Machine Learning: How Machines Learn from Data

Stephen Huang

June 8, 2019

Contents

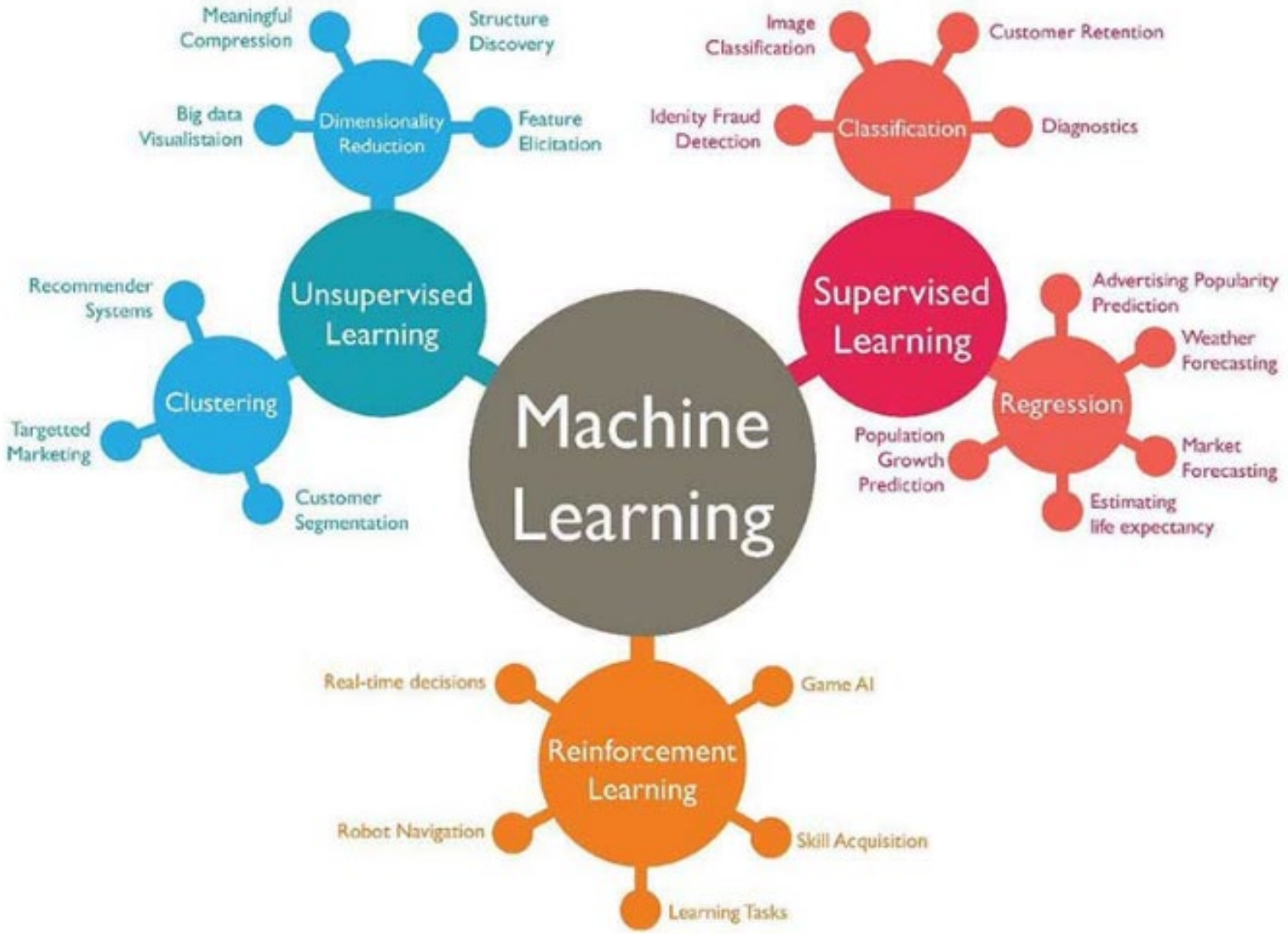
1. Introduction to ML
2. Supervised Learning
   - Regression Analysis
   - Classification: Support Vector Machine
3. Unsupervised Learning
   - Clustering: K-Means
4. ML Process
5. Summary

