SVM Tutorial

1. Question and Answer for SVM (Lectures and Homework)

- Supervised Learning Algorithms
    - Decision Tree – Nonlinear decision boundaries
    - Linear SVM – Linear separators
    - Linear SVM with Kernel – Nonlinear boundaries
    - Perceptron – Linear separator
    - Multilayer Perceptron – Nonlinear decision boundaries
    - Logistic Regression – Linear separator

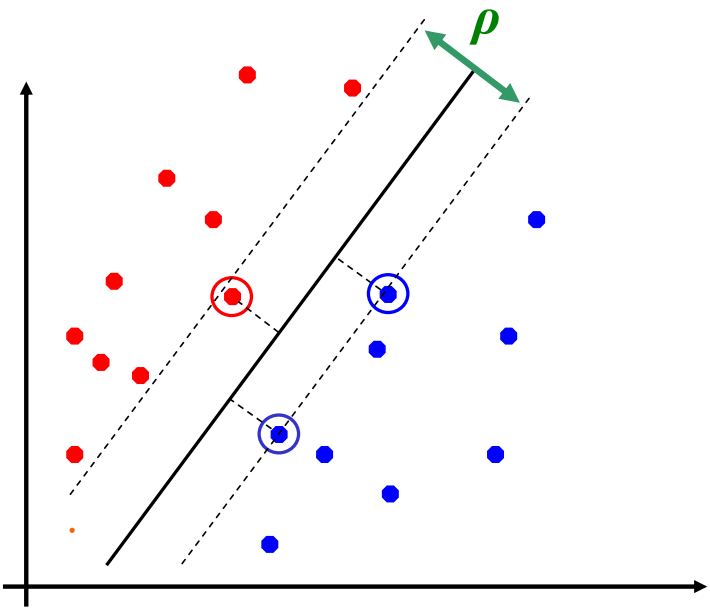
- SVM is a Maximum Margin classifier
    - Maximizing the margin has some proven benefits in terms of generalization performance.

- What is the point of accuracy, precision, and recall?
    - Accuracy is just the number of correct decisions divided by all decisions made
    - Give an example of why this can be misleading?
    - This leads to other metrics such as precision, recall, and many others.

2. Support Vector Machine Algorithm is a Maximum Margin Classifier

SVM finds the hyperplane which:

a. Separates the two classes and
b. Has the maximum margin

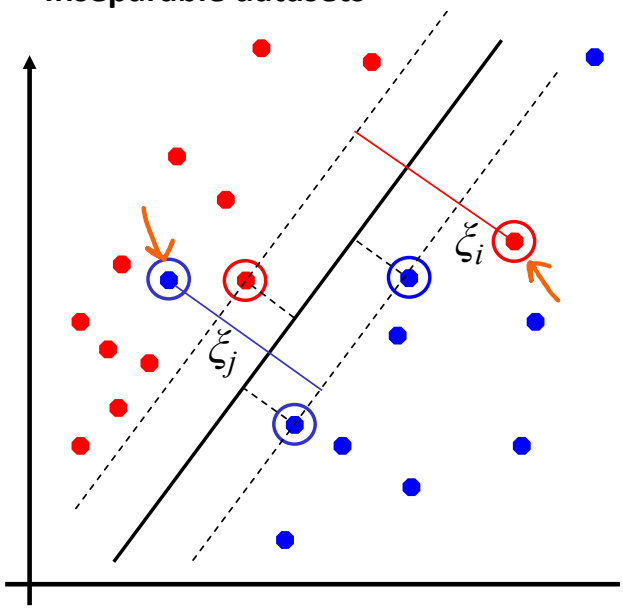


This dataset is
Linearly Separable

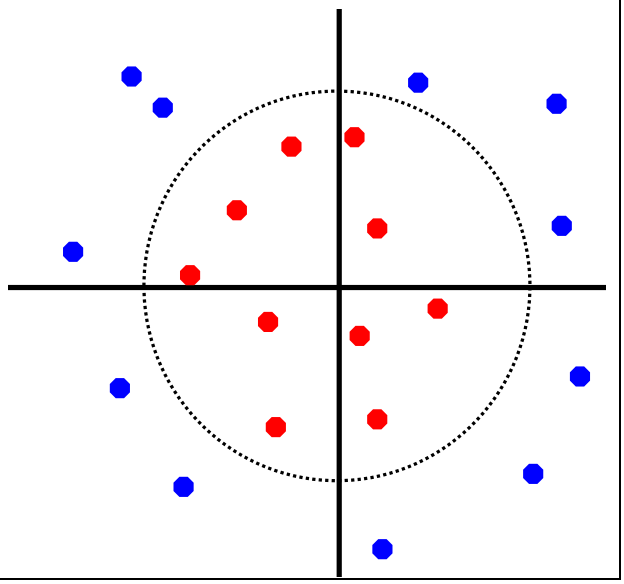
This basic SVM is called the
Hard Margin SVM

