Problem 1

1.1

For the hard-SVM in this problem, the Lagrangian function is

$$L(\boldsymbol{w}, w_0, \boldsymbol{\mu}) = \|\boldsymbol{w}\|_2^2 + \sum_{i=1}^{n} \mu_i g_i(x_i) \tag{1}$$

where

$$g_i(x_i) = 1 - y_i(\boldsymbol{w}^T \phi(x_i) + w_0) \tag{2}$$

the optimized $\boldsymbol{w}^*, w_0^*$ requires

$$\begin{aligned} &\nabla_{w_0} L(\boldsymbol{w}^*, w_0^*, \boldsymbol{\mu}) = 0 \\ \Longrightarrow &\sum_{i=1}^{n} \mu_i y_i = 0 \end{aligned} \tag{3}$$

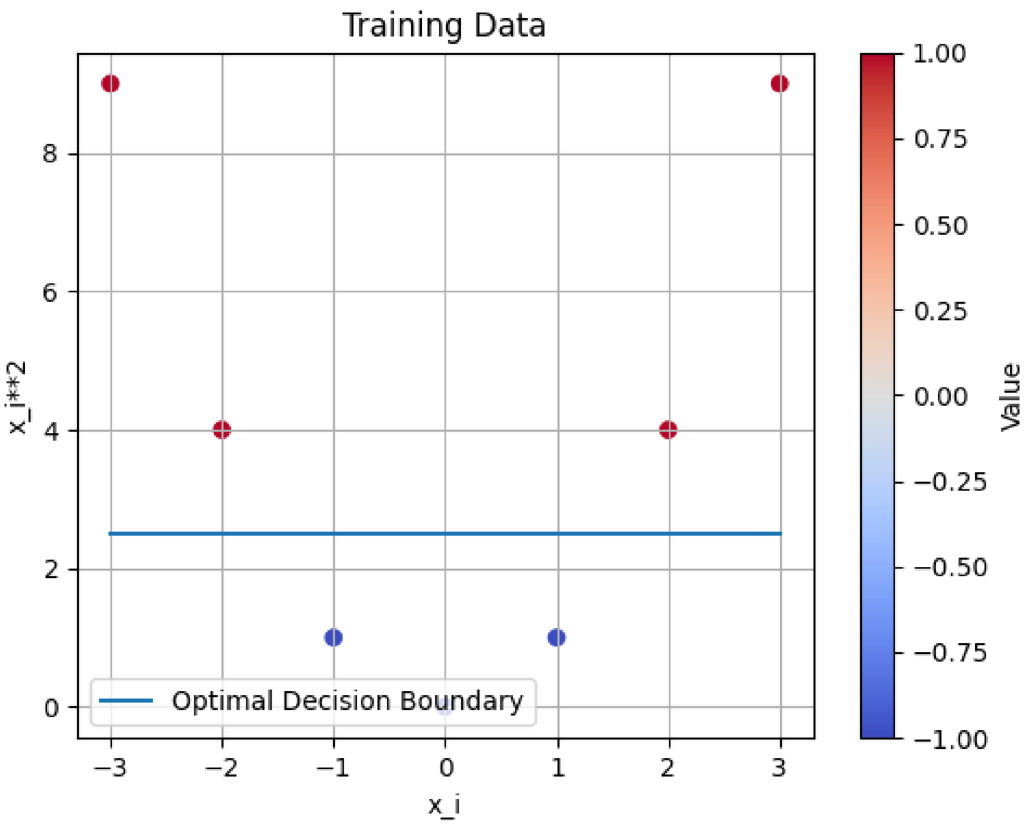
and

$$\begin{aligned} &\nabla_{\boldsymbol{w}} L(\boldsymbol{w}^*, w_0^*, \boldsymbol{\mu}) = 0 \\ \Longrightarrow &\boldsymbol{w}^{*T} - \sum_{i=1}^{n} \mu_i y_i [\phi(x_i)]^T = 0 \end{aligned} \tag{4}$$

There's also the KKT condition

$$\mu_i g_i(x_i) = 0 \tag{5}$$

Using package `cvxpy`, we can finally get the training data and corresponding decision boundary on $\mathbb{R}^2$:

