

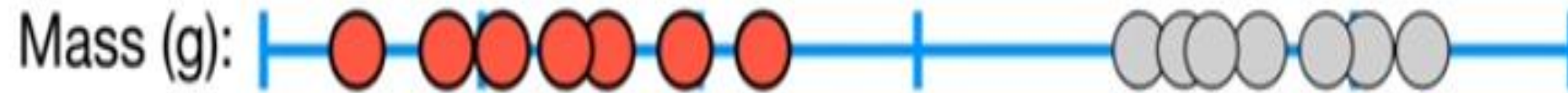
# Support Vector Machine

---

- SVM is a supervised learning model for classification and regression analysis. When it is used for Classification, it is called Support Vector Classifier.
- The algorithm involves finding a hyper-plane in a higher dimensional space which can be used to separate the two different class for binary classification. That is it provides a decision boundary for classifying data points
- The criteria for finding this hyperplane is based on the so-called “widest street approach” that has the largest margin: i.e. largest distance to the nearest training data points of any class
- <https://www.youtube.com/watch?v=N1vOgolbjSc>
- <https://www.youtube.com/watch?v=efR1C6CvhmE>

## Maximum Margin Classifier in 1-dimension

---



The **red dots** represent mice are **not obese**...

## Maximum Margin Classifier in 1-dimension

---

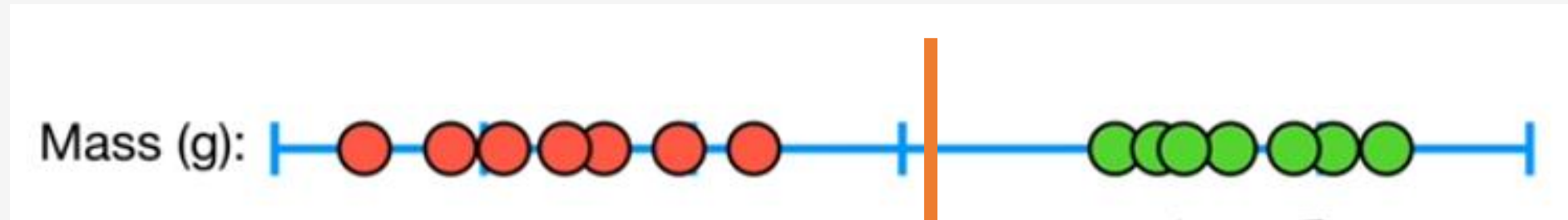


...and the **green dots** represent mice are **obese**.