Extensions of Hard Margin SVM

a. Soft Margin SVM – Used for Linearly Inseparable datasets



b. Kernel Based SVM – Used for Nonlinear datasets



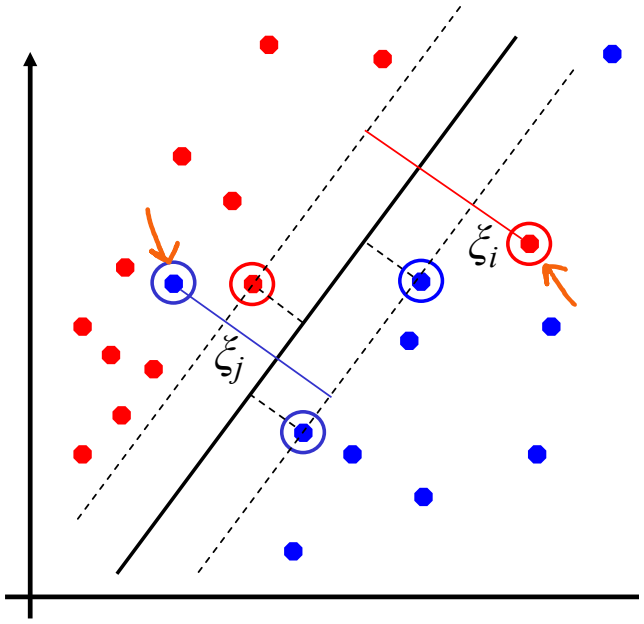
3. Soft Margin SVM

Soft Margin SVM is used for Linearly Inseparable datasets.

It allows for some misclassifications in order to increase the margin.

In Soft Margin SVM, the C parameter controls the penalty for misclassifications relative to maximizing the margin:

a. **Large C means higher penalty for misclassifications → Overfitting**
b. **Small C means smaller penalty for misclassifications → Underfitting**



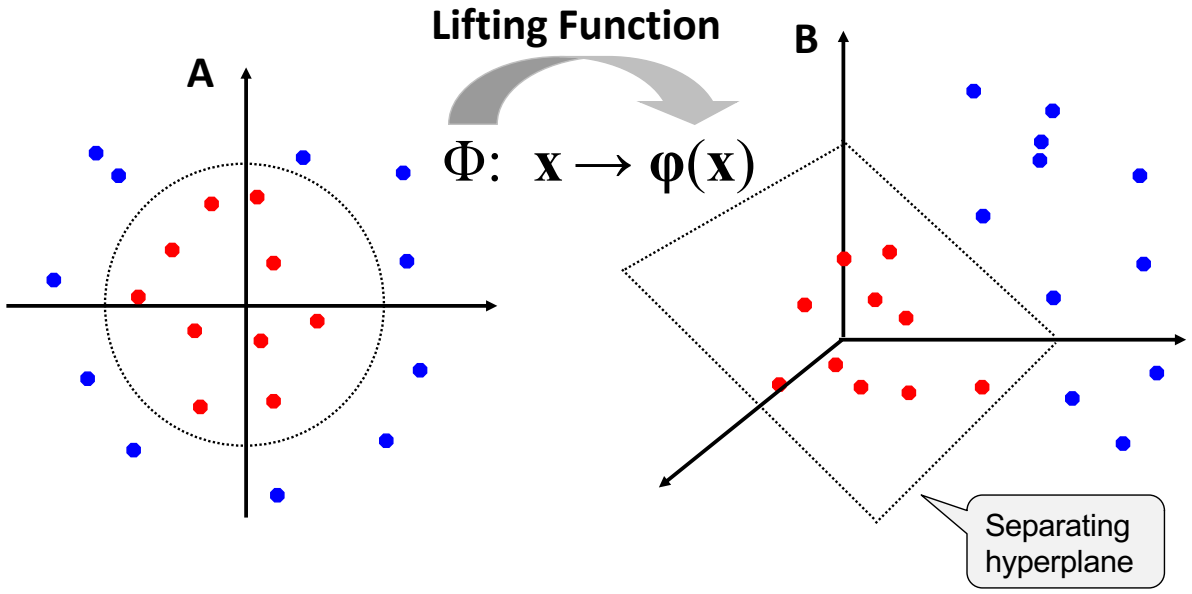
4. SVM with Kernels

SVM with Kernels is used for Nonlinear datasets.

