Assignment 2, HT2020
Support Vector Machine - Breast Cancer Detection

Assignment in partial fulfilment of the requirements for the course

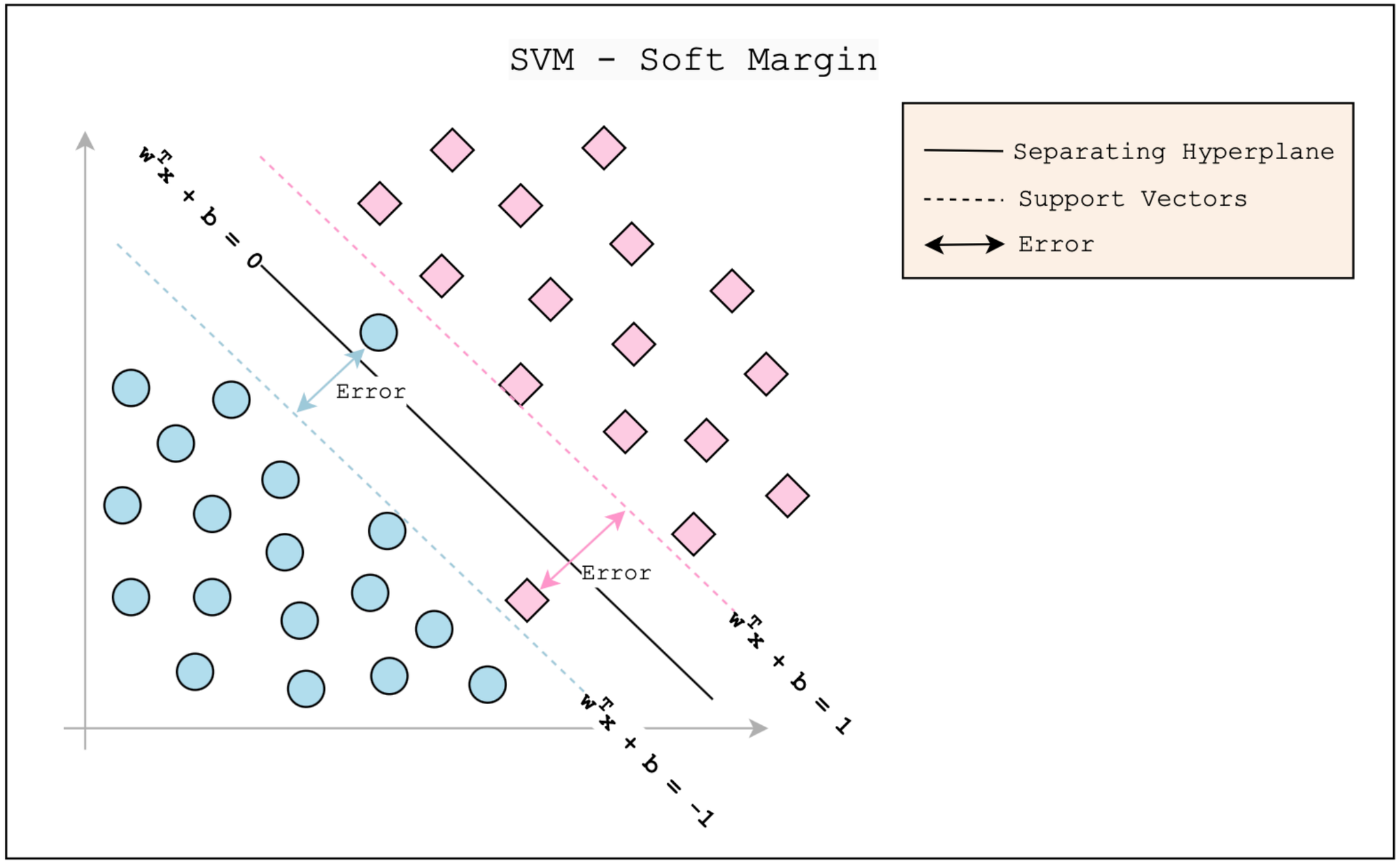
