Python102

Python for Data Science Bootcamp

(6.2) Machine Learning Basics with Python Part 2

AIAT Academy

Machine Learning Basics

- Machine Learning Basic Part 1
  - Machine Learning with Python using Scikit Learn
  - Linear Regression
  - Logistic Regression
- **Machine Learning Basic Part 2**
  - **Support Vector Machine (SVM)**
  - **K means Clustering**
- Machine Learning Basic Part 3
  - Natural Language Processing (NLP)
  - Neural Network and Deep Learning

Support Vector Machines (SVM)

Support Vector Machines (SVM)

- **Support Vector Machine (SVMs)**
  - Supervised learning with associated learning algorithms
  - Analyse data and recognize pattern, used for classification and regression analysis

Support Vector Machines (SVM)

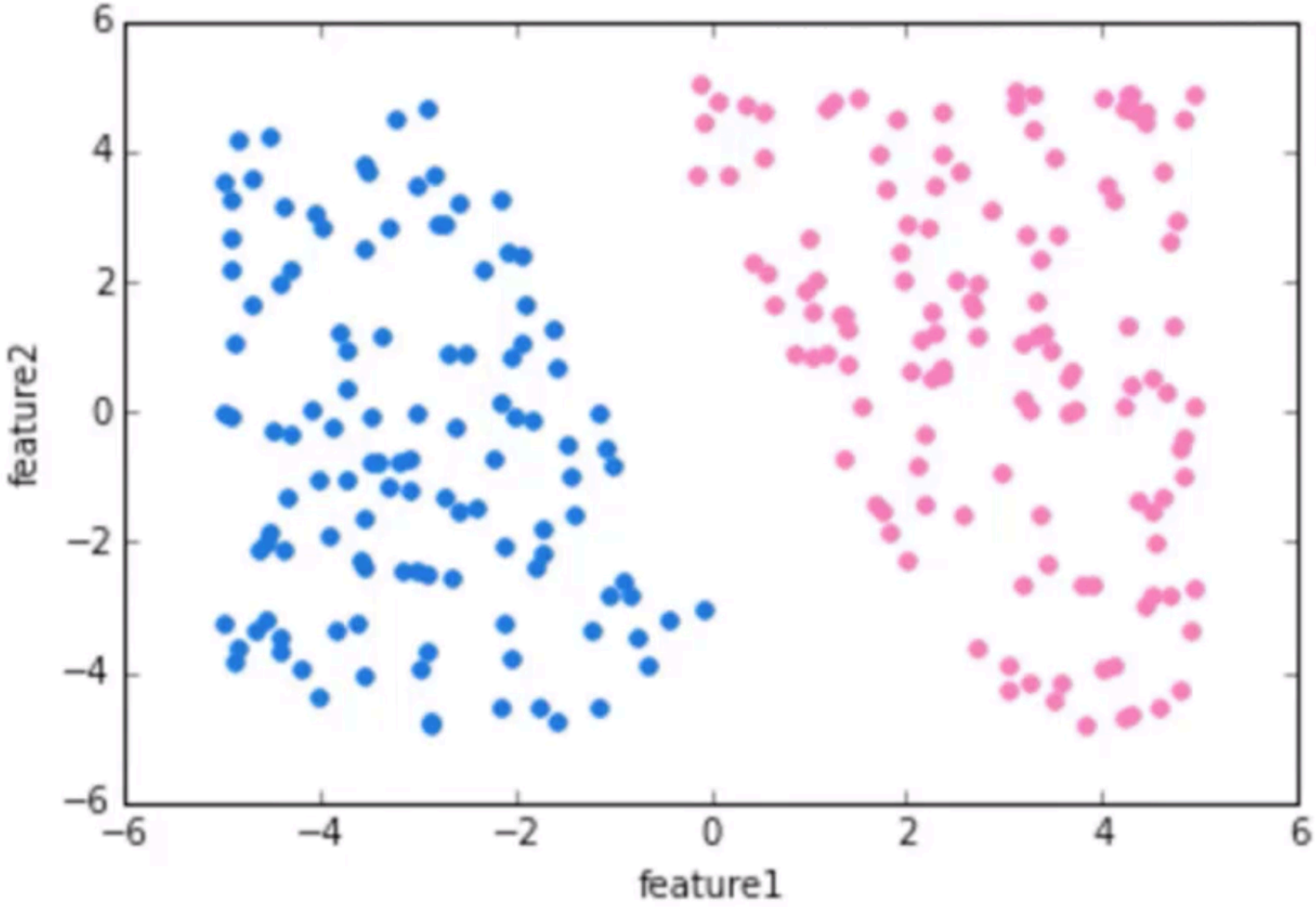
- **Support Vector Machine (SVMs)**
  - Supervised learning with associated learning algorithms
  - Analyse data and recognize pattern, used for classification and regression analysis
- An SVM model is a representation of the example of points space
  - The examples of the separate categories are divided by a clear gap that is as wide as possible