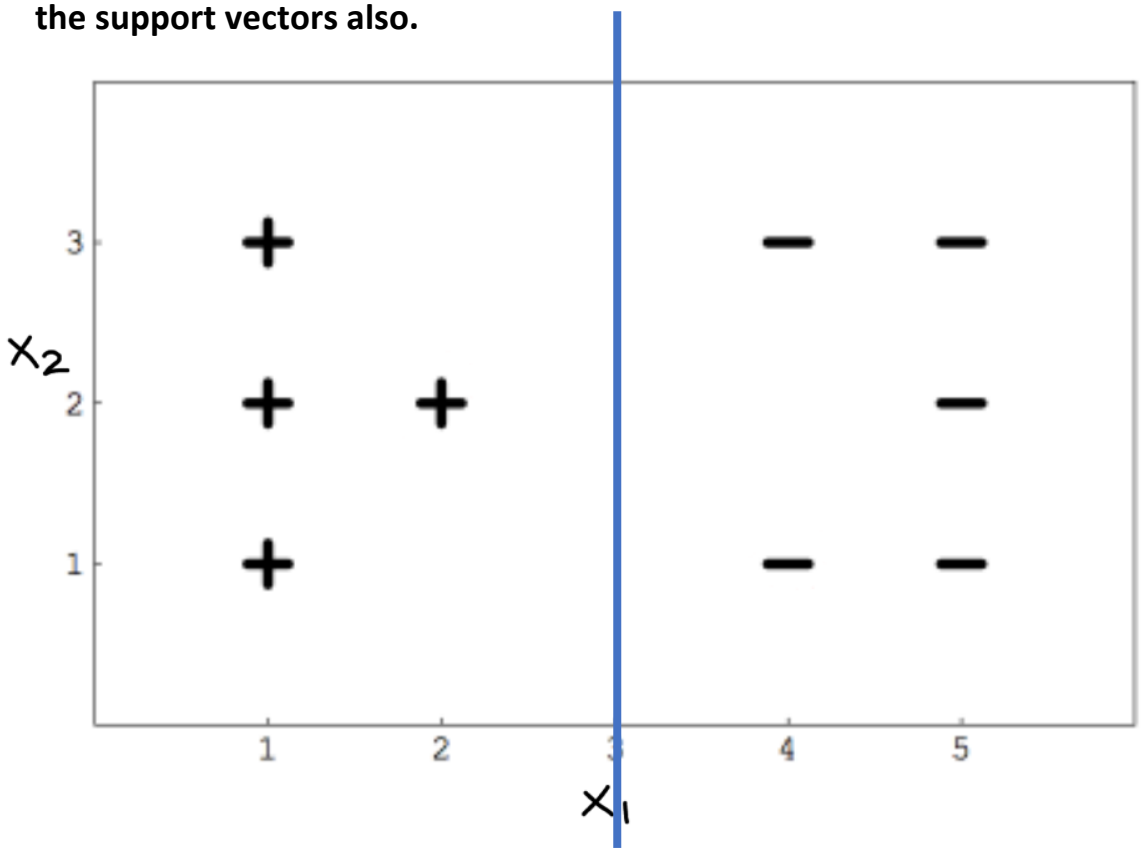
It works by lifting the dataset into a higher dimensional space in which the dataset is possibly linearly separable.

Lifting the dataset into higher dimensional space is achieved by adding features.

