SVM Problems Solutions:

1)

1 Show that a linear separator can always be found in this embedded space, regardless of radius and where the data is centered

Proof:

It is easy to see that for the original data, consider data that is distributed such that anything within distance ***r*** of the center ***(a, b)*** is of class +1, and anything greater than distance ***r*** of the center is of class -1, there is no linear separator, but the data can be classified correctly with $\text{sign}(r^2 - (x_1 - a)^2 - (x_2 - b)^2)$. And it is easy to find that,

$$\text{sign}(r^2 - (x_1 - a)^2 - (x_2 - b)^2) = \text{sign}\left((r^2 - a^2 - b^2) + 2ax_1 + 2bx_2 + (-x_1^2) + (-x_2^2)\right)$$

So, it is obviously that we have a linear separator for the feature map $\phi(x_1, x_2) = (1, x_1, x_2, x_1x_2, x_1^2, x_2^2)$ with,

$$\underline{w} = \left((r^2 - a^2 - b^2), 2a, 2b, 0, -1, -1\right);\ \text{r, a, b} \in \mathbb{R}$$

So, regardless of radius ***r*** and center ***(a, b)***, there is always a linear separator in this embedded space.

2 show that if there is an ellipsoidal separator, regardless of center, width, orientation a separator can be found in the quadratic feature space using this kernel.

Proof:

As the same reason in the **1**, and we already know the function of ellipsoidal in high dimensional coordinate system, then we have the function of the separator for the original data set:

$$f(x_1, x_2, \dots, x_n) = \text{sign}\left(1^2 - \frac{(x_1 - a_1)^2}{{c_1}^2} - \frac{(x_2 - a_2)^2}{{c_2}^2} - \cdots - \frac{(x_n - a_n)^2}{{c_n}^2}\right)$$

Where $(a_1, a_2, \dots, a_n)$ is the center, n is the dimension.

So it is easy to find that we have a linear separator for the feature map $\phi(x_1, x_2, \dots x_n) = (1, x_1, x_2, \dots, x_n, x_1^2, x_2^2, \dots, x_n^2)$ with,

$$\underline{w} = \left(\left(1^2 - \left(\frac{a_1}{c_1}\right)^2 - \left(\frac{a_2}{c_2}\right)^2 - \cdots - \left(\frac{a_n}{c_n}\right)^2\right), 2\frac{a_1}{c_1^2}, \dots, 2\frac{a_n}{c_n^2}, -\frac{1}{c_1^2}, -\frac{1}{c_2^2}, \dots, -\frac{1}{c_n^2}\right);$$

where a_i, c_i, n are positive real number.

So it is easy to see that there is a separator in the quadratic feature space $\phi(\underline{x^i})$.

2)

Suppose we have two ellipsoidal regions denoted as **A** and **B**, from the proof of question 1, we can have two separate formula to denote the two-dimensional data set that satisfied the following: the positive class lay within one ellipsoidal region and the negative class was everywhere else. So we

can write the following formula:
For **A**: the data within **A** is positive class else negative,

$$f_A(x_1, x_2) = \text{sign}\left(1^2 - \frac{(x_1 - a_1)^2}{{c_1}^2} - \frac{(x_2 - a_2)^2}{{c_2}^2}\right)$$

For **B**: the data within **B** is positive class else negative,

$$f_B(x_1, x_2) = \text{sign}\left(1^2 - \frac{(x_1 - b_1)^2}{{d_1}^2} - \frac{(x_2 - b_2)^2}{{d_2}^2}\right)$$

So we can conclude that we have three situations that satisfied our question.

1. The data point is within **A** and it is not in **B** so we have $f_A > 0$ and $f_B < 0, f_A \cdot f_B < 0$.
2. The data point is not in **A** and it is within **B** so we have $f_A < 0$ and $f_B > 0, f_A \cdot f_B < 0$.
3. The data point is not in **A** and it is not in **B** so we have $f_A < 0$ and $f_B < 0, f_A \cdot f_B > 0$.

So we can have the following relationship to describe the condition mentioned in the question:

$$f'(x_1, x_2) = -\big(f_A(x_1, x_2) \times f_B(x_1, x_2)\big)$$

If $f' > 0$, data point is either in **A** or in **B**, else $f' < 0$.
So,

$$f'(x_1, x_2) = \text{sign}\left(\left(1^2 - \frac{(x_1 - a_1)^2}{{c_1}^2} - \frac{(x_2 - a_2)^2}{{c_2}^2}\right) \times \left(1^2 - \frac{(x_1 - b_1)^2}{{d_1}^2} - \frac{(x_2 - b_2)^2}{{d_2}^2}\right)\right)$$

Therefore, because it is easy to show that $f'(x_1, x_2)$ can be a linear separator for feature map $\phi(x_1, x_2) = \left(1, x_1, x_2, x_1x_2, x_1^2, x_2^2, x_1x_2^2, x_1^2x_2, x_1^3, x_2^3, x_1^2x_2^2, x_1x_2^3, x_1^3x_2, x_1^4, x_2^4\right)$, we can know that the kernel $K(\underline{x}, y) = \left(1 + \underline{x} \cdot y\right)^4 = \phi(x_1, x_2) \cdot \phi(x_1, x_2)$ will recover a separator for this question.