Optimisation 1TD184

UPPSALA
UNIVERSITET

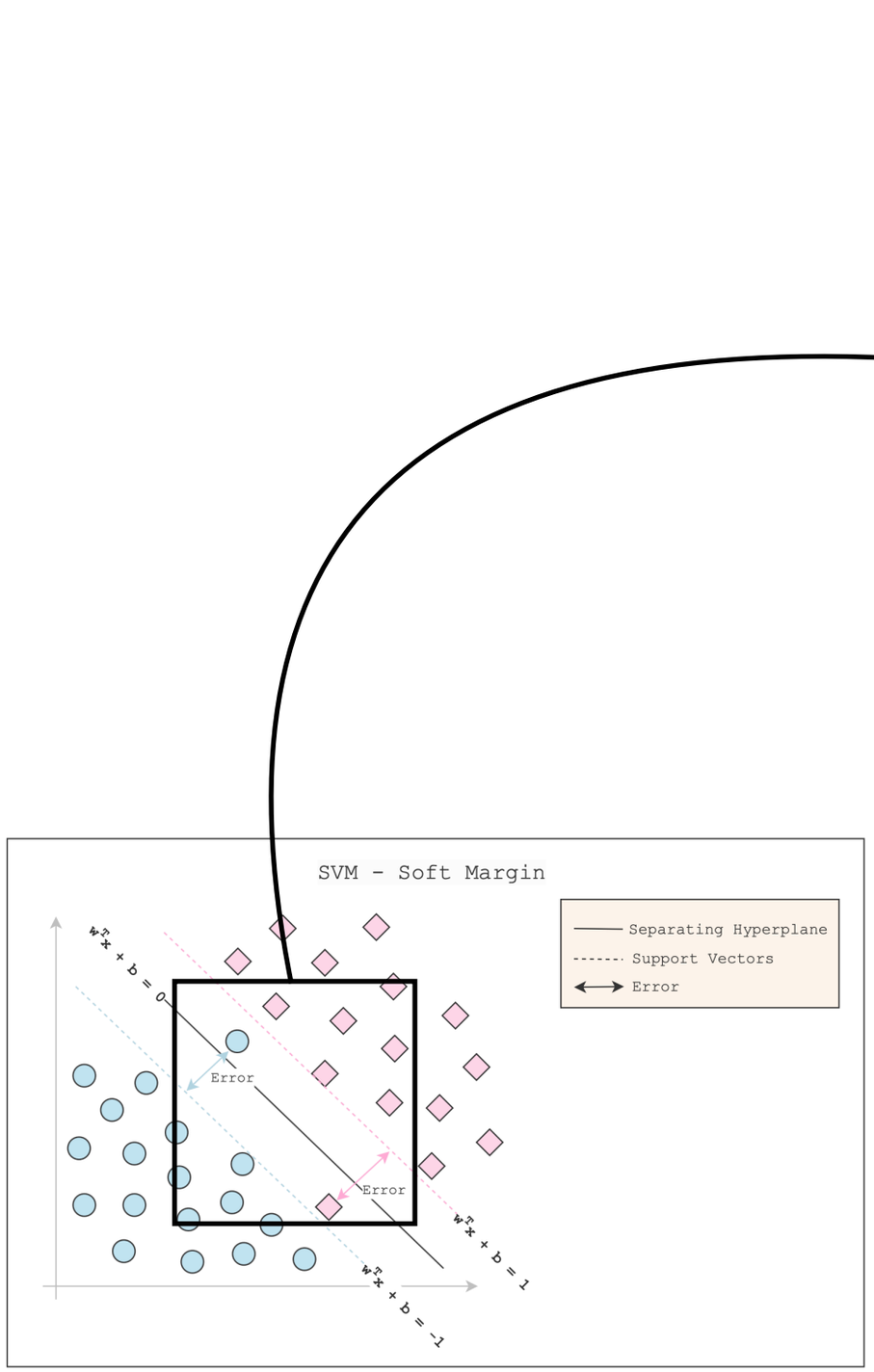
Department of Information Technology

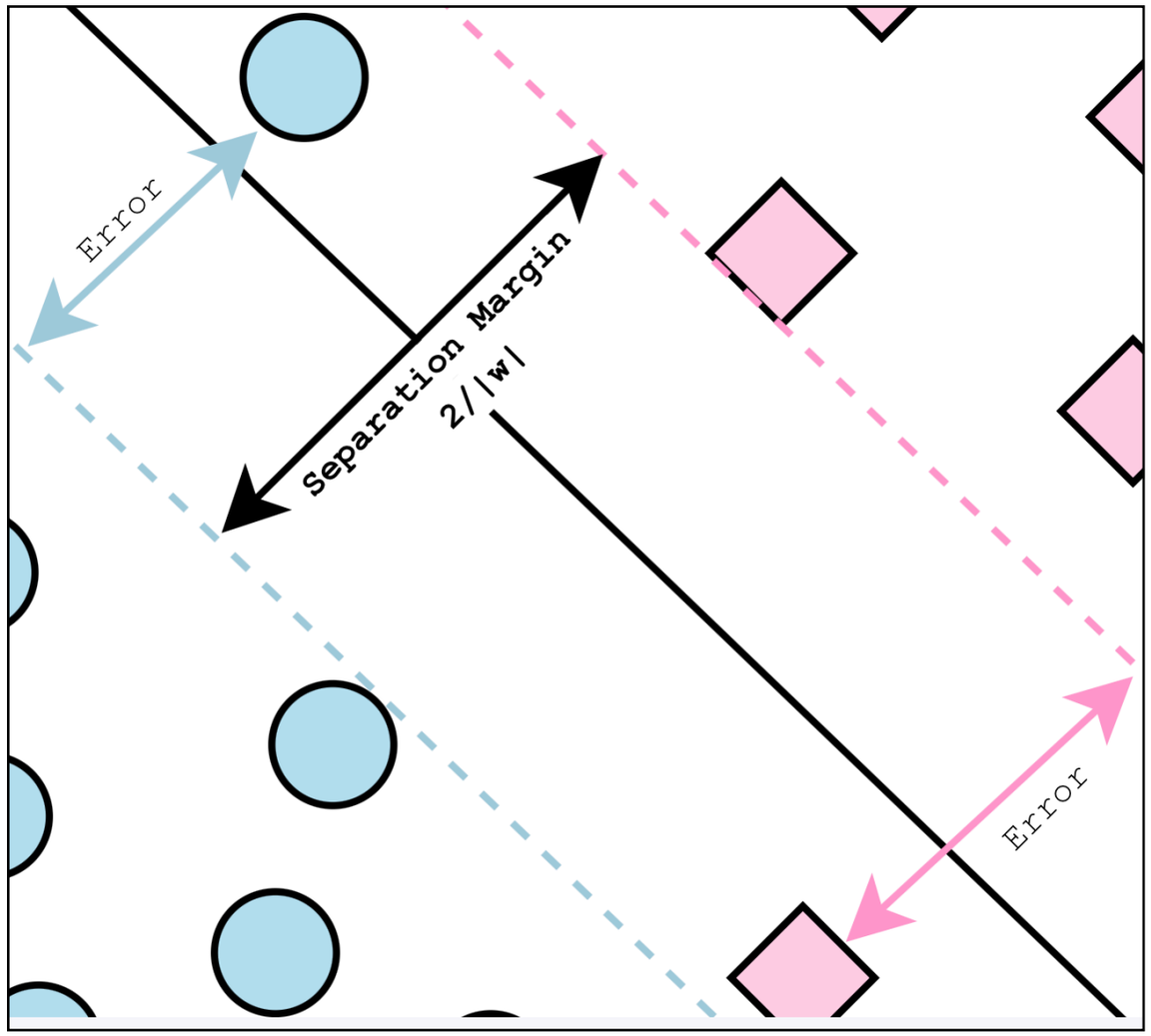
Marcello Pietro Vendruscolo

December 4th, 2020


SVM - Soft Margin
Separating Hyperplane
Support Vectors
Error
$w^T x + b = 0$
Error
$w^T x + b = 1$
$w^T x + b = -1$