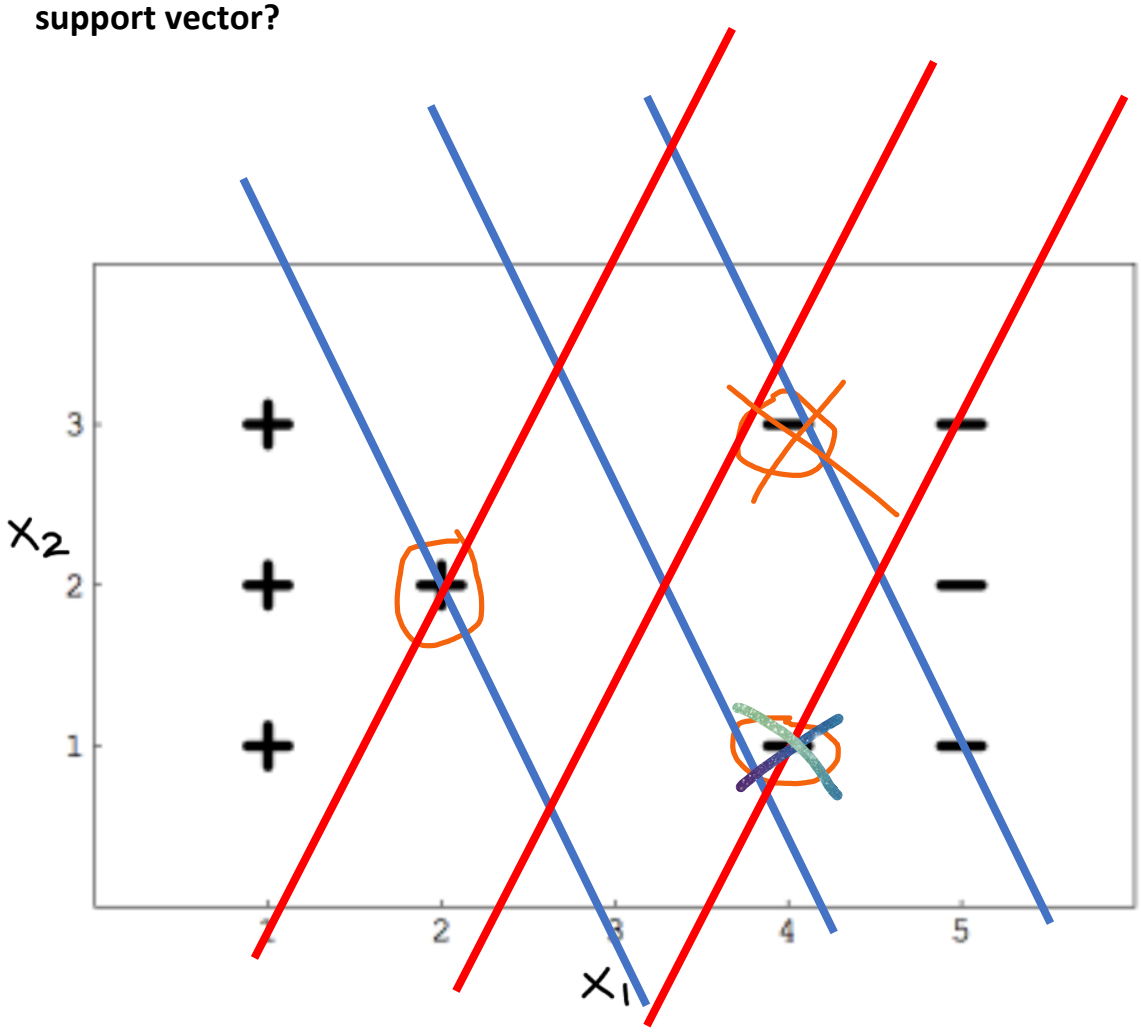
- General idea: the original feature space is mapped to some higher-dimensional feature space where the training set is separable:



5. **Consider the dataset below. What is the Hard Margin SVM solution? Identify the support vectors also.**



6. **Now if we remove any of the support vectors, will the decision boundary change?**
   Will the decision boundary change if we remove a data point that is not a support vector?