3)

Suppose we have two circle regions denoted as **A** (inner circle with radius ***r*** and center (a_1, b_1)) and **B** (outer circle which contains **A** with radius ***R*** and center (a_2, b_2)), from the proof of question 1, we can have two separate formula to denote the two-dimensional data set that satisfied the following: the positive class lay within **A** and the negative class within circular band between **A** and **B** and positive everywhere else. So we can write the following formula separately:
Just for **A**: the data within **A** is positive class else negative,

$$f_A(x_1, x_2) = \text{sign}(r^2 - (x_1 - a_1)^2 - (x_2 - b_1)^2)$$

For **B**: the data within **B** is positive class else negative,

$$f_B(x_1, x_2) = \text{sign}(R^2 - (x_1 - a_2)^2 - (x_2 - b_2)^2)$$

So we can conclude that we have three situations that satisfied our question.

1. The data point is within **A** and **B** so we have $f_A > 0$ and $f_B > 0, f_A \cdot f_B > 0$.
2. The data point is in the band which means it is not in **A** and it is within **B** so we have $f_A < 0$ and $f_B > 0, f_A \cdot f_B < 0$.
3. The data point is not in **A** and it is not in **B** so we have $f_A < 0$ and $f_B < 0, f_A \cdot f_B > 0$.

So we can have the following relationship to describe the condition mentioned in the question:

$$f'(x_1, x_2) = -\big(f_A(x_1, x_2) \times f_B(x_1, x_2)\big)$$

If $f' > 0$, data point is either in **A** and **B** or not in **A** and not in **B**, else $f' < 0$.
So,

$$f'(x_1, x_2) = \text{sign}\big((r^2 - (x_1 - a_1)^2 - (x_2 - b_1)^2) \times (R^2 - (x_1 - a_2)^2 - (x_2 - b_2)^2)\big)$$

Therefore, because it is easy to show that $f'(x_1, x_2)$ can be a linear separator for feature map $\phi(x_1, x_2) = \big(1, x_1, x_2, x_1x_2, x_1^2, x_2^2, x_1x_2^2, x_1^2x_2, x_1^3, x_2^3, x_1^2x_2^2, x_1x_2^3, x_1^3x_2, x_1^4, x_2^4\big)$, we can know that the kernel $K(\underline{x}, y) = \left(1 + \underline{x} \cdot y\right)^4$ will recover a separator for this question.

4)

We already know that for XOR problems,

$\underline{x}$	y
$(-1, -1)$	-1
$(-1, +1)$	$+1$
$(+1, -1)$	$+1$
$(+1, +1)$	-1

1 $K\left(\underline{x}, \underline{y}\right) = \left(1 + \underline{x} \cdot \underline{y}\right)^2 = \phi(\underline{x}) \cdot \phi\left(\underline{y}\right)$

So we can assume $\underline{x} = \underline{y} = [x_1, x_2]^T$ and use $\underline{a}, \underline{b}$ to replace $\underline{x}, \underline{y}$ in case of confusion and from the lecture note we already know that,

$$\phi(\underline{a}) = \phi(x_1, x_2) = \left[1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, x_1^2, x_2^2\right]^T$$

The objective function which we need to maximize for the dual problem becomes

$$L_D(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2}\sum_{i=1}^{4}\sum_{j=1}^{4} \alpha_i\alpha_j y_i y_j k_{ij}$$

Where

$$k_{ij} = K\left(\underline{x}_i, \underline{x}_j\right) = \left(1 + \underline{x}_i^T \cdot \underline{x}_j\right)^2 = 1 + x_{i1}^2x_{j1}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2} + x_{i2}^2x_{j2}^2 + 2x_{i1}x_{j1} + 2x_{i2}x_{j2}$$

with constraints $\sum_{i=0}^{n} \alpha_i y_i = 0$ and $\forall\, i, \alpha_i \geq 0$.