1. Introduction to ML

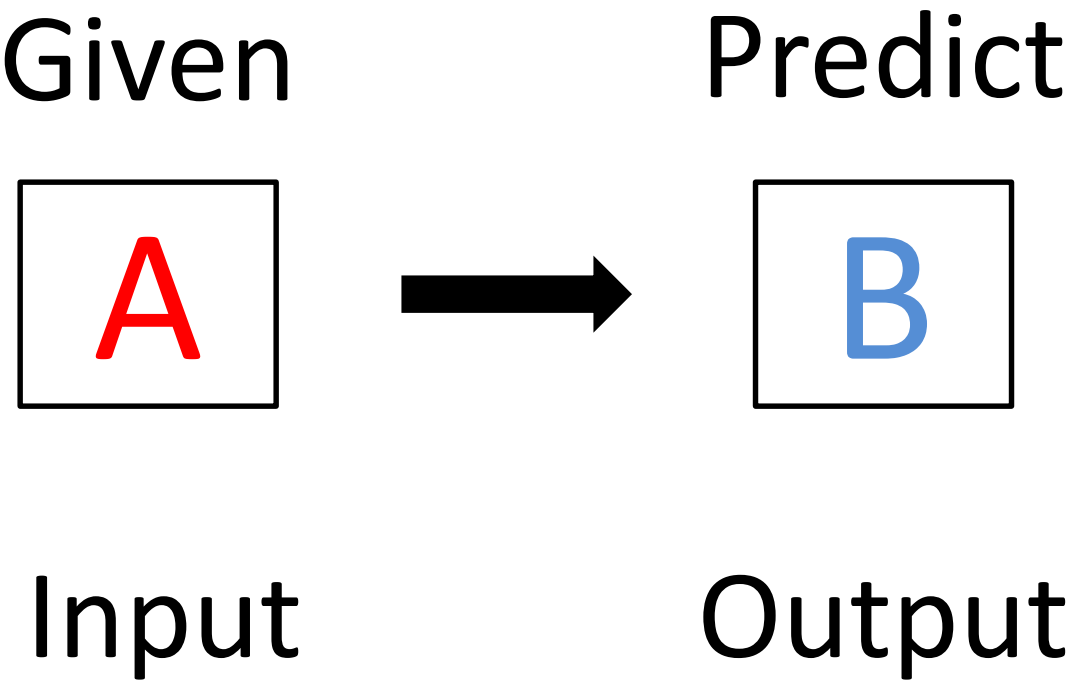
Field of study that gives the computer the ability to learn without being explicitly programed.