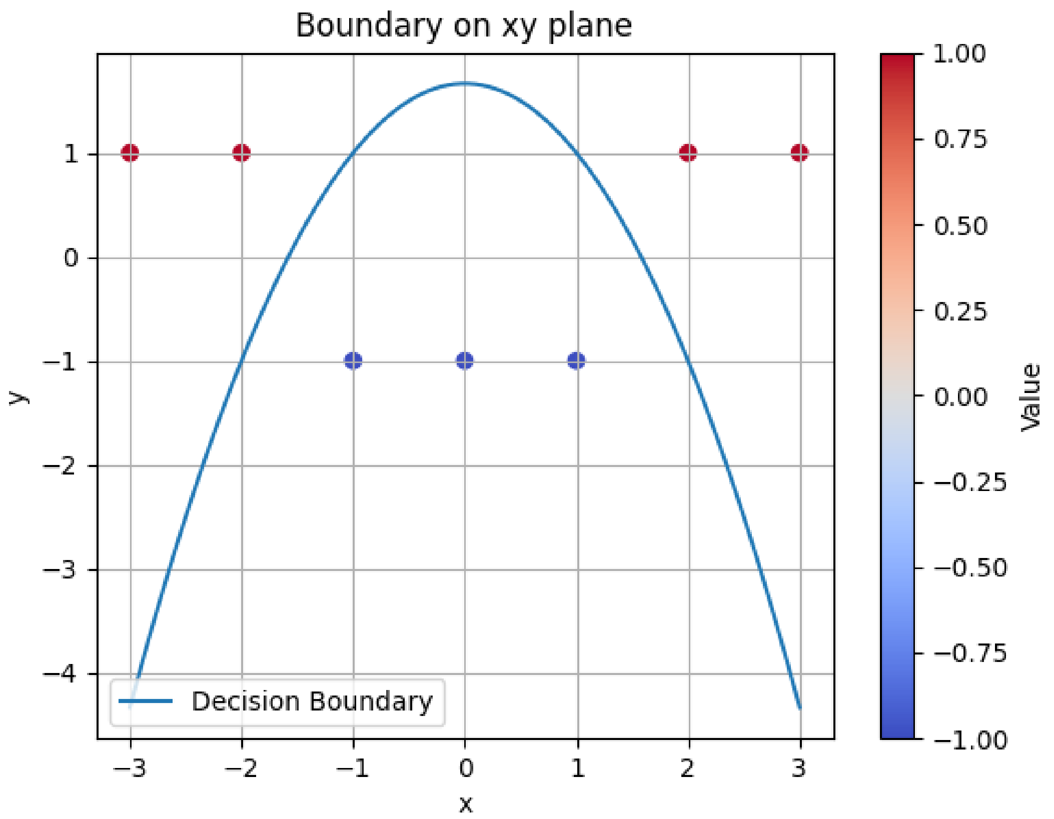


The value of margin we achieved

```
margin = 2.9999999978375125
```

1.2

Similarly, the decision boundary on the (x, y) plane:



1.3

The support vectors of the classifier:

```
support vectors: [[-2  1]
 [-1 -1]
 [ 1 -1]
 [ 2  1]]
```

Problem 2

The Primal Hard-SVM problem can be written as following: consider the Lagrangian

$$L(\boldsymbol{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{w}\|^2 + \sum_{i=1}^{n} \alpha_i [1 - y_i(\boldsymbol{w}^T \cdot \boldsymbol{x}_i + b)] \tag{6}$$

The KKT conditions

$$\begin{gathered} \frac{\partial L}{\partial \boldsymbol{w}} = 0 \\ \frac{\partial L}{\partial b} = 0 \\ \alpha_i \geq 0 \\ \alpha_i [1 - y_i(\boldsymbol{w}^{*T} \cdot \boldsymbol{x}_i + b^*)] = 0 \end{gathered} \tag{7}$$

rewrite the conditions in detail

$$
\begin{aligned}
\boldsymbol{w}^* &= \sum_{i=1}^{n} \alpha_i y_i \boldsymbol{x}_i \\
\sum_{i=1}^{n} \alpha_i y_i &= 0 \\
\alpha_i > 0 &\implies \boldsymbol{w}^{*T} \boldsymbol{x}_i + b = y_i
\end{aligned} \tag{8}
$$