Support Vector Machines (SVM)

- **Support Vector Machine (SVMs)**
  - Supervised learning with associated learning algorithms
  - Analyse data and recognize pattern, used for classification and regression analysis
- An SVM model is a representation of the example of points space
  - The examples of the separate categories are divided by a clear gap that is as wide as possible

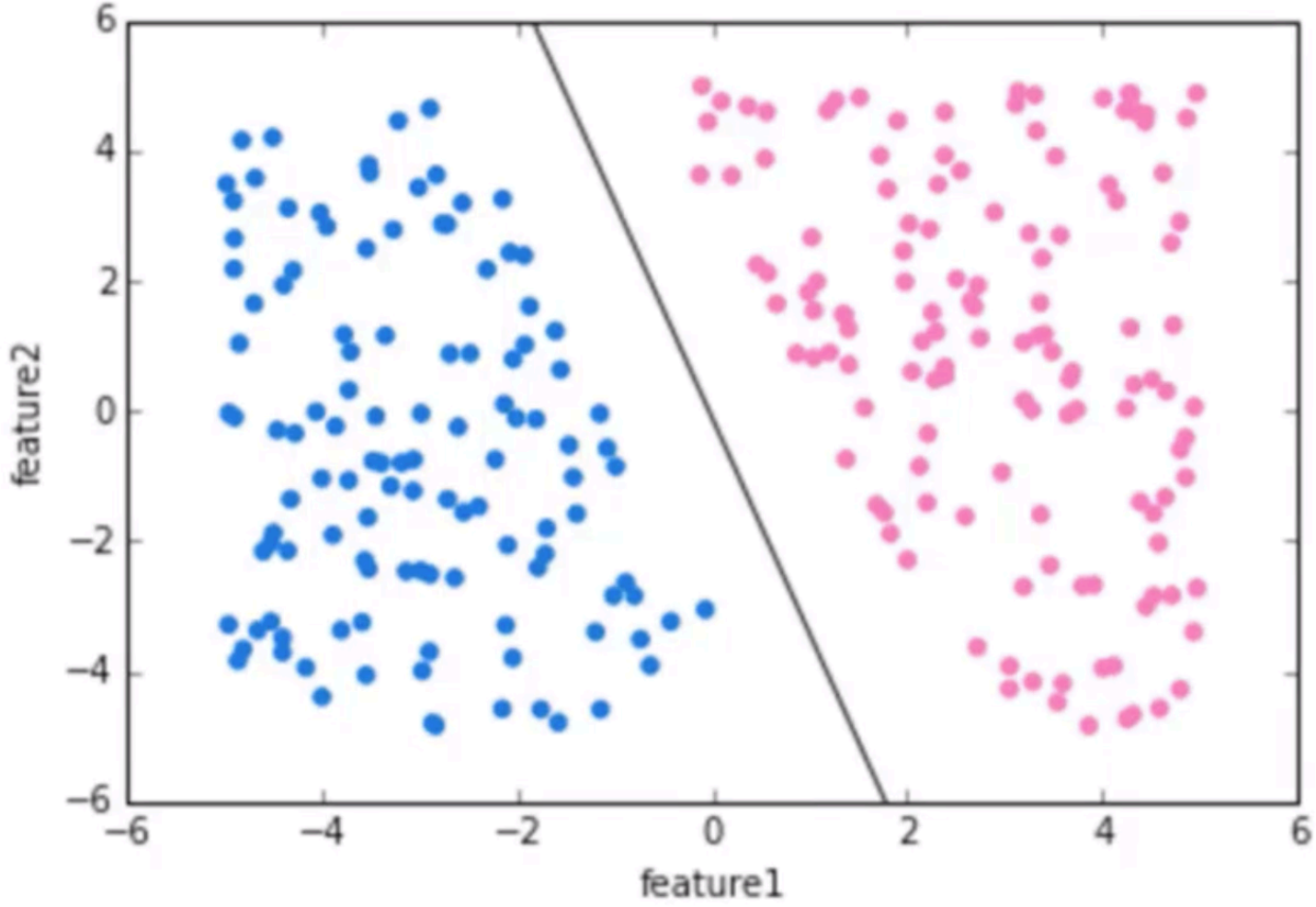
Support Vector Machines (Example)