Arthur Samuel, 1959

Types of ML


Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

Supervised Learning

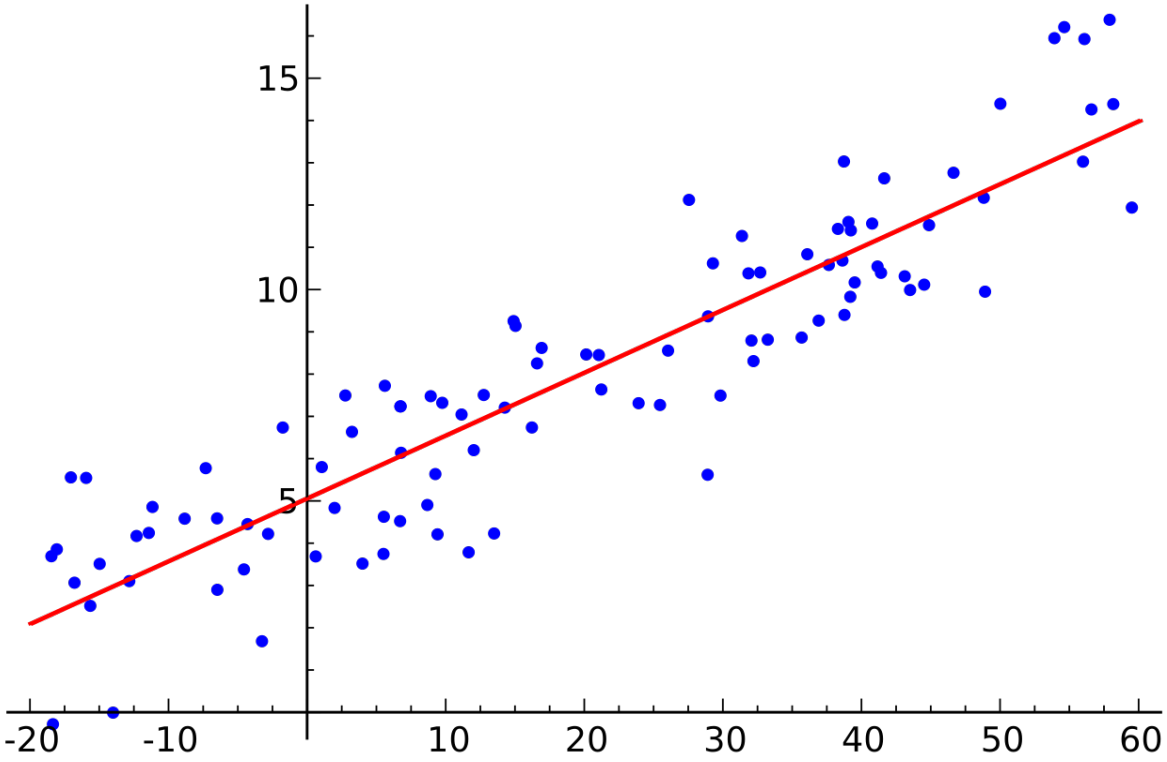


Supervised Learning

Input (A)	Output (B)	Application
Email	Spam or not	Spam Filter
English	Chinese	Speech translation
Ad, User info	Click/Purchase	Online advertisement
Image of a product	Defect or not	Product inspection

2.1 Regression Analysis

- Regression is the supervised learning task for modeling and predicting continuous, numeric variables.



Real Estate Example

Size of House (Sq Ft)	# of BRs	Price ($1,000)
523	1	115
645	1	150
708	2	210
1034	3	280
2290	4	355
2545	4	440

Size -> Price

Size of House (Sq Ft)	# of BRs	Price ($1,000)
523	1	115
645	1	150
708	2	210
1034	3	280
2290	4	355
2545	4	440