# Maximum Margin Classifier in 1-dimension

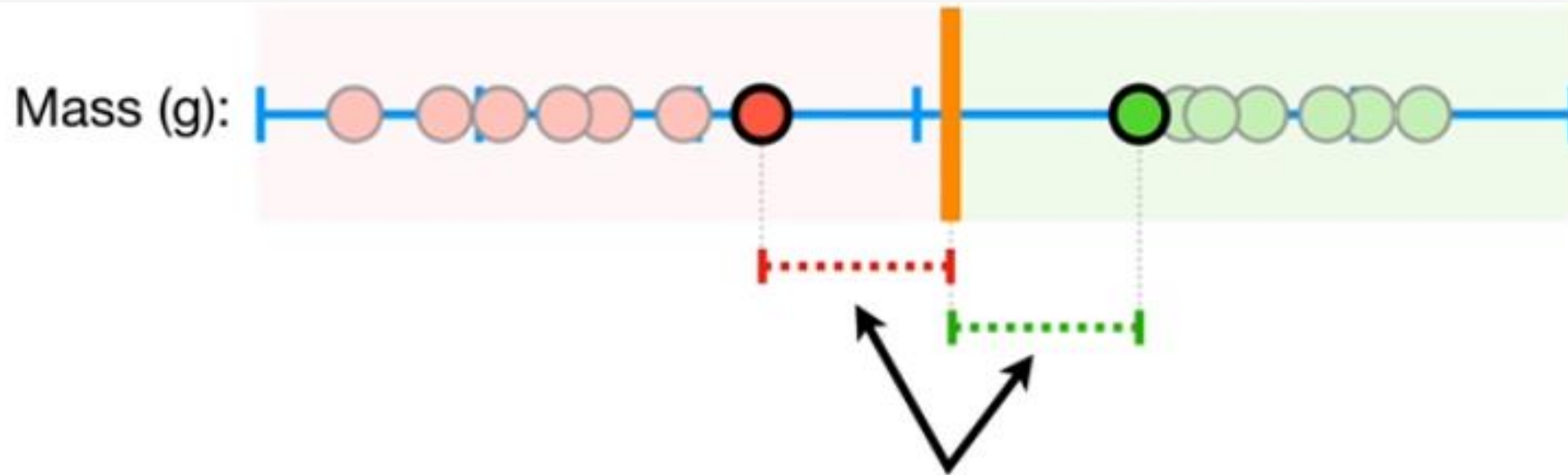
---

How to choose the threshold to decide whether the Mice is obese



## Maximum Margin Classifier in 1-dimension

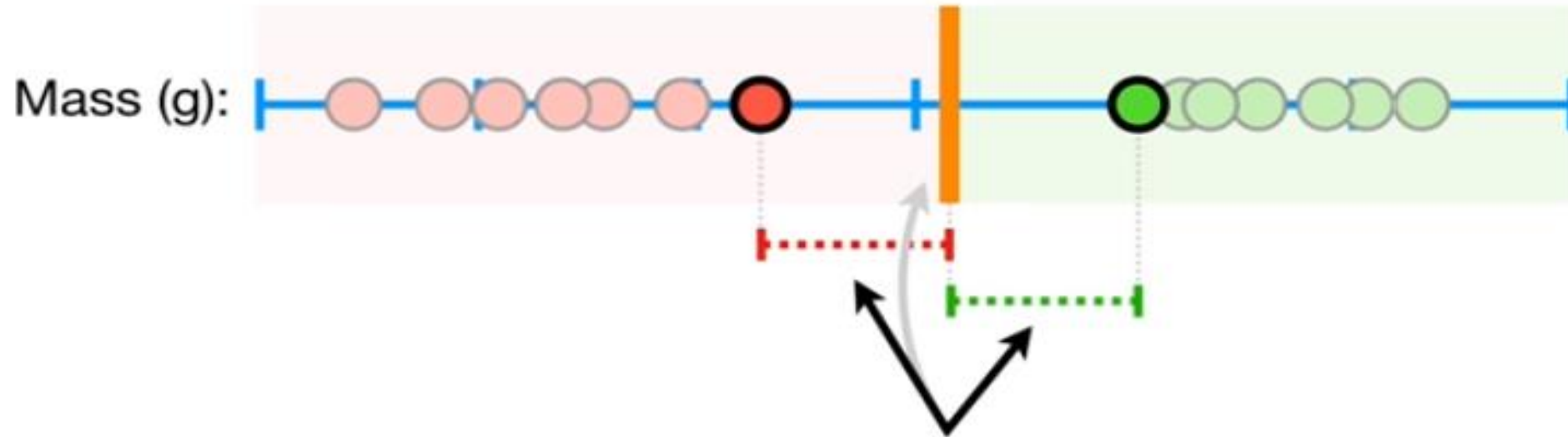
---



The shortest distance between the observations and the threshold is called the **margin**.

## Maximum Margin Classifier in 1-dimension

---

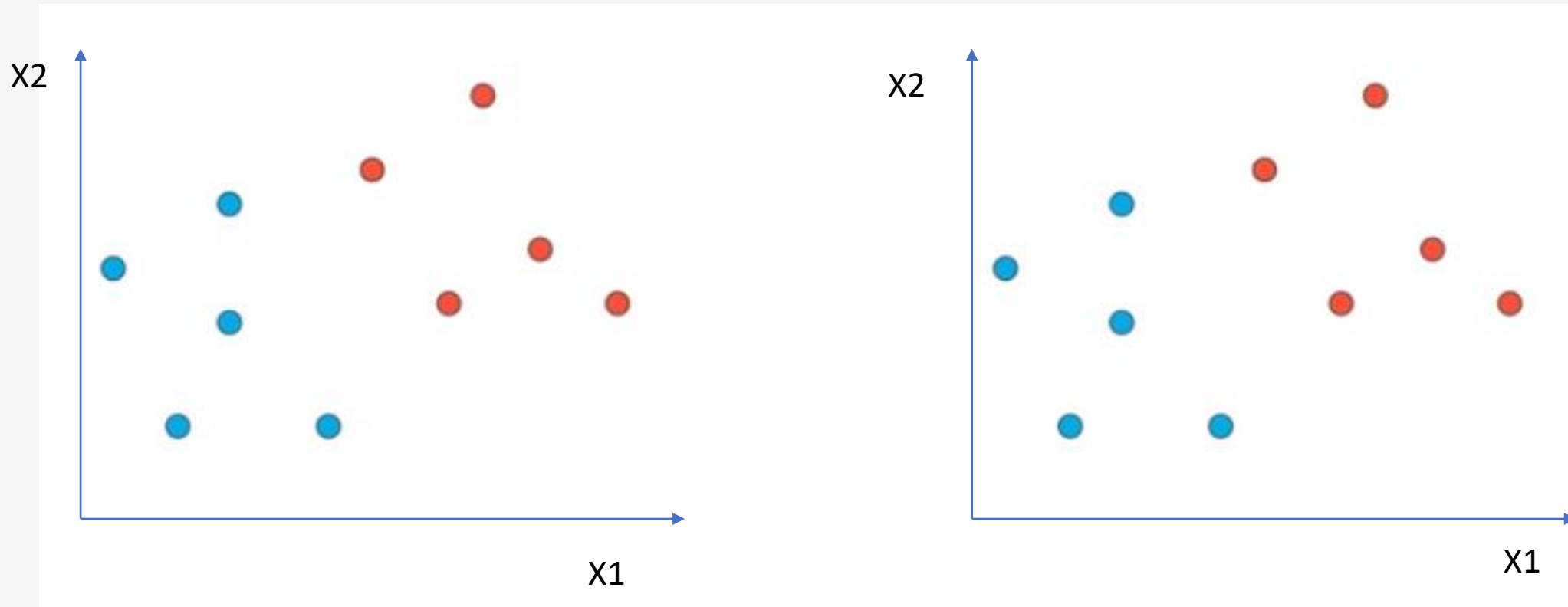


When the threshold is halfway  
between the two observations, the  
**margin** is as large as it can be.