$$y = mx + b$$

$y \rightarrow x_2$

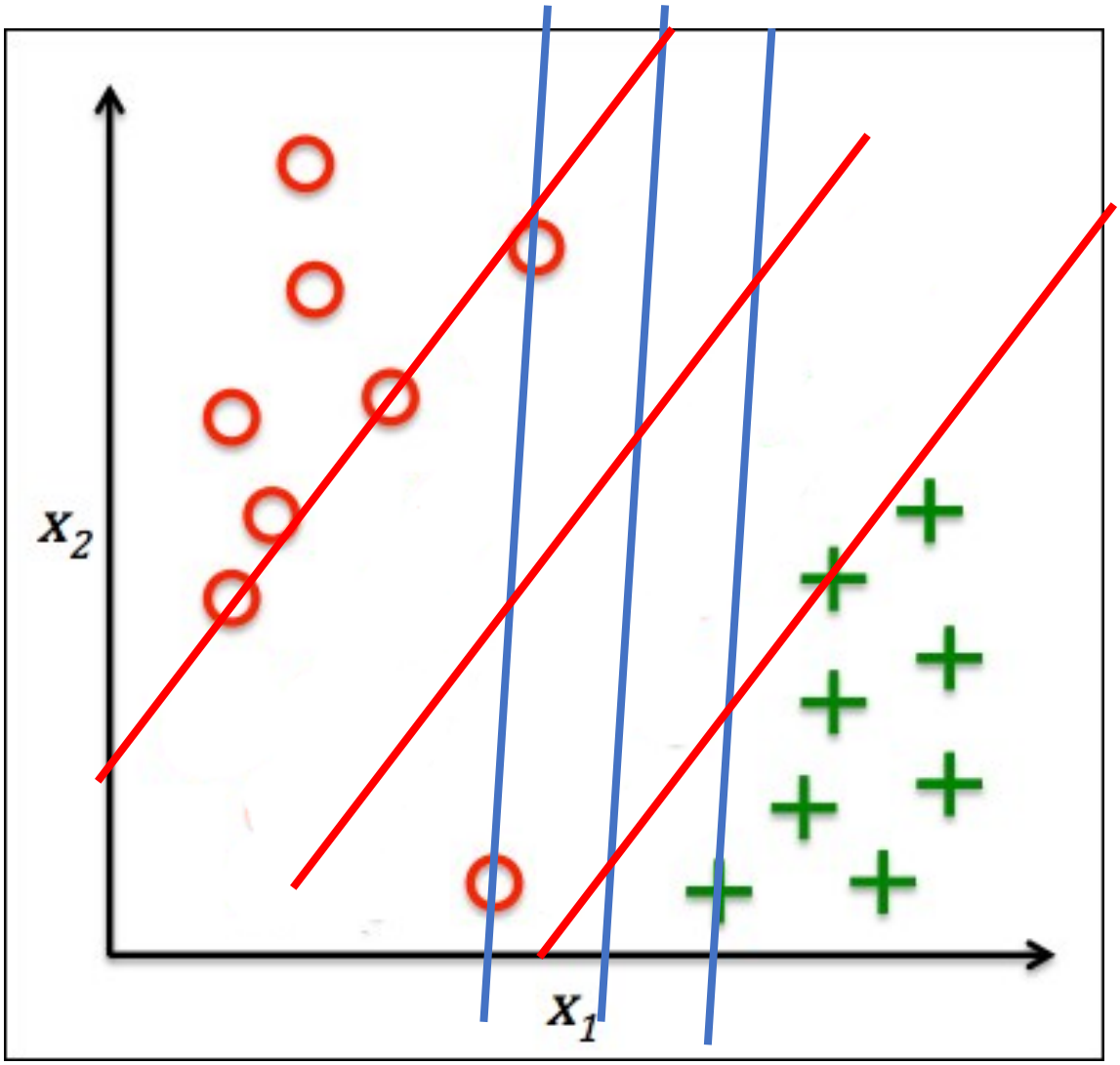
$x \rightarrow x_1$

$$-y + mx + b = 0$$

$$\begin{bmatrix} -1 & m \end{bmatrix} \begin{bmatrix} y \\ x \end{bmatrix} + b = 0$$

7. **Consider the dataset below. Consider using the SVM with soft margin classifier with parameter C.**
    a. **Draw the linear classifier when C is large.**
    b. **Draw the linear classifier when C is small.**

Recall:
Large C means higher penalty for misclassifications → Overfitting
Small C means smaller penalty for misclassifications → Underfitting



Poll Questions:

1. What do you mean by generalization error in context of machine learning algorithms? (Single Choice)
   Answer 1: The threshold amount of error in the model.
   Answer 2: How accurately the model can predict outcomes for unseen data
   Answer 3: How much extra training data is required to train the model perfectly.

2. What do we mean by a hard margin SVM? (Single Choice)
   Answer 1: The SVM allows very low error in classification
   Answer 2: The SVM allows high amount of error in classification
   Answer 3: None of the above

3. When the C parameter is set to infinite, which of the following holds true? (Single Choice)
   Answer 1: The optimal hyperplane if exists, will be the one that completely separates the data.
   Answer 2: The soft-margin classifier will separate the data in all instances.
   Answer 3: None of the above

4. The effectiveness of an SVM depends upon: (Single Choice)
   Answer 1: Selection of Kernel
   Answer 2: Kernel Parameters
   Answer 3: Soft Margin Parameter C
   Answer 4: All of the above

5. What would happen when you use very small C (C~0)? (Single Choice)
   Answer 1: Data will be correctly classified
   Answer 2: Misclassification would happen
   Answer 3: Can't say
   Answer 4: None of these

6. How is your semester going? (Single Choice)
   Answer 1: Awesome
   Answer 2: Okay, but could be better.
   Answer 3: I want to reset!