Problem





Maximise Separation Margin:

$$minimise f(w,b) = \frac{1}{2} w^T w + C \sum_{i=1}^{m} \xi_i \qquad \text{Eq.(I)}$$

$$s.t. : y_i(w^T x_i + b) \geq 1 - \xi_i, i = 1...m, \xi_i \geq 0$$

1. Solver for quadratic objective functions with linear constraints:

Eq.(II)

$$\min_x \frac{1}{2} x^T H x + f^T x \text{ such that } \begin{cases} A \cdot x \leq b, \\ Aeq \cdot x = beq, \\ lb \leq x \leq ub. \end{cases}$$

2. Function call: `x = quadprog(H,f,A,b,Aeq,Beq,lb,ub,x0,options)`

$$minimise f(w,b) = \frac{1}{2}w^T w + C\sum_{i=1}^{m}\xi_i \quad \texttt{Eq.(I)}$$

$$s.t. : y_i(w^T x_i + b) \geq 1 - \xi_i, i = 1...m, \xi_i \geq 0$$

$$\texttt{Eq.(II)} \quad \min_x \frac{1}{2}x^T Hx + f^T x \text{ such that } \begin{cases} A \cdot x \leq b, \\ Aeq \cdot x = beq, \\ lb \leq x \leq ub. \end{cases}$$

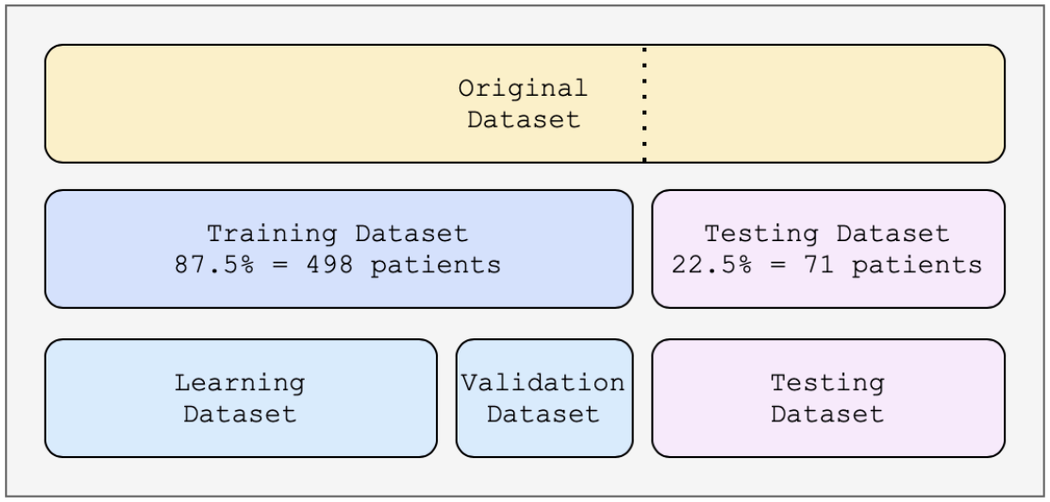
From `Eq.(I)` and `Eq.(II)`:

$$\begin{cases} \frac{1}{2}w^T w = \frac{1}{2}x^T Hx \\ C\sum_{i=1}^{m}\xi_i = f^T x \end{cases}$$