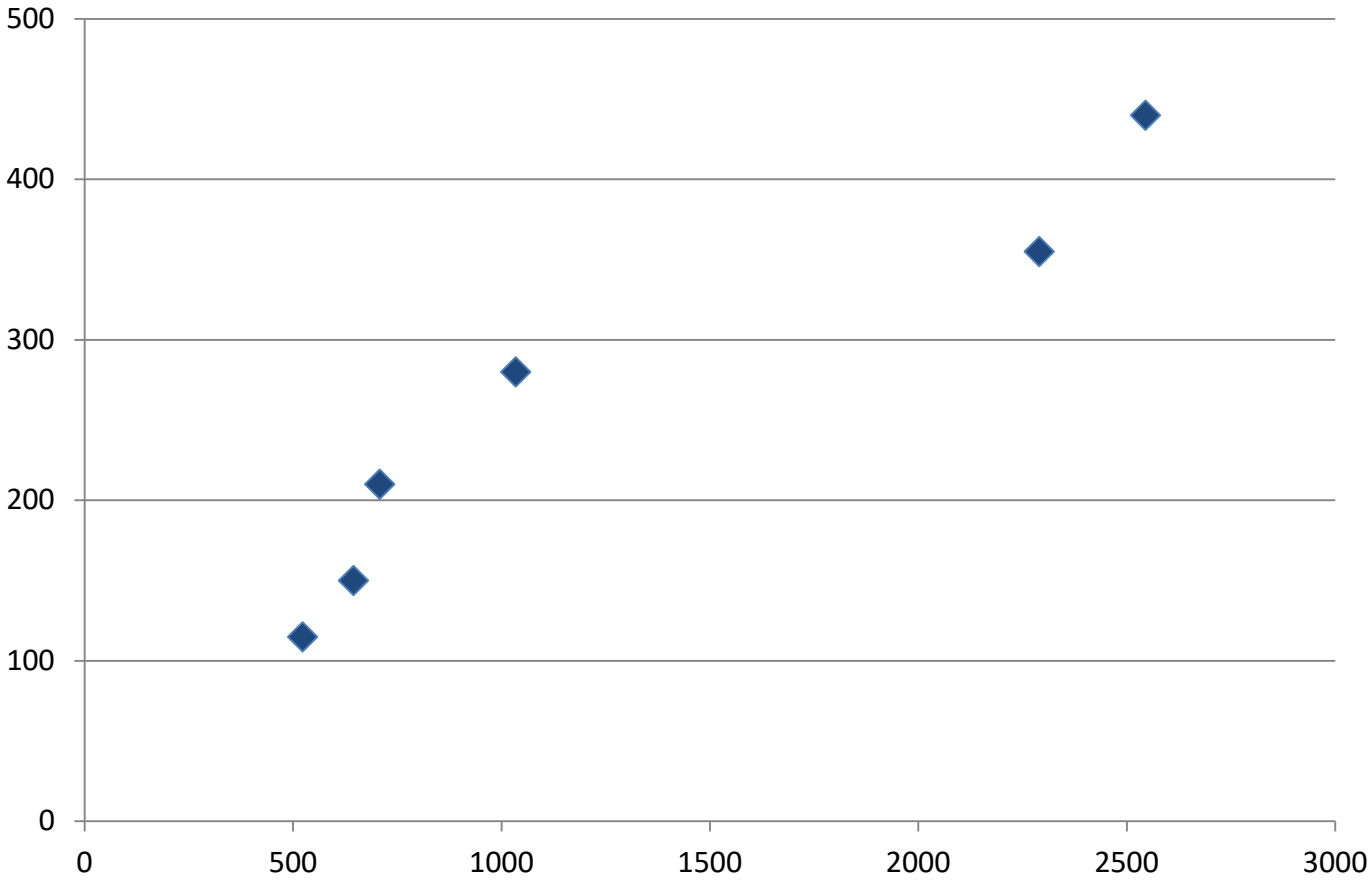
Size, #BR -> Price

Size of House (Sq Ft)	# of BRs	Price ($1,000)
523	1	115
645	1	150
708	2	210
1034	3	280
2290	4	355
2545	4	440