# Maximum Margin Classifier in 2-dimension

---

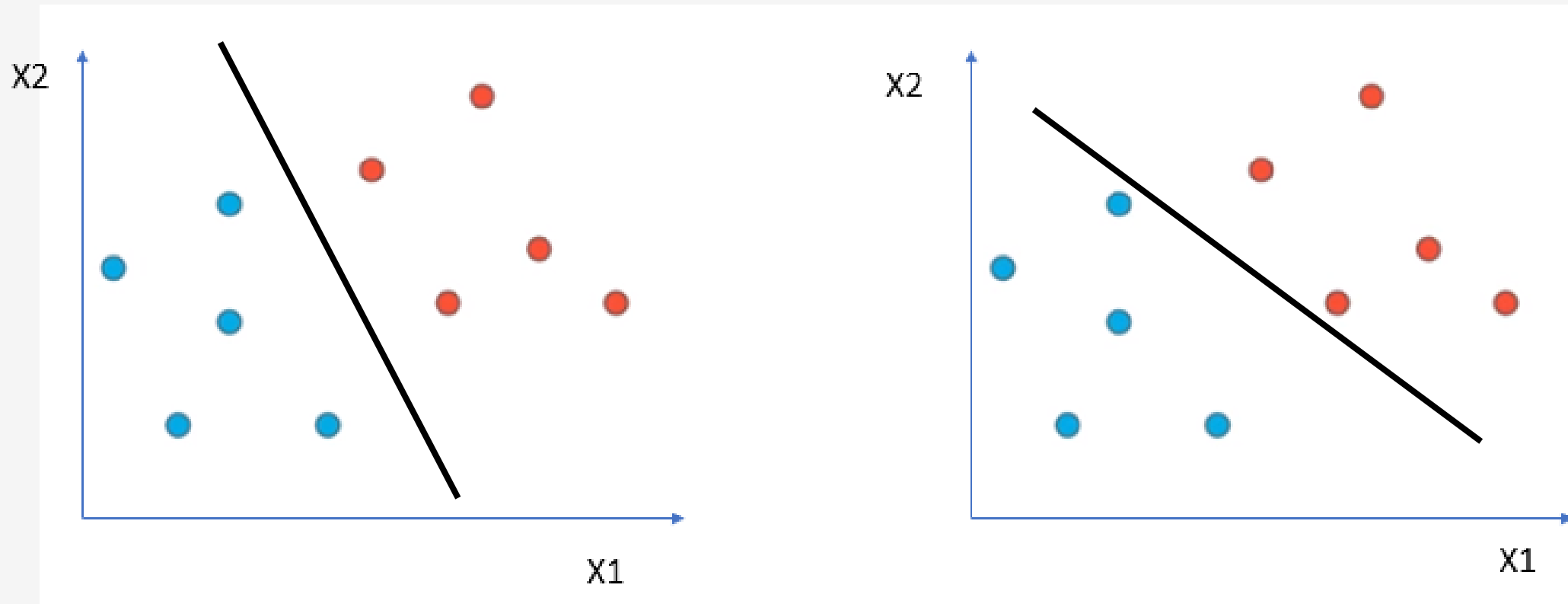
Draw a line that can separate the blue and red dots. The line will serve as the decision boundary.