- Imagine the labelled training data below



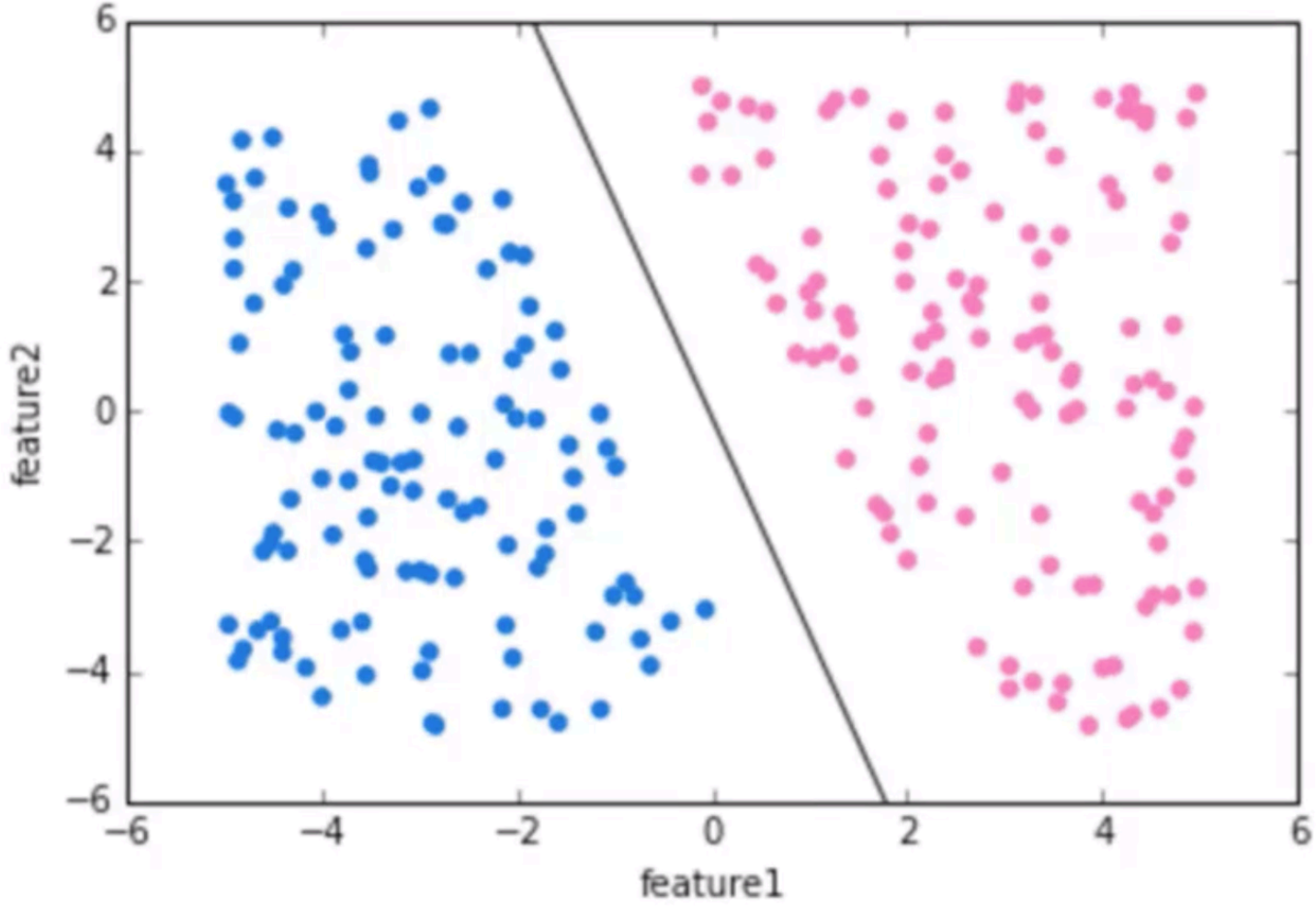
Support Vector Machines (Example)