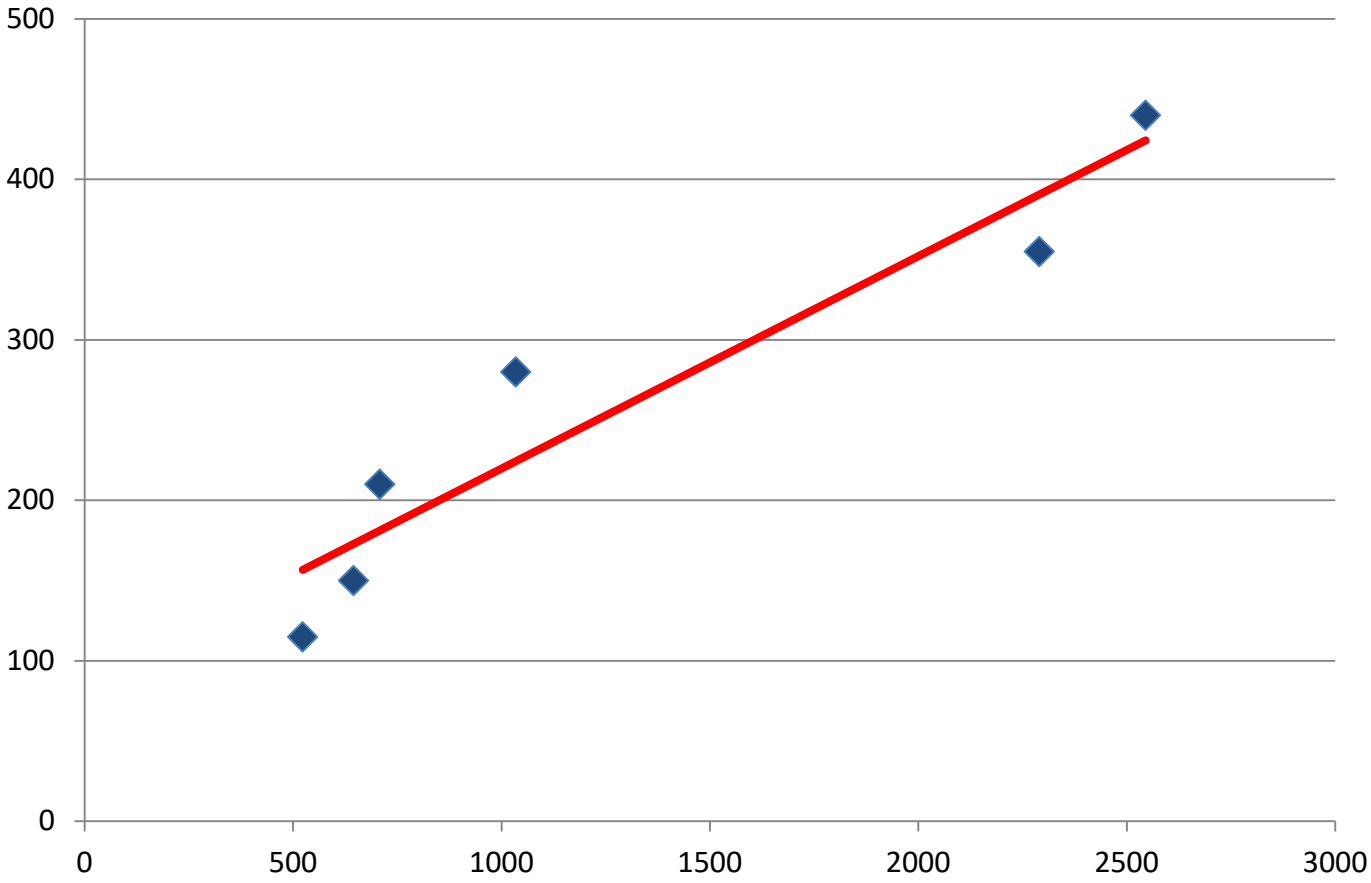
Price -> Sq Ft

Size of House (Sq Ft)	# of BRs	Price ($1,000)
523	1	115
645	1	150
708	2	210
1034	3	280
2290	4	355
2545	4	440