# Maximum Margin Classifier in 2-dimension

---

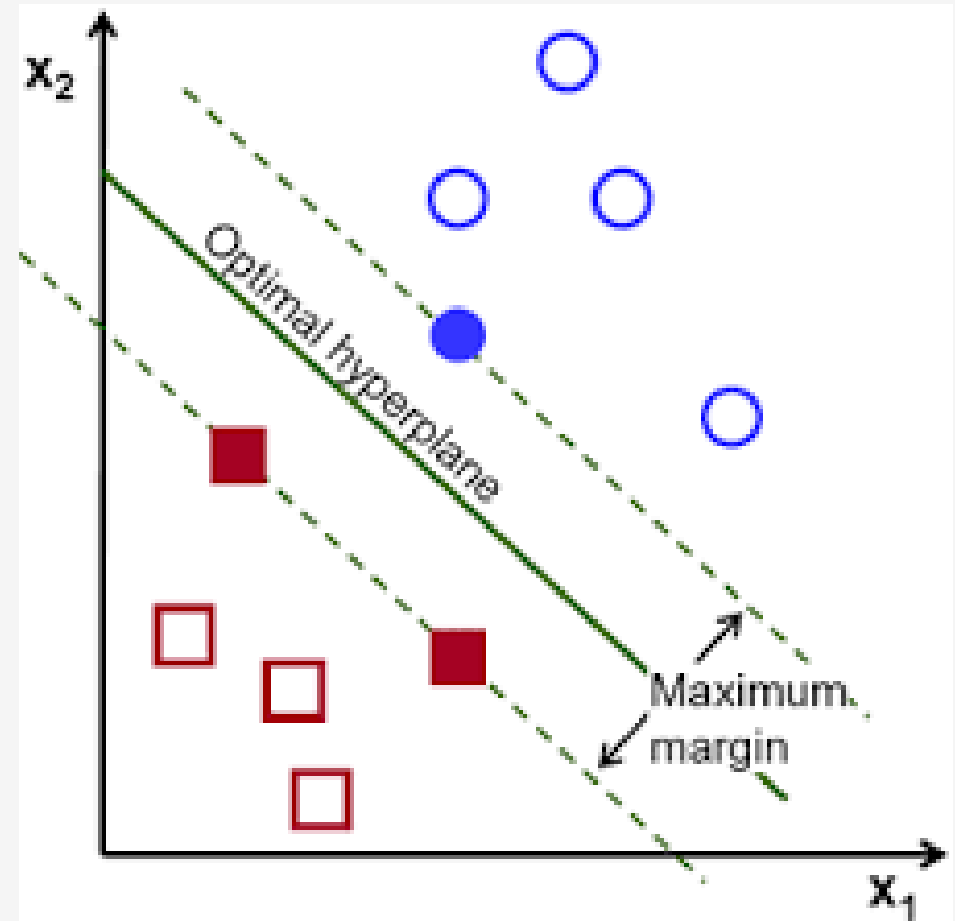
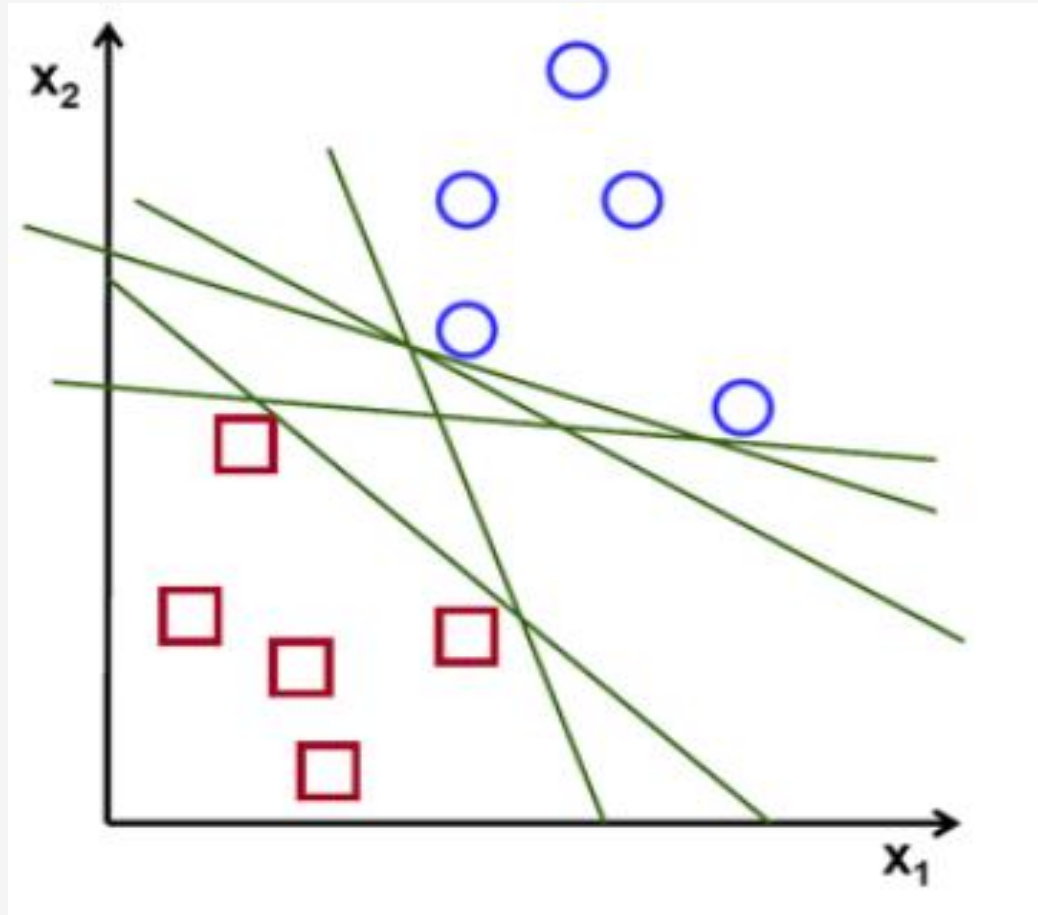
Which line will be a better line?