- We can draw a separating "hyperplane" between the classes



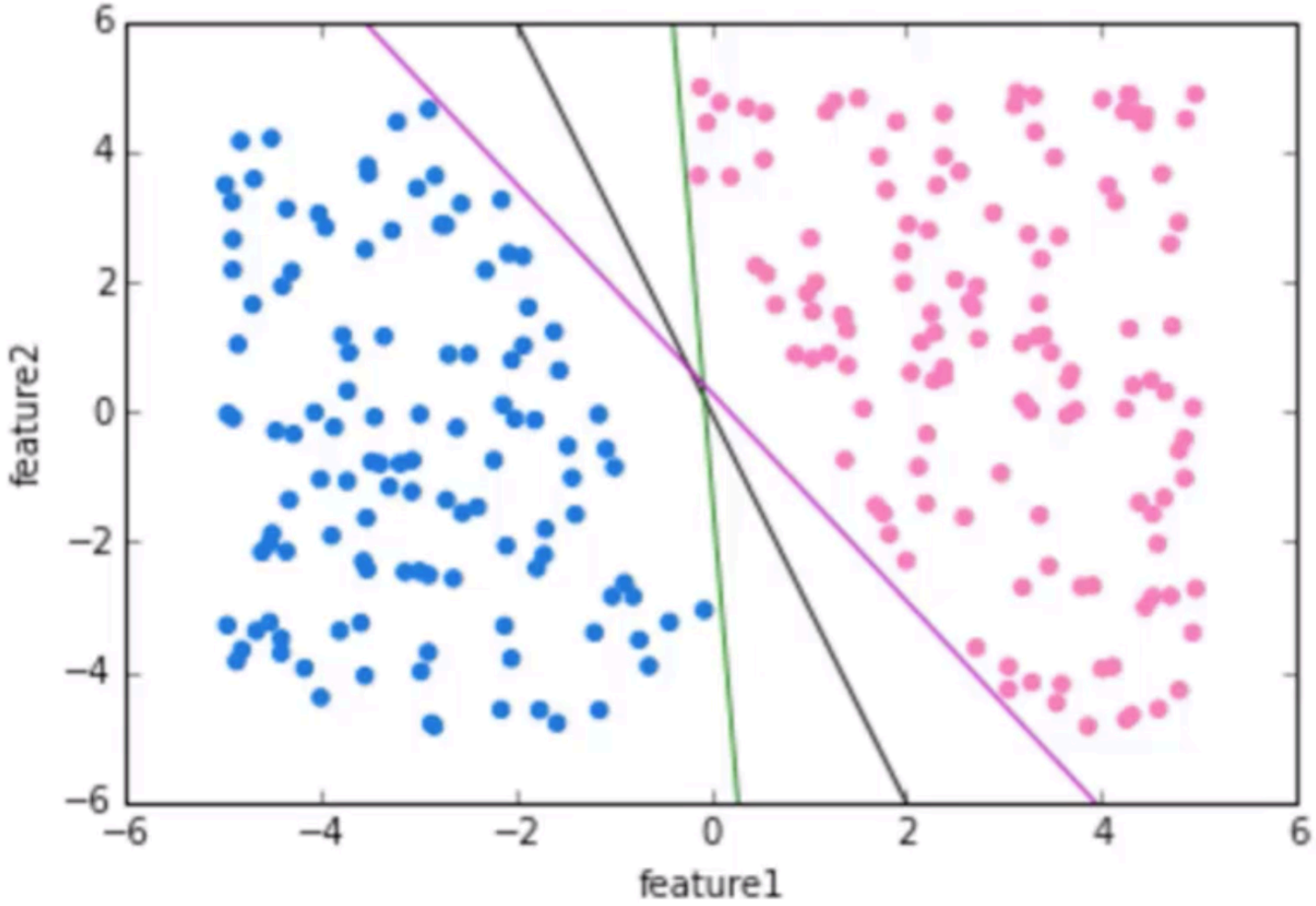
Support Vector Machines (Example)

