq 0$ $\xi_u \geq 0$ (primal), $\forall x \in [1, N]$

$\lambda_u(t_u(\phi^T(x_u)w - b) - 1 + \xi_u) = 0$ $\mu_u \cdot \xi_u = 0$ (comp. slack.), $\forall x \in [1, N]$

(e). $\dfrac{\partial}{\partial w}\left(\frac{1}{2}\|w\|_2^2 + \cancel{C\sum_{n=1}^{N}\xi_n} - \sum_{n=1}^{N}\lambda_n\left(t_n\left(\phi^T(x_n)w - b\right) - 1 + \xi_n\right) - \cancel{\sum_{n=1}^{N}\mu_n\xi_n}\right) =$

$$= \frac{\partial}{\partial w}\left(\frac{1}{2}w^T w\right) - \frac{\partial}{\partial w}\left(\sum_{n=1}^{N}\lambda_n\left(t_n\left(\phi^T(x_n)w - b\right) - 1 + \xi_n\right)\right) =$$

$$= \frac{1}{2}\cdot 2w^T - \sum_{n=1}^{N}\lambda_n\left(t_n\left(\frac{\partial}{\partial w}\left(\phi^T(x_n)\cdot w\right) - \cancel{b}\right) - \cancel{1} + \cancel{\xi_n}\right) =$$

$$= w^T - \sum_{n=1}^{N}\lambda_n t_n \phi^T(x_n)$$

$$\frac{\partial}{\partial w} = 0 \Leftrightarrow w^T = \sum_{n=1}^{N}\lambda_n t_n \phi^T(x_n) \Leftrightarrow \boxed{w = \sum_{n=1}^{N}\lambda_n t_n \phi(x_n)} \ (1)$$

• $\dfrac{\partial}{\partial b}\left(\cancel{\frac{1}{2}\|w\|_2^2} + \cancel{C\sum_{n=1}^{N}\xi_n} - \sum_{n=1}^{N}\lambda_n\left(t_n\left(\phi^T(x_n)w - b\right) - 1 + \xi_n\right) - \cancel{\sum_{n=1}^{N}\mu_n\xi_n}\right) =$

$$= -\frac{\partial}{\partial b}\left(\sum_{n=1}^{N}\lambda_n\left(t_n\left(\cancel{\phi^T(x_n)w} - b\right) - \cancel{1} + \cancel{\xi_n}\right)\right) =$$

$$= -\sum_{n=1}^{N} -\lambda_n t_n = \sum_{n=1}^{N}\lambda_n t_n$$

$$\frac{\partial}{\partial b} = 0 \Leftrightarrow \boxed{\sum_{n=1}^{N}\lambda_n t_n = 0} \ (2)$$

• $\dfrac{\partial}{\partial \xi}\left(\cancel{\frac{1}{2}\|w\|_2^2} + C\sum_{n=1}^{N}\xi_n - \sum_{n=1}^{N}\lambda_n\left(\cancel{t_n\left(\phi^T(x_n)w - b\right)} - \cancel{1} + \xi_n\right) - \sum_{n=1}^{N}\mu_n\xi_n\right) =$

$$= \frac{\partial}{\partial \xi}\left(C\sum_{n=1}^{N}\xi_n\right) - \sum_{n=1}^{N}\frac{\partial}{\partial \xi}\left(\xi_n\right) - \sum_{n=1}^{N}\frac{\partial}{\partial \xi}\mu_n\xi_n =$$

$$= C - \lambda_n - \mu_n$$

$$\frac{\partial}{\partial \xi} = 0 \Leftrightarrow \boxed{\lambda_n = C - \mu_n} \ (3)$$