# Maximum Margin Classifier in 2-dimension

Which of the green line separate the blue circle from the red square data points?



## Support Vector Machine Terminology

---

So SVM is to try to find the hyperplane that maximize the margin between the plane and its nearest data points which are called the support vectors.

Watch <https://www.youtube.com/watch?v=N1vOgolbjSc>

- Hyperplane
- Maximum Margin Classifier
- Margin, Soft Margin
- Support Vectors
- C – parameters
- Kernel tricks

# Support Vector Machine

---

Learning by doing

# Support Vector Machine (continued with C-parameters and Kernel Trick)

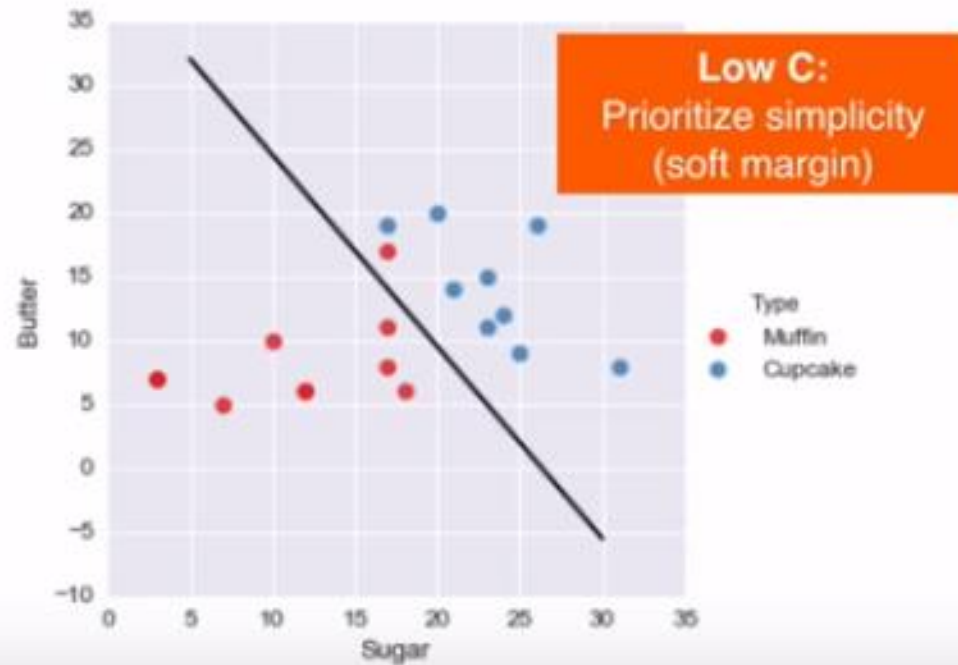
---

- C-Parameter
  - Default value is 1, but can be changed inside the SVM call
  - Low C-Parameter
    - allows mis-classification
    - Soft margin
    - Less complicated model, high bias, low variance, may underfit
  - High C-Parameter value
    - Try to fit as much as possible, allows no mis-classification
    - Hard margin
    - More complicated model, low bias, high variance, may overfit

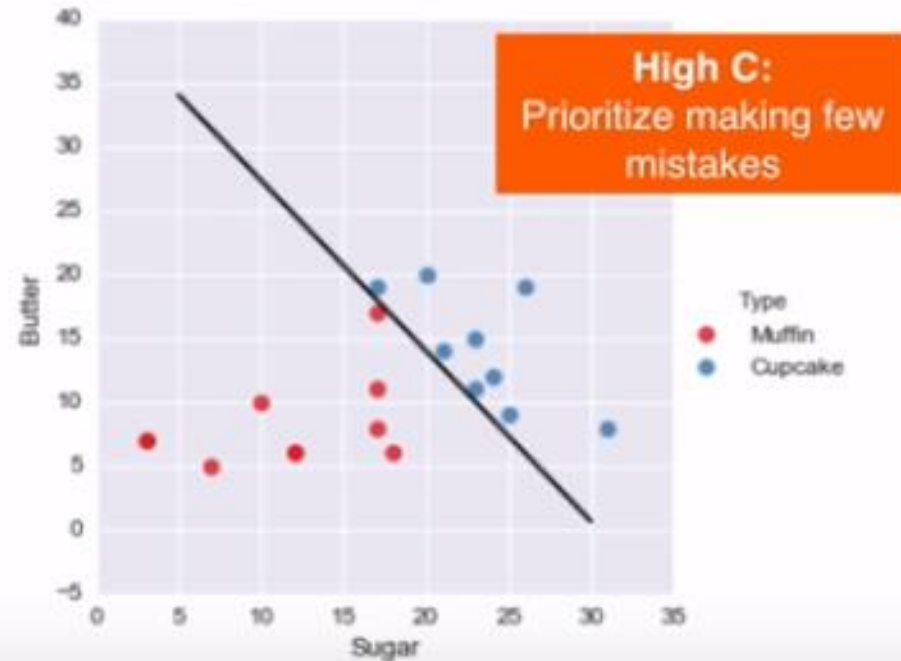
# Support Vector Machine (continued)