- We can draw a separating "hyperplane" between the classes



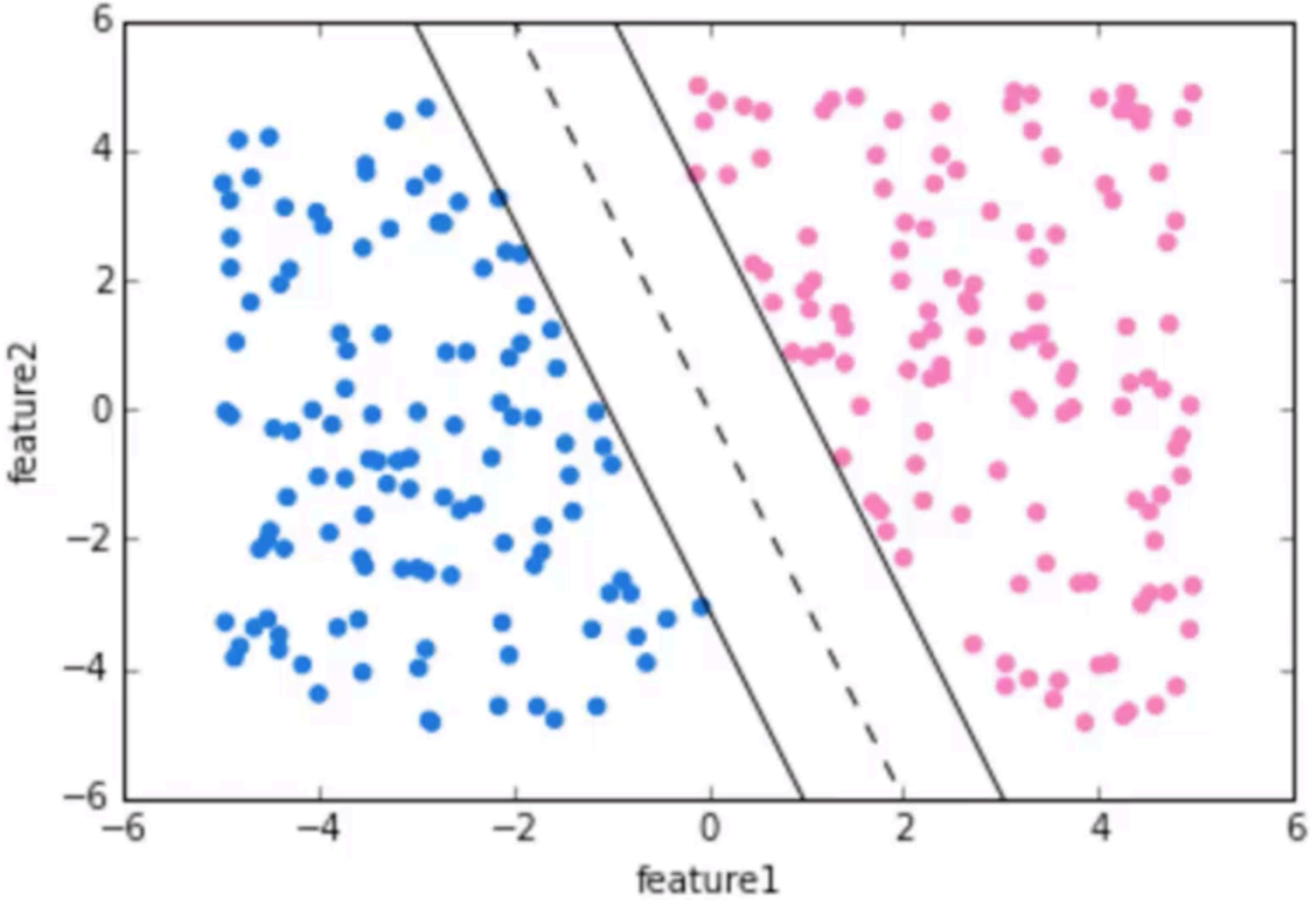
Support Vector Machines (Example)

- But we have many options of hyperplanes that separate perfectly



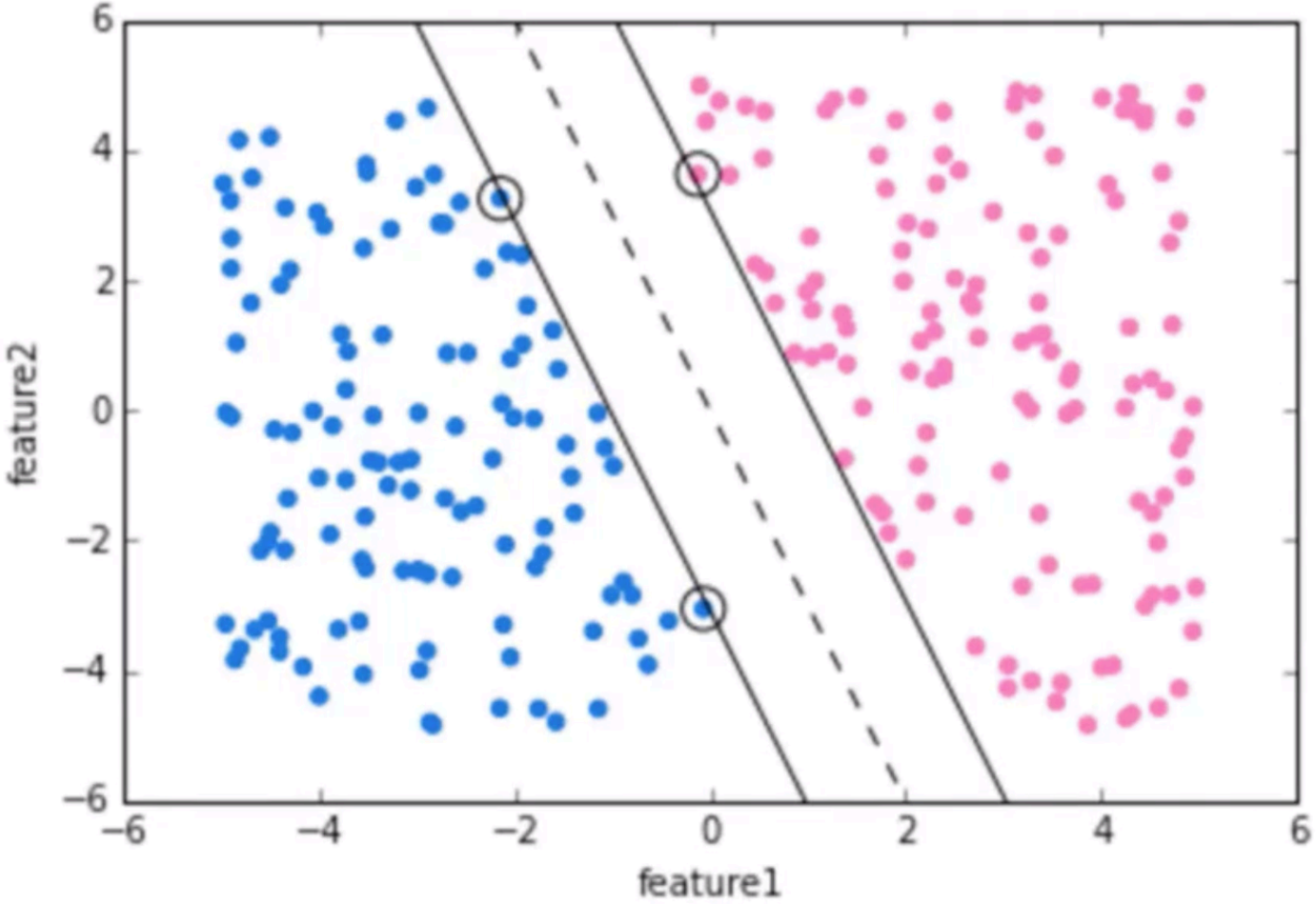
Support Vector Machines (Example)