substitute the two equations back to the Lagrangian of the Primal problem

$$
\begin{aligned}
&L(\boldsymbol{w}^*, b, \alpha) \\
=&\frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \sum_{i=1}^{n}\alpha_i - \sum_{i=1}^{n}\alpha_i y_i b - \sum_{i=1}^{n}\alpha_i y_i \boldsymbol{w}^T\boldsymbol{x}_i \\
=&\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_j^T\boldsymbol{x}_i - \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_j^T\boldsymbol{x}_i + \sum_{i=1}^{n}\alpha_i \\
=&\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_j^T\boldsymbol{x}_i
\end{aligned} \tag{9}
$$

which is the option function of the second problem. With the KKT condition, the problem now become the dual problem, i.e. we need to solve the problem below:

$$
\begin{aligned}
&\max_{\boldsymbol{\alpha}} \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_j^T\boldsymbol{x}_i \\
s.t. &\sum_{i=1}^{n}\alpha_i y_i = 0 \\
&\alpha_i \geq 0, \quad 1 \leq i \leq n
\end{aligned} \tag{10}
$$

Whenever we get the solution $\boldsymbol{\alpha}$ of the problem above, we can get that

$$
\begin{aligned}
\boldsymbol{w}^* &= \sum_{i=1}^{n} \alpha_i^* y_i \boldsymbol{x}_i \\
\alpha_i^* > 0 &\implies \boldsymbol{w}^{*T} \boldsymbol{x}_i + b = y_i
\end{aligned} \tag{11}
$$

And if $\boldsymbol{w}^{*T}\boldsymbol{x}_i + b = y_i$, we can obtain that

$$
\begin{aligned}
&\sum_{i=j}^{n}\alpha_j^* y_j y_i \boldsymbol{x}_j^T\boldsymbol{x}_i = 1 \quad \forall i \;\; s.t. \;\boldsymbol{w}^{*T}\boldsymbol{x}_i + b = y_i \\
\implies &\max_{\boldsymbol{\alpha}} \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_j^T\boldsymbol{x}_i \\
&=\frac{1}{2}\sum_{\boldsymbol{w}^{*T}\boldsymbol{x}_i+b=y_i}\alpha_i^*
\end{aligned} \tag{12}
$$

here we've used the contrapositive of the condition $\alpha_i > 0 \implies \boldsymbol{w}^{*T}\boldsymbol{x}_i + b = y_i$. For what we are searching for is the maximum, and the only constraint now is

$$
\sum_{\boldsymbol{w}^{*T}\boldsymbol{x}_i+b=y_i}\alpha_i^* y_i = 0 \qquad y_i \in \{-1, 1\} \tag{13}
$$

we can always find a group of $\alpha_i^* > 0$ to satisfy the constraint, which gives the maximum of the option function. So we finally proved that

$$
\alpha_i^* > 0 \impliedby \boldsymbol{w}^{*T}\boldsymbol{x}_i + b = y_i \tag{14}
$$

i.e. we've proved the relation of the solutions of both problems:

$$
\begin{aligned}
\boldsymbol{w}^* &= \sum_{i=1}^{n} \alpha_i^* y_i \boldsymbol{x}_i \\
\alpha_i^* > 0 &\iff \boldsymbol{w}^{*T} \boldsymbol{x}_i + b = y_i
\end{aligned}
\tag{15}
$$

Q.E.D.

Problem 3

The general parameter groups and the corresponding accuracy are listed below

Kernel Function	Penalty strength	Accuracy
Linear	0.1	0.98
Linear	1	0.96
Linear	10	0.92
Polynomial	0.1	0.94
Polynomial	1	0.93
Polynomial	10	0.96
RBF	0.1	0.82
RBF	1	0.94
RBF	10	0.96
Sigmoid	0.1	0.45
Sigmoid	1	0.47
Sigmoid	10	0.4

It's clear that the linear model with penalty strength `0.1` performs the best, and the corresponding support vectors are in the folder `Support Vector`, which are plotted below:

The reason why they are support vectors might be that these faces are quite clean and mostly laughing. The wrinkles might increases the similarity between the faces between men and women, making these faces the support vectors.