## C Parameter: Comparison

```
# Fit the SVM model with a LOW C
model = svm.SVC(kernel='linear', C=2**-5)
model.fit(sugar_butter, type_label)
```



```
# Fit the SVM model with a HIGH C
model = svm.SVC(kernel='linear', C=2**5)
model.fit(sugar_butter, type_label)
```



# Support Vector Machine (Kernel Trick)

---

Default is linear, but can use Polynomial, RBF (Radial Basis Function), or Gaussian

## Kernel Trick: Code

Original Code  
(linear)

```
# Fit basic SVC model (linear kernel)
model = svm.SVC(kernel='linear')
model.fit(sugar_butter, type_label)
```

Updated Code  
(RBF)

```
# Fit the SVC model with radial kernel
model = svm.SVC(kernel='rbf', C=1, gamma=2**-5)
model.fit(sugar_butter, type_label)
```

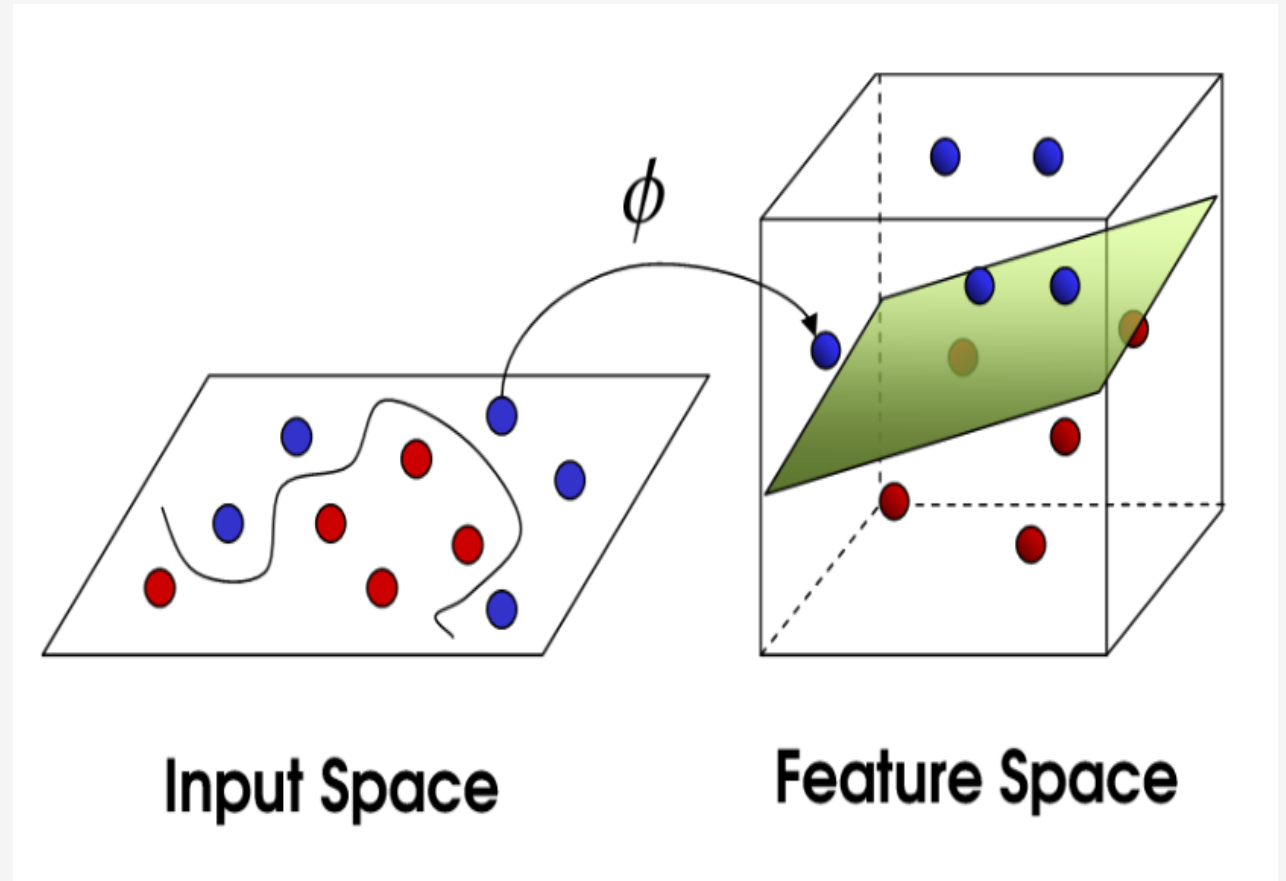
# Support Vector Machine (Kernel Trick)

---

Kernel Trick is a trick to transform the dataset from a lower dimension space to a higher dimension so that at higher dimension, the dataset can be separated by a linear hyperplane

