Task 1

Linear separability Linear separability describes the fact that two data sets can be separated by a hyperplane such that the data sets are on the opposite sites of the hyperplane and no element of one set is on the site of the other set.

Slack variables Slack variables are introduced to describe if and how data points are 'wrongly' assigned, e. g. if they are closer to the separating hyperplane than $\frac{1}{||w||}$. I. e. it can be between the separating hyperplane and the support vectors or it can also be completely on the other site with a distance of $> \frac{2}{||w||}$. Slack variables allow to use linear separation with penalty (cf. the trade-off between margin and number of mistakes on training data).

Kernel functions In order to separate two data sets, sometimes the dimension of the data space has to be increased, e. g. from 2D to 3D. This is done via a feature transformation function ϕ. The kernel is just the dot product of the feature transformation vectors $\phi^\top \phi$ and can be used to transform the data into higher dimensional space instead of directly using the feature transformation itself. So the kernel functions transform the data space into higher dimensional space and is also responsible for the distribution of the data in this space.

Task 2

1.)

Perceptron classifier function: $\hat{f}(x_i) = w^\top x_i$

2.)

$$\begin{aligned} \hat{f}(x_0)y_0 &= (\begin{pmatrix}1 & -1 & 0.5\end{pmatrix} \begin{pmatrix}1\\0\\0\end{pmatrix}) \cdot (-1) \\ &= -1 \end{aligned}$$

$$\begin{aligned} w &= \begin{pmatrix}1\\-1\\0.5\end{pmatrix} - 0.6 \begin{pmatrix}1\\0\\0\end{pmatrix} \\ &= \begin{pmatrix}0.4\\-1\\0.5\end{pmatrix} \end{aligned}$$

Start over with $i = 0$

$$\begin{aligned} \hat{f}(x_0)y_0 &= (\begin{pmatrix}0.4 & -1 & 0.5\end{pmatrix} \begin{pmatrix}1\\0\\0\end{pmatrix}) \cdot (-1) \\ &= -0.4 \end{aligned}$$

$$
\begin{aligned}
w &= \begin{pmatrix} 0.4 \\ -1 \\ 0.5 \end{pmatrix} - 0.6 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} -0.2 \\ -1 \\ 0.5 \end{pmatrix}
\end{aligned}
$$

Start over with $i = 0$

$$
\begin{aligned}
\hat{f}(x_0)y_0 &= (\begin{pmatrix} -0.2 & -1 & 0.5 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}) \cdot (-1) \\
&= 0.2
\end{aligned}
$$

Go over to $i = 1$

$$
\begin{aligned}
\hat{f}(x_1)y_1 &= (\begin{pmatrix} -0.2 & -1 & 0.5 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}) \cdot 1 \\
&= 0.3
\end{aligned}
$$

Go over to $i = 2$

$$
\begin{aligned}
\hat{f}(x_2)y_2 &= (\begin{pmatrix} -0.2 & -1 & 0.5 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}) \cdot 1 \\
&= -1.2
\end{aligned}
$$

$$
\begin{aligned}
w &= \begin{pmatrix} -0.2 \\ -1 \\ 0.5 \end{pmatrix} + 0.6 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} 0.4 \\ -0.4 \\ 0.5 \end{pmatrix}
\end{aligned}
$$

Start over with $i = 0$

$$
\begin{aligned}
\hat{f}(x_0)y_0 &= (\begin{pmatrix} 0.4 & -0.4 & 0.5 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}) \cdot (-1) \\
&= -0.4
\end{aligned}
$$

$$\begin{aligned} w &= \begin{pmatrix} 0.4 \\ -0.4 \\ 0.5 \end{pmatrix} - 0.6 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} -0.2 \\ -0.4 \\ 0.5 \end{pmatrix} \end{aligned}$$

Start over with $i = 0$

$$\begin{aligned} \hat{f}(x_0)y_0 &= (\begin{pmatrix} -0.2 & -0.4 & 0.5 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}) \cdot (-1) \\ &= 0.2 \end{aligned}$$

Go over to $i = 1$

$$\begin{aligned} \hat{f}(x_1)y_1 &= (\begin{pmatrix} -0.2 & -0.4 & 0.5 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}) \cdot 1 \\ &= 0.3 \end{aligned}$$