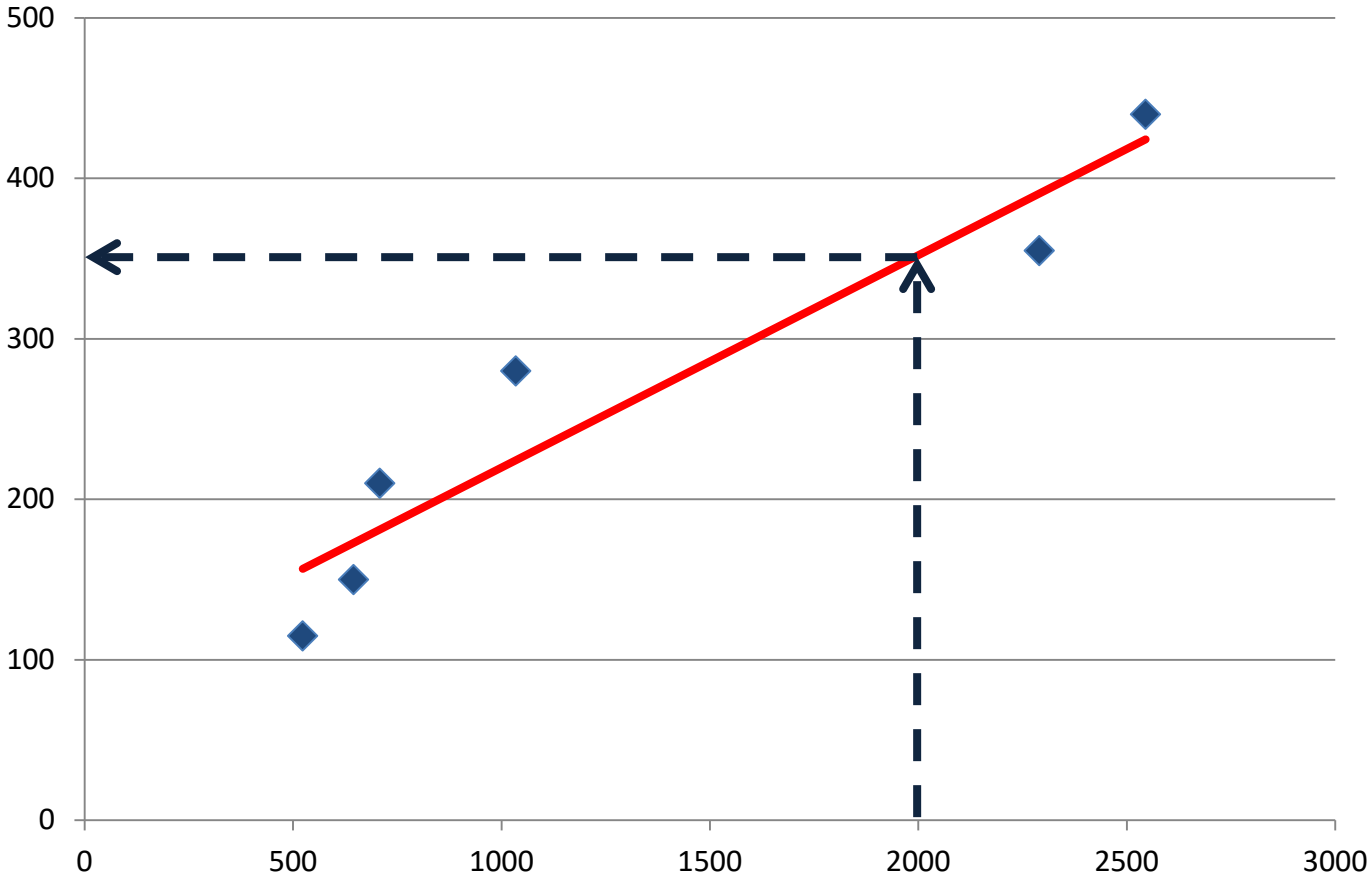
Regression



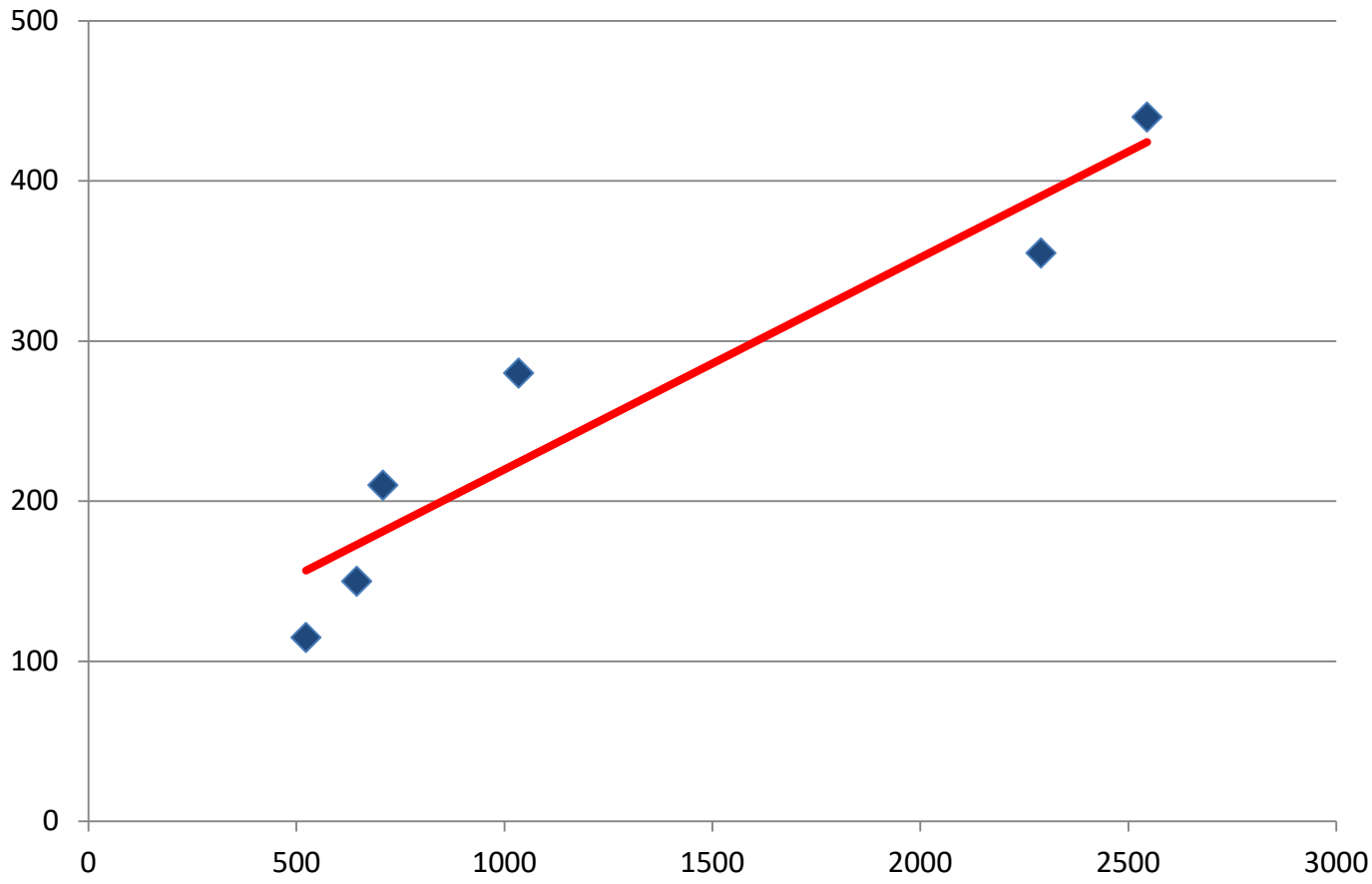
Linear Regression