From `Constraints`:

$$\begin{cases} y_i(w^T x_i + b) + \xi_i \geq 1 \\ A \cdot x \leq b \end{cases} \qquad \begin{cases} \xi_i \geq 0 \\ lb \leq x \leq ub \end{cases}$$

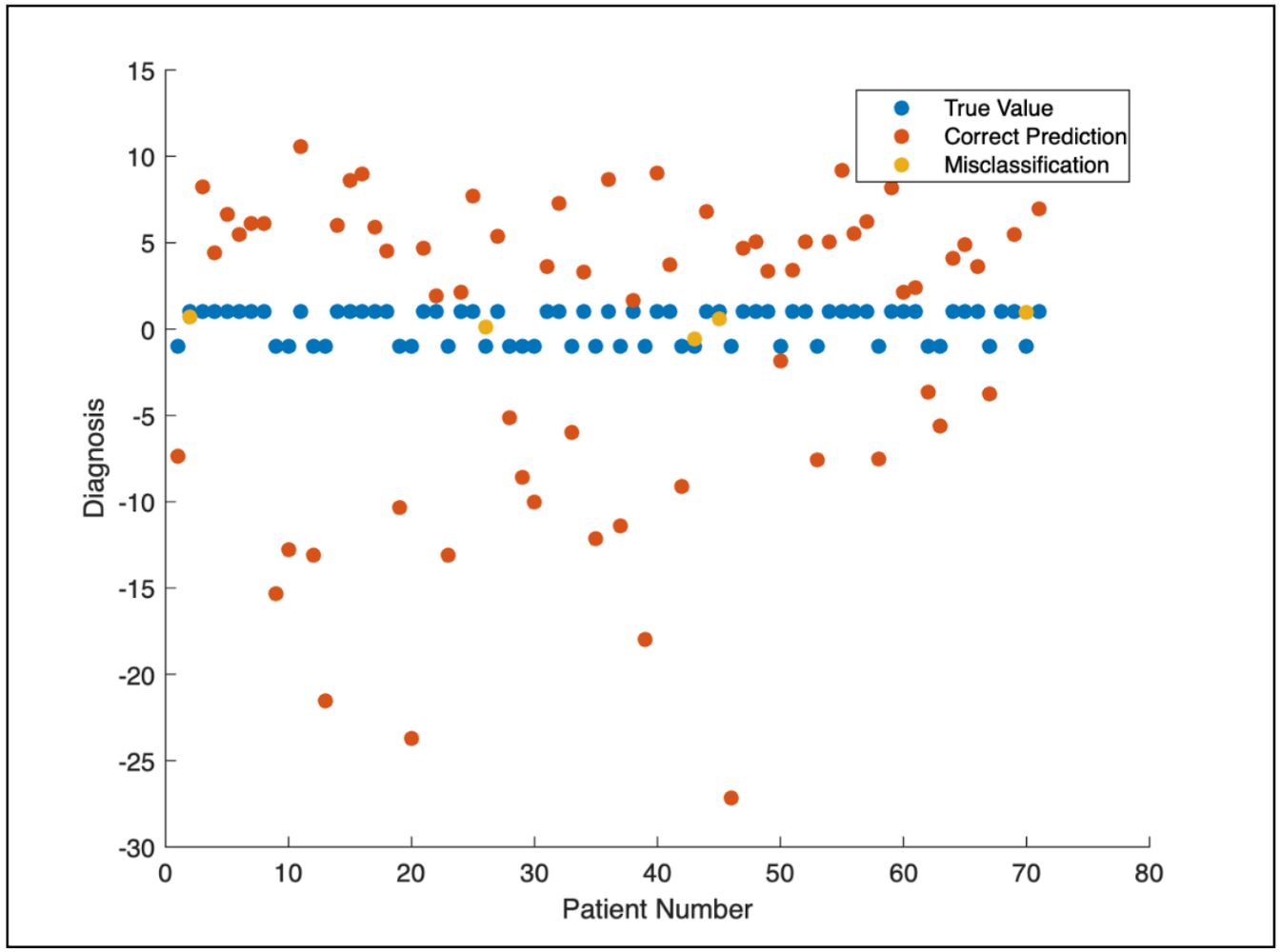
```
H = diag([ones(1, n), zeros(1, l + 1)]);
f = [zeros(1,n), 0, C * ones(1,l)]';

%Constraints
p = diag(y_matrix_training) * x_matrix_training;
A = -[p y_matrix_training eye(l)];
c = -ones(l,1);

%Bound
lb = [-inf * ones(n+1,1); zeros(l,1)];

options = optimoptions(@quadprog,'MaxIterations',500);
z = quadprog(H,f,A,c,[],[],lb,[],[],options);
w = z(1:n,:);
b = z(n+1,:);
eps = z(n+2:end,:);
```