- We would like to choose a hyperplane that maximizes the margin between classes



Support Vector Machines (Example)

- The vector points that the margin lines touch are knows as Support Vectors



Support Vector Machine with Python
Colab

AiAT ACADEMY

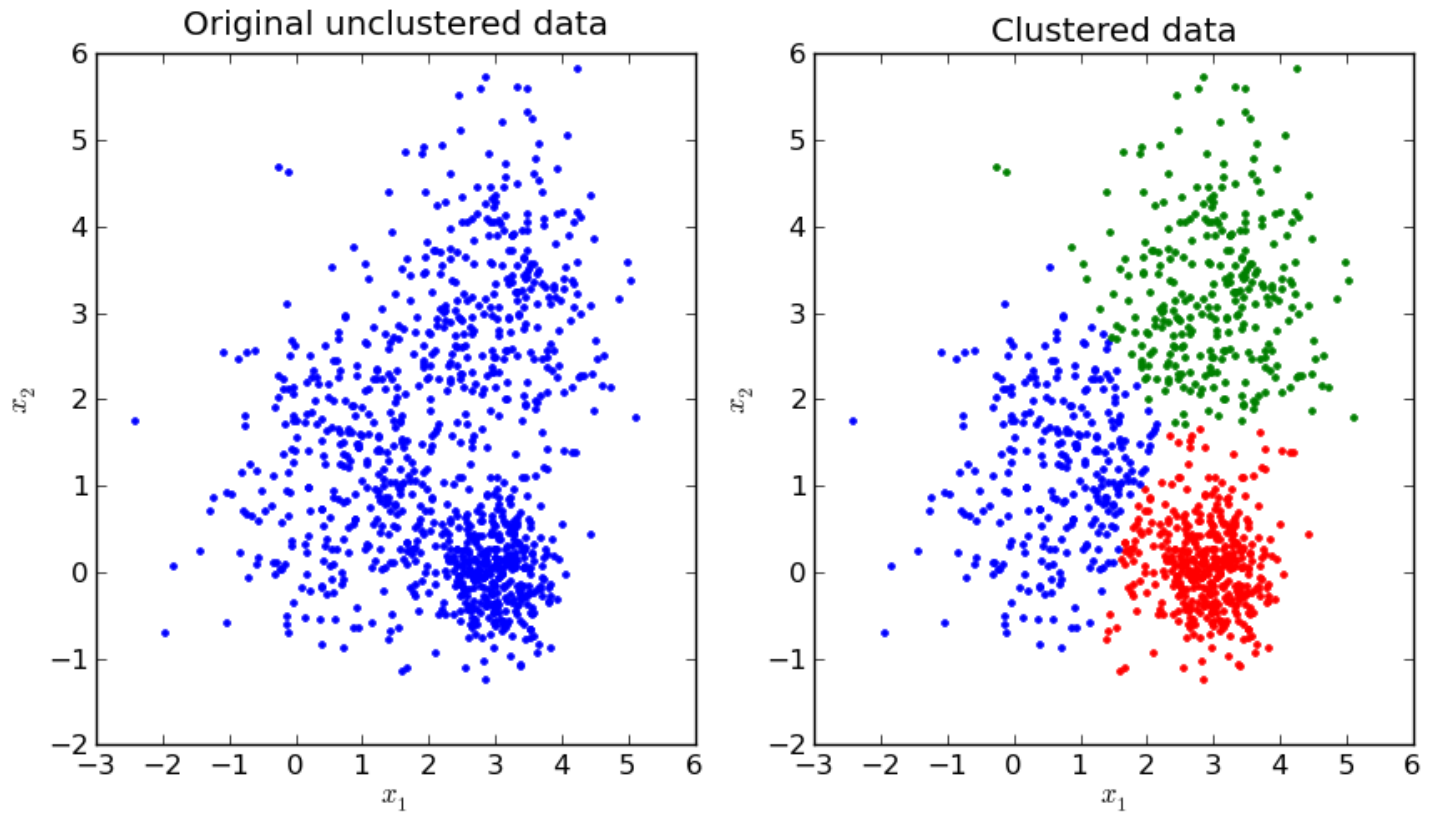
K means Clustering

K means Clustering

- **K Means**