So, (1), (2), (3) are our stationary points

$$L(\mu,\lambda,w,b,\xi) = \frac{1}{2}\|w\|_2^2 + C\sum_{n=1}^{N}\xi_n - \sum_{n=1}^{N}\lambda_n\left(t_n(\phi^T(x_n)w - b) - 1 + \xi_n\right) - \sum_{n=1}^{N}\mu_n\xi_n =$$

$$= \frac{1}{2}w^Tw - \sum_{n=1}^{N}\lambda_n t_n \phi^T(x_n)w - \sum_{n=1}^{N}\lambda_n t_n b + C\sum_{n=1}^{N}\xi_n + \sum_{n=1}^{N}\lambda_n(1-\xi_n) - \sum_{n=1}^{N}\mu_n\xi_n =$$

$$\overset{1}{\underset{2}{=}} \left(\frac{1}{2}w^T - \overset{1}{w^T}\right)w - b\cdot\cancel{\sum_{n=1}^{N}\lambda_n t_n} + C\sum_{n=1}^{N}\xi_n - \sum_{n=1}^{N}\lambda_n + \sum_{n=1}^{N}\lambda_n - \sum_{n=1}^{N}\lambda_n\xi_n - \sum_{n=1}^{N}\mu_n\xi_n =$$

$$\overset{3}{=} -\frac{1}{2}w^Tw + C\sum_{n=1}^{N}\xi_n + \sum_{n=1}^{N}\lambda_n - \sum_{n=1}^{N}(\overset{3}{C} - \mu_n)\xi_n - \sum_{n=1}^{N}\mu_n\xi_n =$$

$$= -\frac{1}{2}w^Tw + \cancel{C\sum_{n=1}^{N}\xi_n} + \sum_{n=1}^{N}\lambda_n - \cancel{C\sum_{n=1}^{N}\xi_n} + \cancel{\sum_{n=1}^{N}\mu_n\xi_n} - \cancel{\sum_{n=1}^{N}\mu_n\xi_n}$$

$$= \sum_{n=1}^{N}\lambda_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\lambda_n\lambda_m t_n t_m \phi^T(x_n)\cdot\phi(x_m)$$

So the dual formulation is:

$$\tilde{L}(\lambda) = \sum_{n=1}^{N}\lambda_n - \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N}\lambda_n\cdot\lambda_m\cdot t_n\cdot t_m\cdot\phi^T(x_n)\cdot\phi(x_m)$$

$$\text{s.t. } \sum_{n=1}^{N}\lambda_n t_n = 0$$

$$\lambda_n = C - \mu_n$$

(f) The explicit form of the kernel is:

$$k(x_n, x_m) = \begin{bmatrix} \cos(x_n^{(1)} - x_n^{(2)}) \\ \sin(x_n^{(1)} - x_n^{(2)}) \end{bmatrix}^T \begin{bmatrix} \cos(x_m^{(1)} - x_m^{(2)}) \\ \sin(x_m^{(1)} - x_m^{(2)}) \end{bmatrix}$$

since $k(x_n, x_m) = \varphi^T(x_n)\,\varphi(x_m)$, and $\varphi(x) = \begin{bmatrix} \cos(x_n^{(1)} - x_n^{(2)}) \\ \sin(x_n^{(1)} - x_n^{(2)}) \end{bmatrix}$

(g) Since we have found λ_n and b we can just use our prediction function $y(x)$ to classify the point. So, to classify x^* in the dual space we will use:

$$\boxed{y(x^*) = \sum_{n=1}^{N} \lambda_n t_n k(x_n, x^*) + b}$$

(h) $\boxed{\text{Case 1}}$: Outside the margin and is correctly classified

In this scenario, we know that:
$$\begin{cases} \lambda_n = 0, \text{ because it is not a support vector \& is correctly classified} \\ \boxed{0 < \xi_n < 1}, \text{ because it's correctly classified} \\ t_n \cdot y(x_n) - 1 + \xi_n = 0, \text{ due to slackness} \\ \boxed{\mu_n = 0}, \text{ due to slackness } (4) \end{cases}$$

Also, from (3), (4) we can say that: $\boxed{\lambda_n = C}$

$\boxed{\text{Case 2}}$: On the margin

We have a support vector in this case, so $\boxed{\lambda_n > 0}$, because $t_n y(x_n) - 1 + \xi_n = 0$.
In addition, $\boxed{\xi_n = 0}$ since we are not crossing the area denoted by the support vector(s). Finally, $\boxed{\mu_n \geq 0}$ and we can't say more about it as $\xi_n = 0$ but μ_n could potentially be 0 as well.