Nice Kernel visualization

<https://www.youtube.com/watch?v=3liCbRZPrZA>



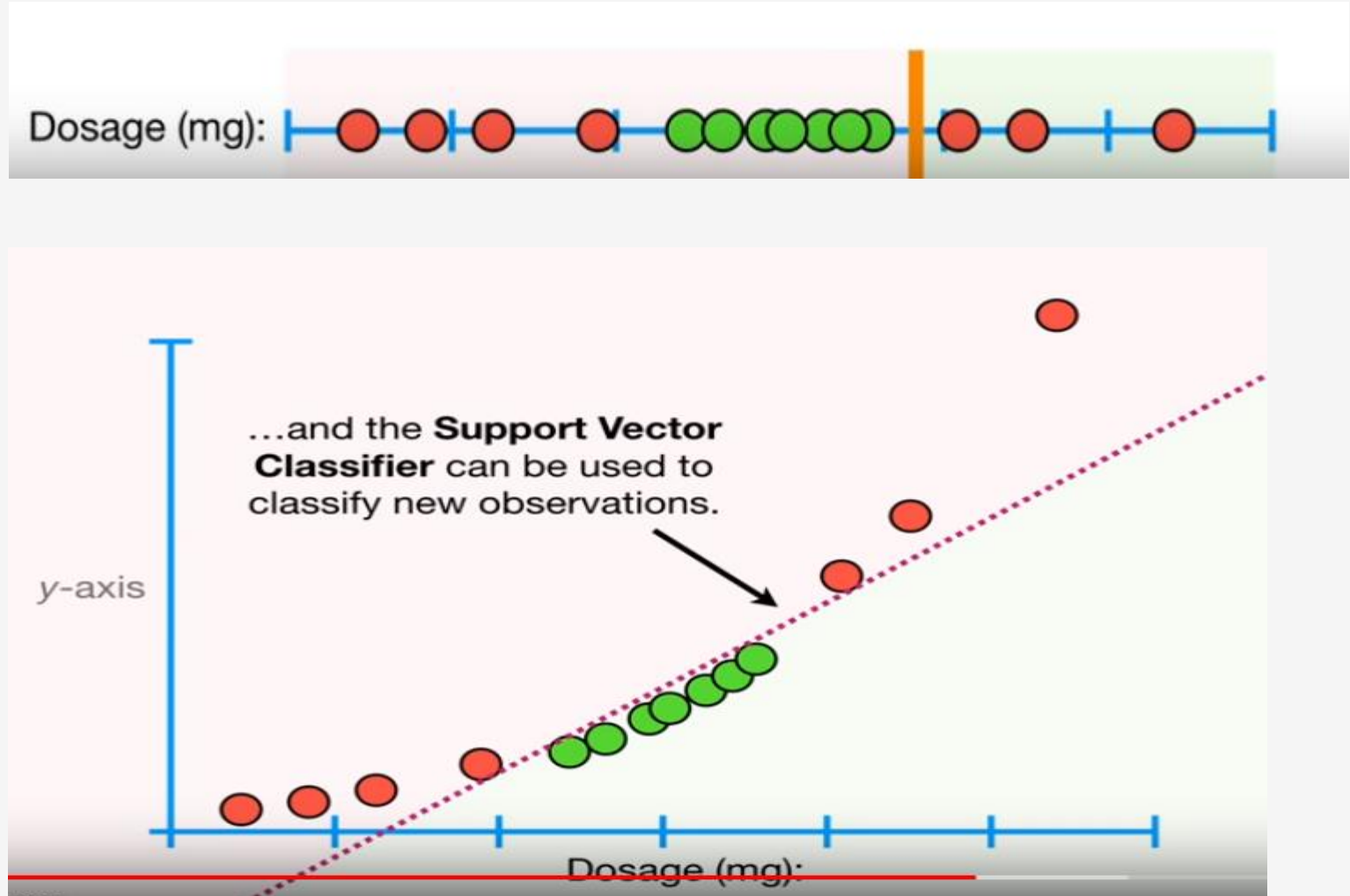
# Support Vector Machine (Kernel Trick)

Another example: <https://www.youtube.com/watch?v=efR1C6CvhmE> starting from 12:10

In 1-dimension, one cannot  
Separate the high and low  
dosage for cured patients