  - Unsupervised learning that will attempt to group similar clusters together in your data
- Typical clustering problems
  - Cluster similar documents
  - Cluster customers based on Features
  - Market segmentation
  - Identify similar physical groups

K means Clustering

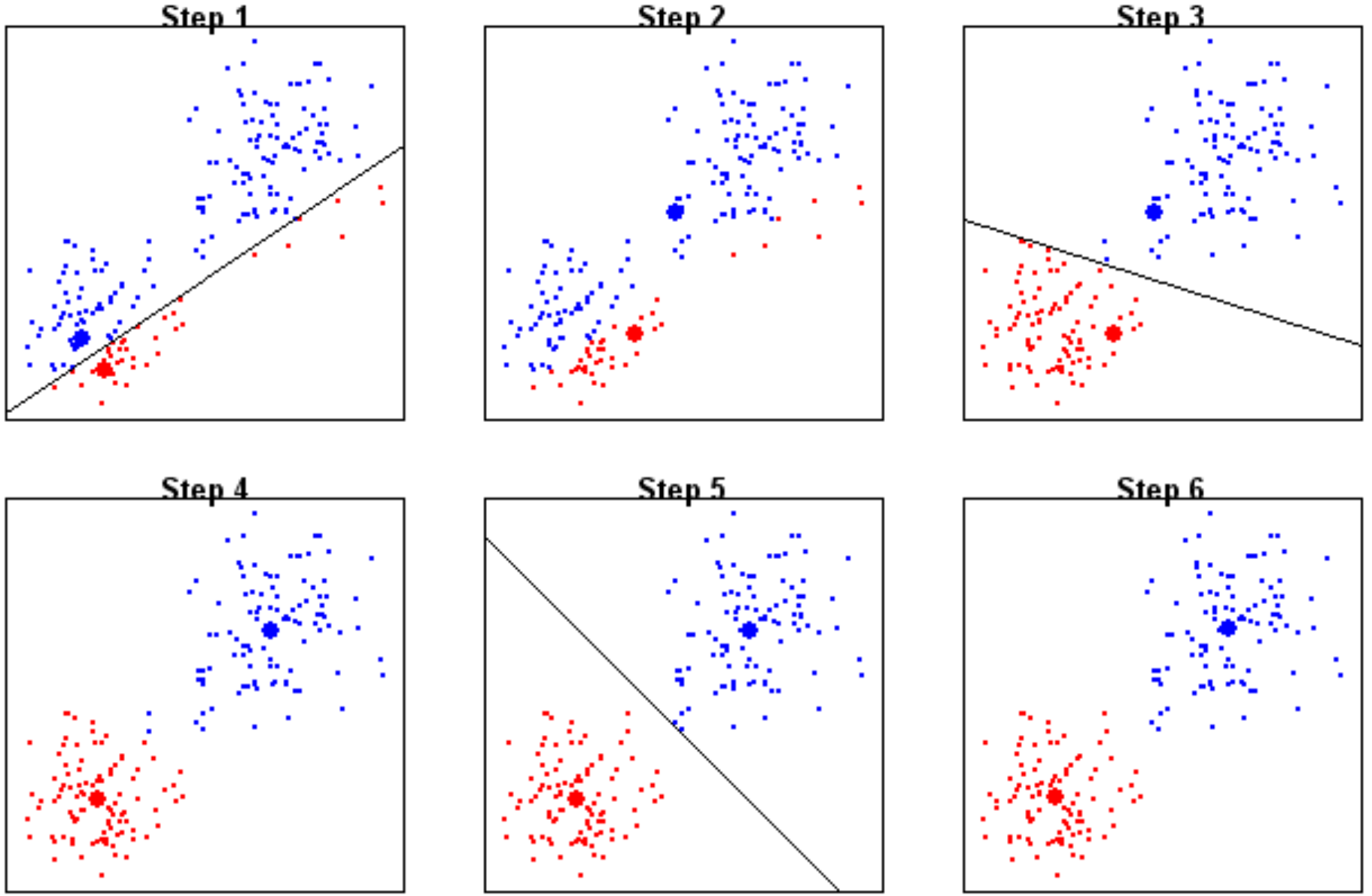
- The goal is to divide data into distinct groups such that observations within each group are similar