$C = 1000$:

Accuracy ~92.5%

Depends on how the division of the dataset happened (i.e., the randomisation seed)

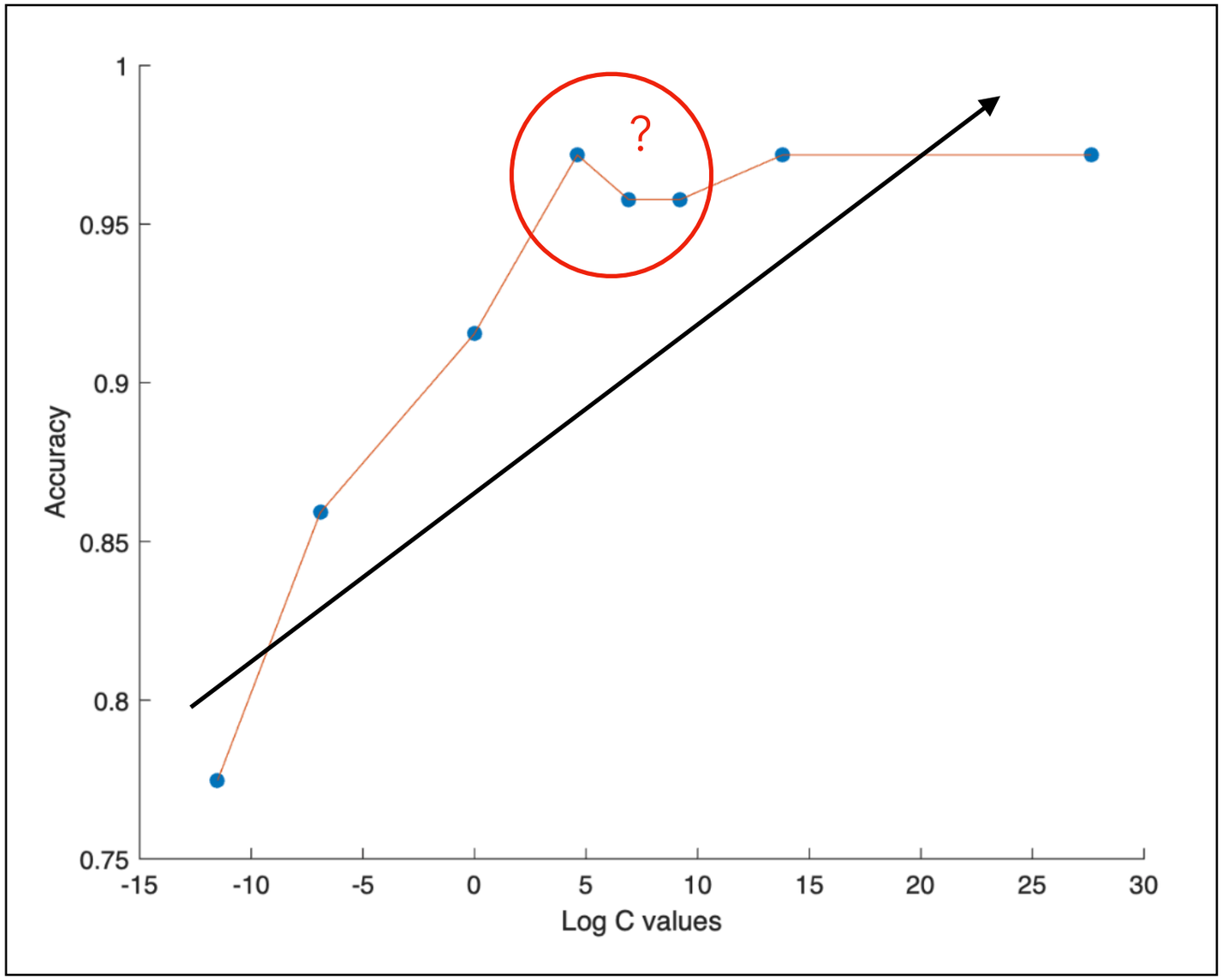


Confusion Matrix

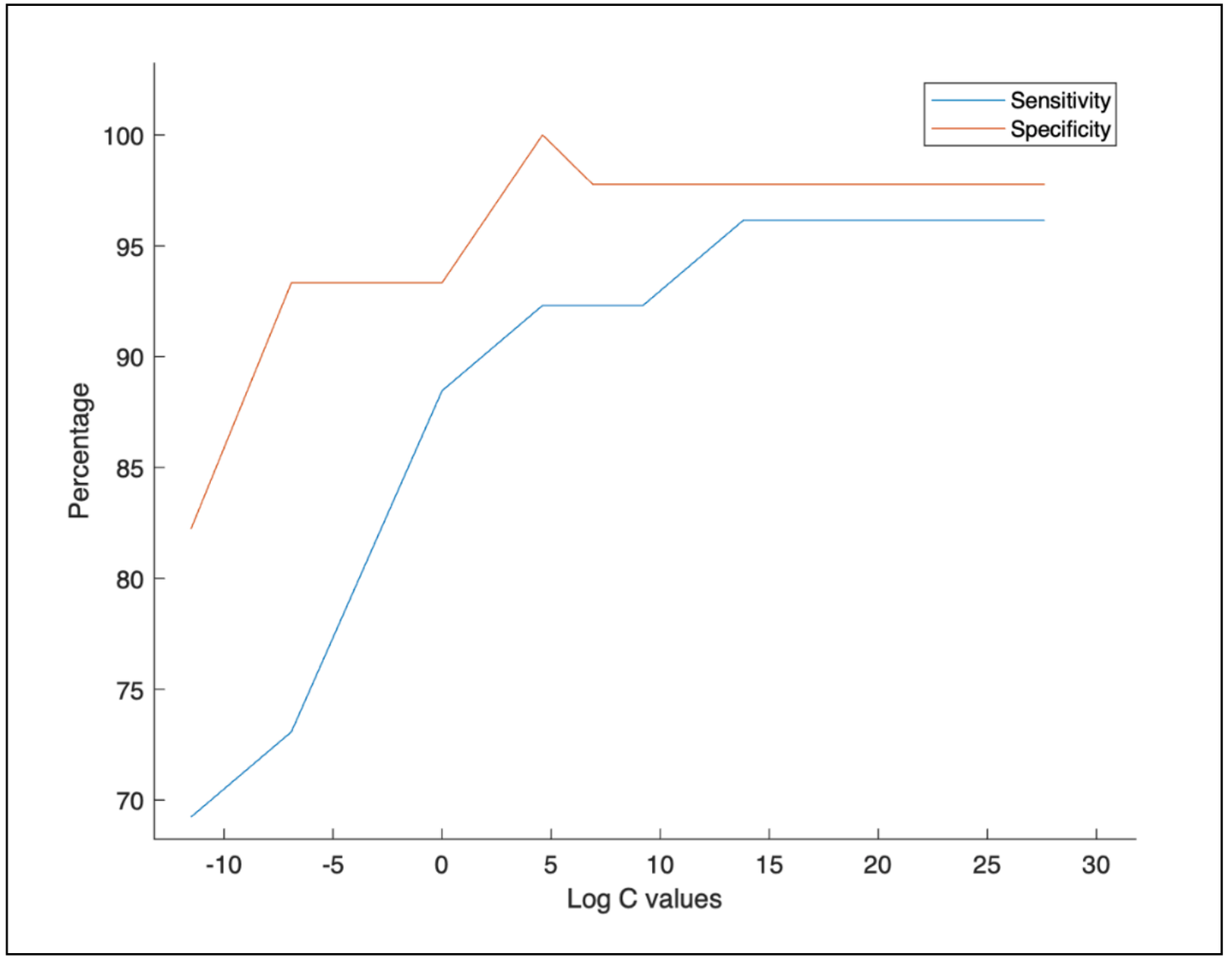
True Condition

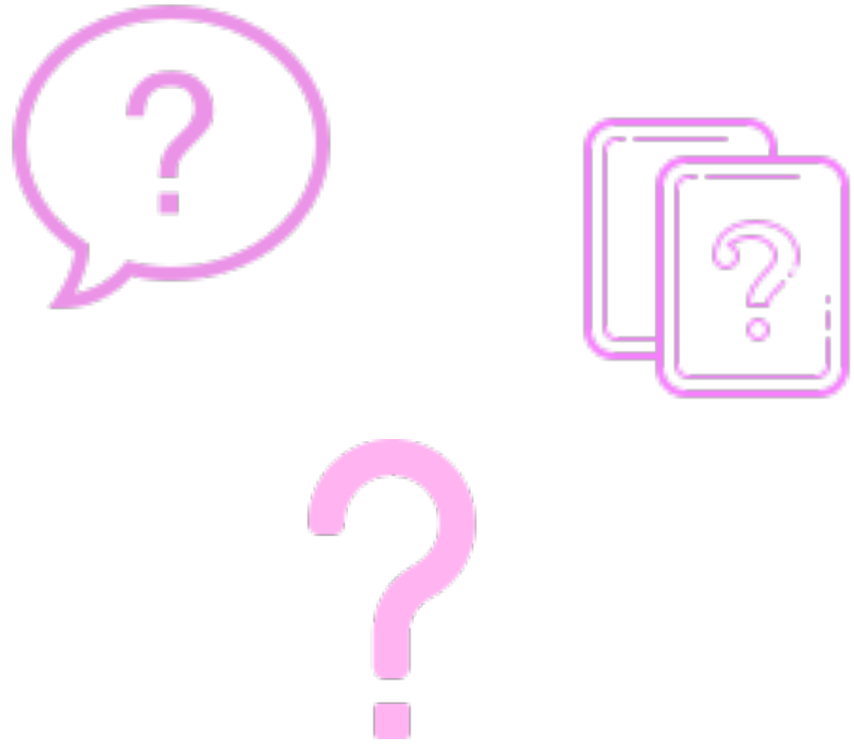
Predicted Condition	Condition Positive (Malignant cells)	Condition Negative (Benign cells)	**Accuracy** $\frac{\sum TruePositive + \sum TrueNegative}{\sum TotalPopulation}$
Predicted condition positive	True Positive	False positive	
Predicted condition negative	False Negative	True Negative	
	Sensitivity $\frac{\sum TruePositive}{\sum ConditionPositive}$	**Specificity** $\frac{\sum TrueNegative}{\sum ConditionNegative}$	

Why can the confusion matrix sometimes be more meaningful than just the misclassification error?


1
0.95
0.9
0.85
0.8
0.75
Accuracy
?
-15
-10
-5
0
5
10
15
20
25
30
Log C values

Task 6


Sensitivity
Specificity
100
95
90
85
80
75
70
Percentage
-10
-5
0
5
10
15
20
25
30
Log C values



Thank you