and we firstly calculate the k_{ij}:

$$k_{11} = K\left(\begin{bmatrix}-1\\-1\end{bmatrix}, \begin{bmatrix}-1\\-1\end{bmatrix}\right) = 1 + 1 + 2 + 1 + 2 + 2 = 9$$

$$k_{12} = K\left(\begin{bmatrix}-1\\-1\end{bmatrix}, \begin{bmatrix}-1\\+1\end{bmatrix}\right) = 1 + 1 - 2 + 1 + 2 - 2 = 1$$

$$k_{13} = K\left(\begin{bmatrix}-1\\-1\end{bmatrix}, \begin{bmatrix}+1\\-1\end{bmatrix}\right) = 1 + 1 - 2 + 1 - 2 + 2 = 1$$

$$k_{14} = K\left(\begin{bmatrix}-1\\-1\end{bmatrix}, \begin{bmatrix}+1\\+1\end{bmatrix}\right) = 1 + 1 + 2 + 1 - 2 - 2 = 1$$

The same way we get $k_{22} = k_{33} = k_{44} = 9$, $k_{21} = k_{23} = k_{24} = k_{31} = k_{32} = k_{34} = k_{41} = k_{42} = k_{43} = 1$.

So,

$$K = \begin{bmatrix} 9 & 1 & 1 & 1 \\ 1 & 9 & 1 & 1 \\ 1 & 1 & 9 & 1 \\ 1 & 1 & 1 & 9 \end{bmatrix}$$

So,

$$L_D(\alpha) = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$$
$$-\frac{1}{2}(9\alpha_1^2 - 2\alpha_1\alpha_2 - 2\alpha_1\alpha_3 + 2\alpha_1\alpha_4 + 9\alpha_2^2 + 2\alpha_2\alpha_3 - 2\alpha_2\alpha_4 + 9\alpha_3^2$$
$$- 2\alpha_3\alpha_4 + 9\alpha_4^2)$$

Optimizing $L_D(\alpha)$ with respect to the Lagrange multipliers yields the following set of simultaneous equations:

$$\frac{\partial L_D(\alpha)}{\partial \alpha_1} = 1 - (9\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4) = 0$$
$$\frac{\partial L_D(\alpha)}{\partial \alpha_2} = 1 - (-\alpha_1 + 9\alpha_2 + \alpha_3 - \alpha_4) = 0$$
$$\frac{\partial L_D(\alpha)}{\partial \alpha_3} = 1 - (-\alpha_1 + \alpha_2 + 9\alpha_3 - \alpha_4) = 0$$
$$\frac{\partial L_D(\alpha)}{\partial \alpha_4} = 1 - (\alpha_1 - \alpha_2 - \alpha_3 + 9\alpha_4) = 0$$

So we get $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \frac{1}{8}$

$$\underline{w} = \sum_{i=1}^{4} \alpha_i y_i \phi(\underline{x_i}) = \left(0,0,0,\frac{1}{\sqrt{2}},0,0\right)$$

The function of separator or boundary is $\underline{w} \cdot \phi(\underline{x})$ which is:

$$-x_1x_2 = 0$$

2 $K\left(\underline{x}, \underline{y}\right) = exp\left(-\left\|\underline{x} - \underline{y}\right\|^2\right) = \phi(\underline{x})\phi\left(\underline{y}\right)$

So we can assume $\underline{x} = \underline{y} = [x_1, x_2]^T$ and use $\underline{a}, \underline{b}$ to replace $\underline{x}, \underline{y}$ in case of confusion

We can expand the kernel by using series $e^n = \sum_{i=0}^{\infty} \frac{n^i}{i!} = 1 + n + \frac{n^2}{2!} + \frac{n^3}{3!} + \cdots$, so

$$K(\underline{a}, \underline{b}) = exp\left(-(a_1^2 + b_1^2 - 2a_1b_1 + a_2^2 + b_2^2 - 2a_2b_2)\right)$$
$$= exp\left(-\|\underline{a}\|^2\right) exp\left(-\|\underline{b}\|^2\right) exp(2\underline{ab})$$

$$exp(2\underline{ab}) = \left(1 + 2\underline{ab} + 4\frac{\underline{ab}^2}{2!} + \cdots\right)$$
$$= \left[\left(1 + \sqrt{2}\underline{a} + 2\frac{\underline{a}^2}{\sqrt{2!}} + \cdots\right) \cdot \left(1 + \sqrt{2}\underline{b} + 2\frac{\underline{b}^2}{\sqrt{2!}} + \cdots\right)\right]$$

So feature map is:

$$\phi(\underline{x}) = exp\left(-\|\underline{x}\|^2\right)\left(1+\sqrt{2}\underline{x}+2\frac{\underline{x}^2}{\sqrt{2!}}+\cdots\right)$$