K means Clustering (Algorithm)

- Choose a number of Cluster "K"
- Randomly assign each point to a cluster
- Until clusters stop changing, repeat the following:
  - For each cluster, compute the cluster centroid by taking the mean vector of points in the cluster
  - Assign each data point to the cluster for which the centroid is the closet

K means Clustering

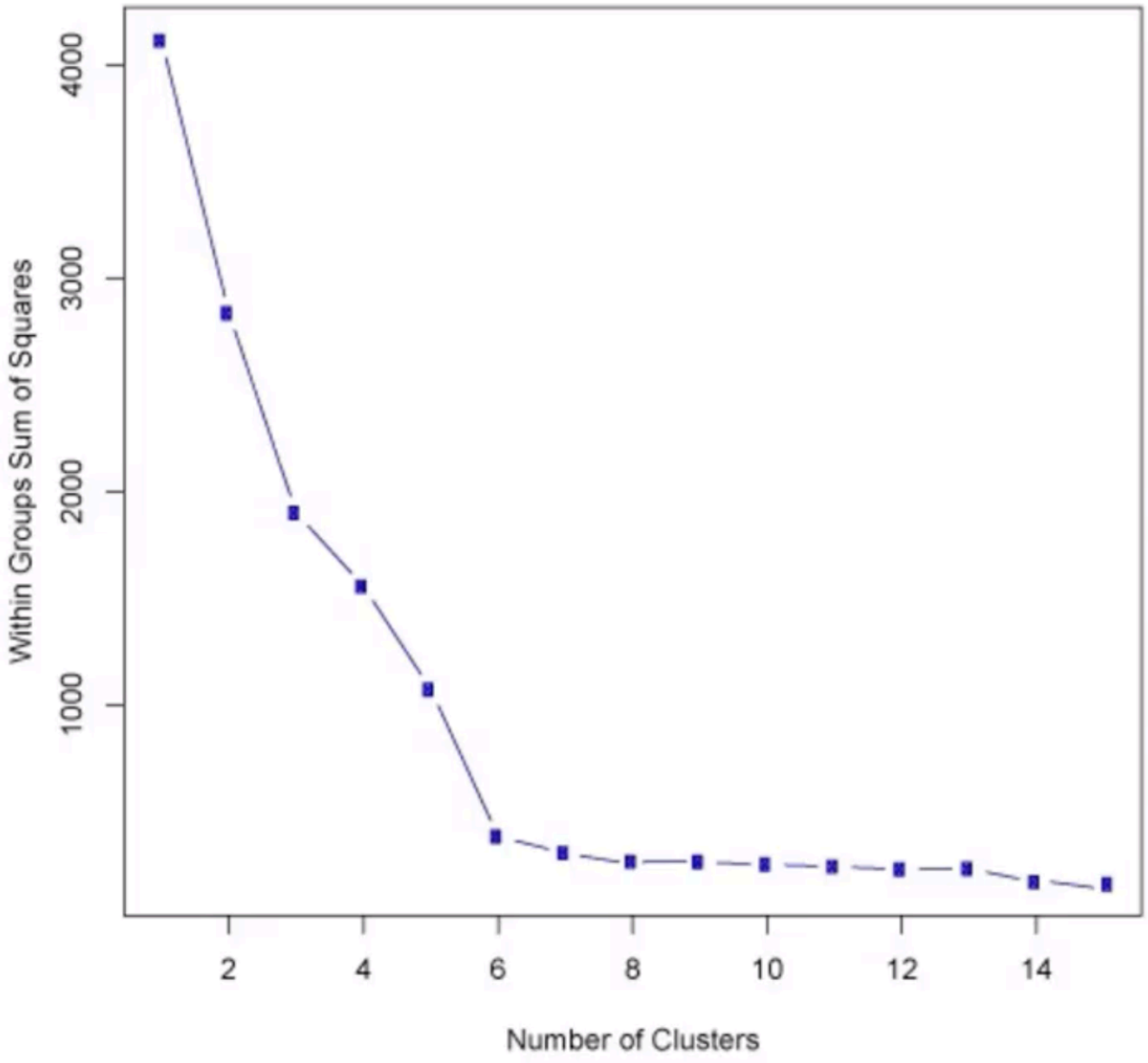


K means Clustering (Choosing K Value)

- There is no easy answer for choosing a "best K value"
- There is a way called "elbow method"
  - Compute sum of squared error (SSE) for some value of k (e.g. 2, 4, 6, 8, etc.)
  - The SSE is defined as the sum if the squared distance between each member of the cluster and its centroid

K means Clustering (Choosing K Value)

- The results will look like "**elbow**"
- In this case six or seven cluster may give us a better result of clusterinf



K means Clustering with Python
Colab