However, if we add another  
feature, Dosage squares

Then in 2-dimension space,  
the  
Data points can be separated  
by a line

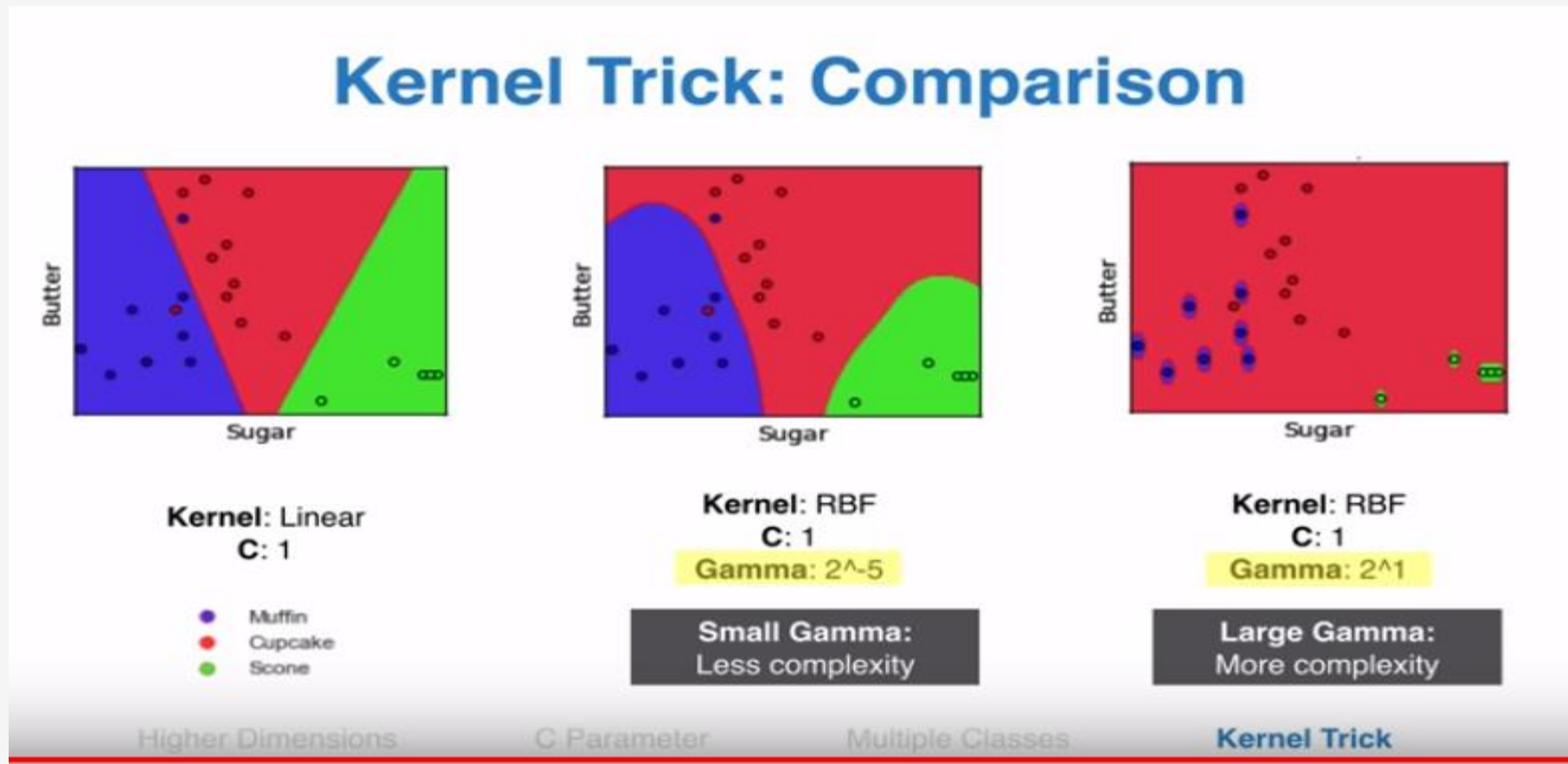




# Support Vector Machine (continued)

Need an additional Gamma parameter when using RBF Kernel

Large gamma: overfit, Low gamma: underfit



## Support Vector Machine (continued)

---

- So how to use the right Kernel, the right C-parameter and Gamma-parameter
- Use Grid search, i.e. think of C and Gamma parameters as two dimension in a grid, run different combination of C and Gamma until you find a good combination so your result (precision and recall) is good enough.
  - We call this fine-tuning your model.
  - However, how do you know this fine-tuning of C and Gamma is good for other datasets
  - => Cross validation comes to the rescue!
- This is time consuming and difficult. So this is one of the drawbacks in using SVM

# Support Vector Machine (summary)

---

- Advantages:

- Works well even when the number of features is much larger than the number of instances. Example in spam filter where a large number of words are the potential signifiers of a message being spam
- Allows a non-linear decision boundary curve. SVM transforms the variables to create new dimensions such that the representation of the classifier is a linear function at higher dimensions

- Disadvantages

- No probability associated with each prediction
- Training the SVMs can be time-consuming when data is large and there are lots of noise, hard to compute the soft margin

# Some useful references

---

- <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>
- Muffin and Cupcakes
- <https://www.youtube.com/watch?v=N1vOgolbjSc>
- StatQuest
- <https://www.youtube.com/watch?v=efR1C6CvhmE>
- Bias and Variance:
- <https://www.youtube.com/watch?v=EuBBz3bl-aA&feature=youtu.be>
- Nice Kernel visualization
- <https://www.youtube.com/watch?v=3liCbRZPrZA>