The objective function which we need to maximize for the dual problem becomes

$$L_D(\alpha) = \alpha_1+\alpha_2+\alpha_3+\alpha_4-\frac{1}{2}\sum_{i=1}^{4}\sum_{j=1}^{4}\alpha_i\alpha_j y_i y_j k_{ij}$$

Where

$$k_{ij} = K\left(\underline{x}_i, \underline{x}_j\right) = exp\left(-\left(\left(x_{i1}-x_{j1}\right)^2+\left(x_{i2}-x_{j2}\right)^2\right)\right)$$

with constraints $\sum_{i=0}^{n}\alpha_i y_i = 0$ and $\forall\, i, \alpha_i \geq 0.$

and we firstly calculate the k_{ij}:

$$k_{11} = K\left(\begin{bmatrix}-1\\-1\end{bmatrix},\begin{bmatrix}-1\\-1\end{bmatrix}\right) = exp\big(-(0+0)\big) = 1$$

$$k_{12} = K\left(\begin{bmatrix}-1\\-1\end{bmatrix},\begin{bmatrix}-1\\+1\end{bmatrix}\right) = exp\big(-(0+4)\big) = e^{-4}$$

$$k_{13} = K\left(\begin{bmatrix}-1\\-1\end{bmatrix},\begin{bmatrix}+1\\-1\end{bmatrix}\right) = exp\big(-(4+0)\big) = e^{-4}$$

$$k_{14} = K\left(\begin{bmatrix}-1\\-1\end{bmatrix},\begin{bmatrix}+1\\+1\end{bmatrix}\right) = exp\big(-(4+4)\big) = e^{-8}$$

The same way we get $k_{22} = k_{33} = k_{44} = 2$, $k_{21} = k_{24} = k_{31} = k_{34} = k_{42} = k_{43} = e^{-4}$, $k_{23} = k_{32} = k_{41} = e^{-8}$.

So,

$$K = \begin{bmatrix} 1 & e^{-4} & e^{-4} & e^{-8} \\ e^{-4} & 1 & e^{-8} & e^{-4} \\ e^{-4} & e^{-8} & 1 & e^{-4} \\ e^{-8} & e^{-4} & e^{-4} & 1 \end{bmatrix}$$

So, we can put it into $L_D(\alpha)$ and optimizing $L_D(\alpha)$ with respect to the Lagrange multipliers yields the following set of simultaneous equations:

$$\frac{\partial L_D(\alpha)}{\partial \alpha_1} = 1-(\alpha_1-e^{-4}\alpha_2-e^{-4}\alpha_3+e^{-8}\alpha_4) = 0$$

$$\frac{\partial L_D(\alpha)}{\partial \alpha_2} = 1-(-e^{-4}\alpha_1+\alpha_2+e^{-8}\alpha_3-e^{-4}\alpha_4) = 0$$

$$\frac{\partial L_D(\alpha)}{\partial \alpha_3} = 1-(-e^{-4}\alpha_1+e^{-8}\alpha_2+\alpha_3-e^{-4}\alpha_4) = 0$$

$$\frac{\partial L_D(\alpha)}{\partial \alpha_4} = 1-(e^{-8}\alpha_1-e^{-4}\alpha_2-e^{-4}\alpha_3+\alpha_4) = 0$$

So we get $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \frac{\left(e^4+1\right)^2}{(e^4-e^{-4})^2}$

So the separator is

$$\sum_i \alpha_i y^i K(\underline{x_i}, \underline{x}) = \frac{(e^4+1)^2}{(e^4-e^{-4})^2}\Big(-exp(-[(x_1+1)^2+(x_2+1)^2]) + exp(-[(x_1+1)^2+(x_2-1)^2]) + exp(-[(x_1-1)^2+(x_2+1)^2]) - exp(-[(x_1-1)^2+(x_2-1)^2])\Big) = 0$$

so the bounder is

$$-exp(-[(x_1+1)^2+(x_2+1)^2]) + exp(-[(x_1+1)^2+(x_2-1)^2]) + exp(-[(x_1-1)^2+(x_2+1)^2]) - exp(-[(x_1-1)^2+(x_2-1)^2]) = 0$$

So, we can defined the preference by using the margin
For Polynomial Kernel:
Margin is

$$\gamma(\underline{w}) = \min_i \frac{|\underline{w} \cdot \underline{x_i}|}{\|\underline{w}\|} = \sqrt{2}$$

For gaussian kernel, we have $K(\underline{x_i}, \underline{x_i}) = \|\phi(\underline{x_i})\|^2 = exp\left(-\|\underline{x_i} - \underline{x_i}\|^2\right) = 1$ for i = 1,2,3,4 , $|\underline{w} \cdot \underline{x_i}| = 1$.
But I don't know how to solve $\underline{w}$.
But I assume that the polynomial kernel is better.