Go over to $i = 2$

$$\begin{aligned} \hat{f}(x_2)y_2 &= (\begin{pmatrix} -0.2 & -0.4 & 0.5 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}) \cdot 1 \\ &= -0.6 \end{aligned}$$

$$\begin{aligned} w &= \begin{pmatrix} -0.2 \\ -0.4 \\ 0.5 \end{pmatrix} + 0.6 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 0.4 \\ 0.2 \\ 0.5 \end{pmatrix} \end{aligned}$$

Start over with $i = 0$

$$\begin{aligned} \hat{f}(x_0)y_0 &= (\begin{pmatrix} 0.4 & 0.2 & 0.5 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}) \cdot (-1) \\ &= -0.4 \end{aligned}$$

$$w = \begin{pmatrix} 0.4 \\ 0.2 \\ 0.5 \end{pmatrix} - 0.6 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$
$$= \begin{pmatrix} -0.2 \\ 0.2 \\ 0.5 \end{pmatrix}$$

Start over with $i = 0$

$$\hat{f}(x_0)y_0 = (\begin{pmatrix} -0.2 & 0.2 & 0.5 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}) \cdot (-1)$$
$$= 0.2$$

Go over to $i = 1$

$$\hat{f}(x_1)y_1 = (\begin{pmatrix} -0.2 & 0.2 & 0.5 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}) \cdot 1$$
$$= 0.3$$

Go over to $i = 2$

$$\hat{f}(x_2)y_2 = (\begin{pmatrix} -0.2 & 0.2 & 0.5 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}) \cdot 1$$
$$= 1$$

Go over to $i = 3$

$$\hat{f}(x_2)y_2 = (\begin{pmatrix} -0.2 & 0.2 & 0.5 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}) \cdot 1$$
$$= 0.5$$

Done.

3.) For XOR the labels are as follows: $y = [-1\ 1\ 1\ -1]^\top$
So, in order for this to work, $\hat{f}(x_o) < 0$ needs to hold. This is achieved by having $w_1 < 0$. Additionally $\hat{f}(x_3) < 0$ also needs to hold. Since $\hat{f}(x_3) = w_1 + w_2 + w_3$ it must hold that $w_2 + w_3 < w_1$ and since $w_1 < 0$ it must hold that either $w_2 < 0$ and/or $w_3 < 0$ (they can't both be ≥ 0, because then $w_2 + w_3 \geq 0 > w_1 \lightning$).
Now, consider $w_2 < 0$. Then the case $x_2 = [1\ 1\ 0]^\top$ fails, because $\hat{f}(x_2)y_2 < 0$.
Consider the other case $w_3 < 0$, there we have the same problem: $x_1 = [1\ 0\ 1]^\top$ fails, because $\hat{f}(x_1)y_1 < 0$.
So, there is no weight vector which satisfies all requirements for the XOR function in order for it to work.

Task 3

$$
\begin{aligned}
\langle \phi(x_i), \phi(x_j) \rangle &= \begin{pmatrix} x_{i1}^2 & \sqrt{2}x_{i1}x_{i2} & x_{i2}^2 \end{pmatrix} \begin{pmatrix} x_{i1}^2 \\ \sqrt{2}x_{i1}x_{i2} \\ x_{i2}^2 \end{pmatrix} \\
&= x_{i1}^2 x_{j1}^2 + 2x_{i1}x_{i2}x_{j1}x_{j2} + x_{i2}^2 x_{j2}^2 \\
&= (x_{i1}x_{j1} + x_{i2}x_{j2})^2 \\
&= \left(\begin{pmatrix} x_{i1} & x_{i2} \end{pmatrix} \begin{pmatrix} x_{j1} \\ x_{j2} \end{pmatrix} \right)^2 \\
&= (\langle \phi(x_i), \phi(x_j) \rangle)^2
\end{aligned}
$$

Task 4

e^x can be expressed by its taylor expansion $e^x = \sum\limits_{n=0}^{\infty} \frac{x^n}{n!}$. If we look at the gaussian kernel $k(x, x') = e^{-\frac{||x-x'||^2}{2\sigma^2}}$ and use the taylor expansion instead, we see that $k(x, x') = \sum\limits_{n=0}^{\infty} \frac{||x-x'||^n}{2\sigma^2 n!}$. Since this is a sum over all polynomials from degree 0 to degree ∞, we are also looking at the dot products of infinite dimensional vectors, which makes the RBF a projection to an inifnite dimensional feature space.