$\boxed{\text{Case 3}}$: On the wrong side of the decision boundary

Since the point is incorrectly classified we know that $\boxed{\xi_n > 1}$ so due to slackness $\boxed{\mu_n = 0}$. Finally, from (3) we can say that $\boxed{\lambda_n = C}$.

(i) We know λ_n so using ③ we know that $\boxed{\mu_n = C - \lambda_n}$. Then, we can use ① to find w:

$$\boxed{w = \sum_{n=1}^{N} \lambda_n t_n \Phi(x_n)}$$

In addition, to find b we need to find a point x_i that is a support vector. Then we will know that for this point it holds that:

$t_n(\Phi^T(x_n)w - b) - 1 = 0$, as $\boxed{\xi_n = 0}$ and $\boxed{\lambda_n > 0}$

In fact, we do this for all the support vectors and solve all of the equations that are similar to the above, in terms of b. Then, we set the average $\bar{b}$. Thus, $\boxed{\bar{b} = (t_n(\Phi^T(x_n)w) - 1)/t_n}$

Finally, for any slack variable we know that it holds:

$\boxed{\xi_n = -t_n(\Phi(x_n)w - b) + 1}$ and $\boxed{\lambda_n = C}$

(j)(ii) To do this transformation we first need to rotate the data. We can see that the data are in ^{a} $-45°$ degrees slope (so it's -1) so if we perform a rotation the data will be much more easily classifiable. So we apply the transformation:

$$R(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

Where $\theta = 45°$, so we get:

$$R(\tfrac{\pi}{4}) = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}$$

Subsequently, we can merely classify the dataset by the function $|x_1 \cdot x_2|$, where the blue class will contain the points the for which $x_1 \cdot x_2 > 0$ and the corresponding points with a negative sign will be assigned to the red class. The whole reason that this works is that after the rotation we centered our data around the core axis' $x=0, y=0$ so just by the sign of a simple multiplication we can conclude about the final classification.

To sum up, $\Phi(X_1, X_2) = |x_1 \cdot x_2| = \begin{cases} x_1 \cdot x_2, & x_1, x_2 \text{ in quadrant I or III} \\ -x_1 \cdot x_2, & x_1 \cdot x_2 \text{ in quadrant II or IV} \end{cases}$

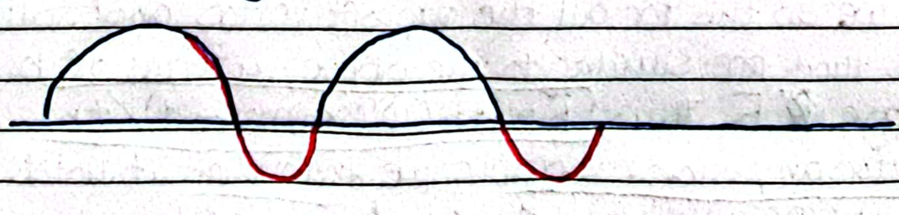
(where the dataset has been rotated by $45°$

(ii) First of all, we need to make this problem simpler and to do that we willl reduce the problem to 1D. We observe that similar to the previous case, our data fall onto $y=-x$ where the intercept b varies. So, we will reduce the problem to 1D by transforming every point with $b = x_1 + x_2$. For every point we will sum its coordinates and let this summation describe its location in 1D, where this space is arranged by the intercept. We can visualize the new dataspace below:

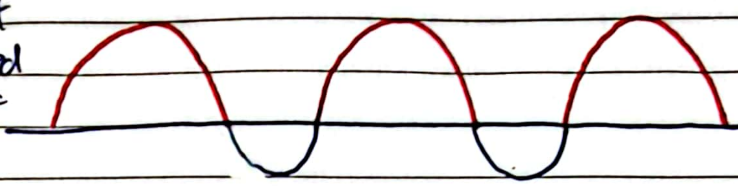
However, as we can see from the plotted points, the data are (naturally) still entangled so we must find a way to separate them. We will do this using a piecewise function of cosine functions with different periods. Note that we will make the assumption that the "width" of the blue data is $3w$ and the "width" of the red data is $5w$. Essentially this means that the period for the red data will be $\frac{2\pi}{3w}$ (remember $T_{period} = \frac{2\pi}{k}$) and the period for the blue data will be $\frac{2\pi}{5w}$.

$$f(x) = \begin{cases} \cos\left(\frac{2\pi}{5w}x\right), & \text{when } x = \frac{3w}{4}(1+2n),\ n \in \mathbb{Z} \\ \cos\left(\frac{2\pi}{3w}x\right), & \text{when } x = \frac{5w}{4}(1+2n),\ n \in \mathbb{Z} \end{cases}$$

The conditions show where $\frac{2\pi}{3w}x$ and $\frac{2\pi}{5w}x$ are equal to $\frac{\pi}{2} + n\pi$, so essentially they are the coordinates of the points where the cosine functions are becoming 0. Graphically, we are doing the following:



(assume that we have tuned the phases of the cosines accordingly



(these will eventually meet because even though we shift between them they keep moving in 'time' even when we don't use them)

But, we need to add something to this solution, as the two conditions will eventually be true simultaneously.(1) In that case, we need to define a state function S which will be 1 if we used $\cos\left(\frac{2\pi}{5w}x\right)$ in t_{i-1} where t_i is when both are zero, and will be 0 if we used $\cos\left(\frac{2\pi}{3w}x\right)$. Then, the function becomes:

$$f(x) = \begin{cases} \cos\left(\frac{2\pi}{5w}x\right), \text{ when } x = \frac{3w}{4}(1+2n),\ n\in\mathbb{Z} \\ \cos\left(\frac{2\pi}{3w}x\right), \text{ when } x = \frac{5w}{4}(1+2n),\ n\in\mathbb{Z} \\ \cos\left(\frac{2\pi}{5w}x\right),\ \left(x \bmod \frac{15\pi}{4}\right)=0 \text{ and } S=0 \\ \cos\left(\frac{2\pi}{3w}x\right),\ \left(x \bmod \frac{15\pi}{4}\right)=0 \text{ and } S=1 \end{cases}$$

(1) $\frac{3w}{4}(1+2n) = \frac{5w}{4}(1+2m) \Leftrightarrow \boxed{3n-5m=1}$

Using the Extended Euclidean Algorithm:

$$\boxed{n=2+5k},\ \boxed{m=1+3k}$$

So, for each integer k there is a pair of n, m where the equation is satisfied and thus both cosines will be 0.

Hence, this will be our transformation!

(iii) This case is completely identical to the previous one, if we just shift the sign of the slope, as now we use $y=x$ insted of $y=-x$. Thus, we can do the exact same procedure, reduce the data in 1D by x_1-x_2 and use the following piecewise function:

$$f(x) = \begin{cases} \cos\left(\frac{2\pi}{5w}x\right), \text{ when } x = \frac{3w}{4}(1+2n),\ n\in\mathbb{Z} \\ \cos\left(\frac{2\pi}{3w}x\right), \text{ when } x = \frac{5w}{4}(1+2n),\ n\in\mathbb{Z} \\ \cos\left(\frac{2\pi}{5w}x\right), \text{ when } \left(x \bmod \frac{15\pi}{4}\right)=0 \text{ and } S=0 \\ \cos\left(\frac{2\pi}{3w}x\right), \text{ when } \left(x \bmod \frac{15\pi}{4}\right)=0 \text{ and } S=1 \end{cases}$$

And that will be our transformation for this case as well.