Under-fitting vs Over-fitting



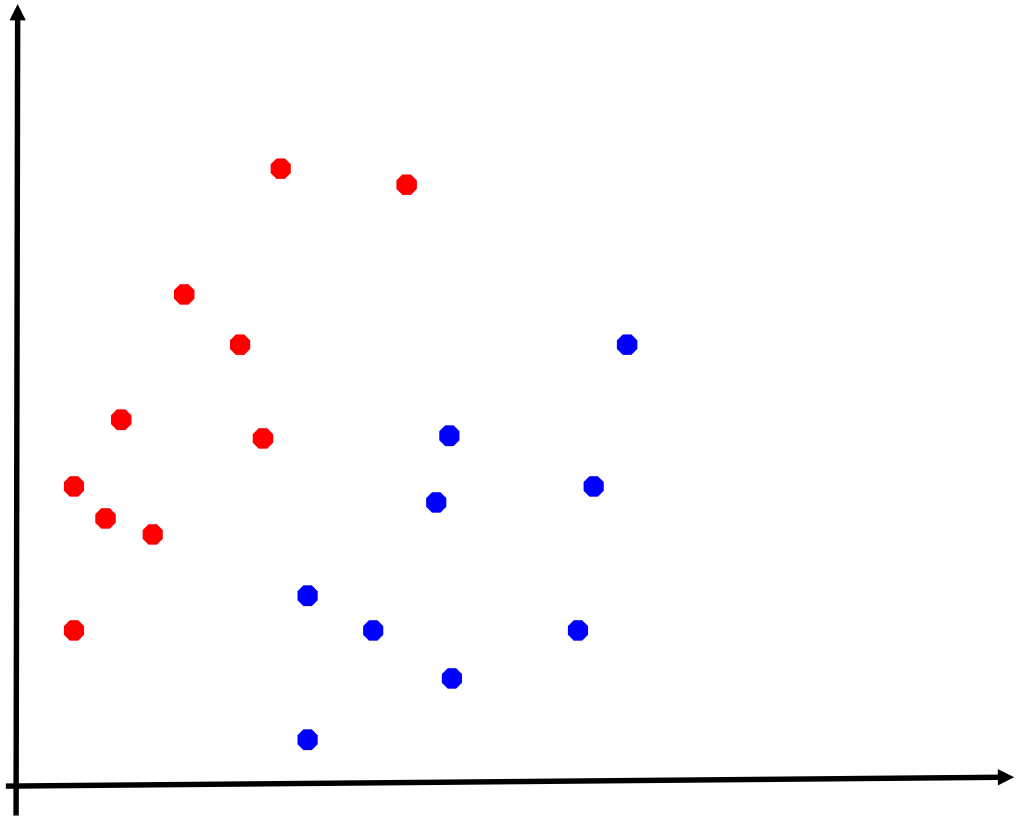
Under-fitting vs Over-fitting



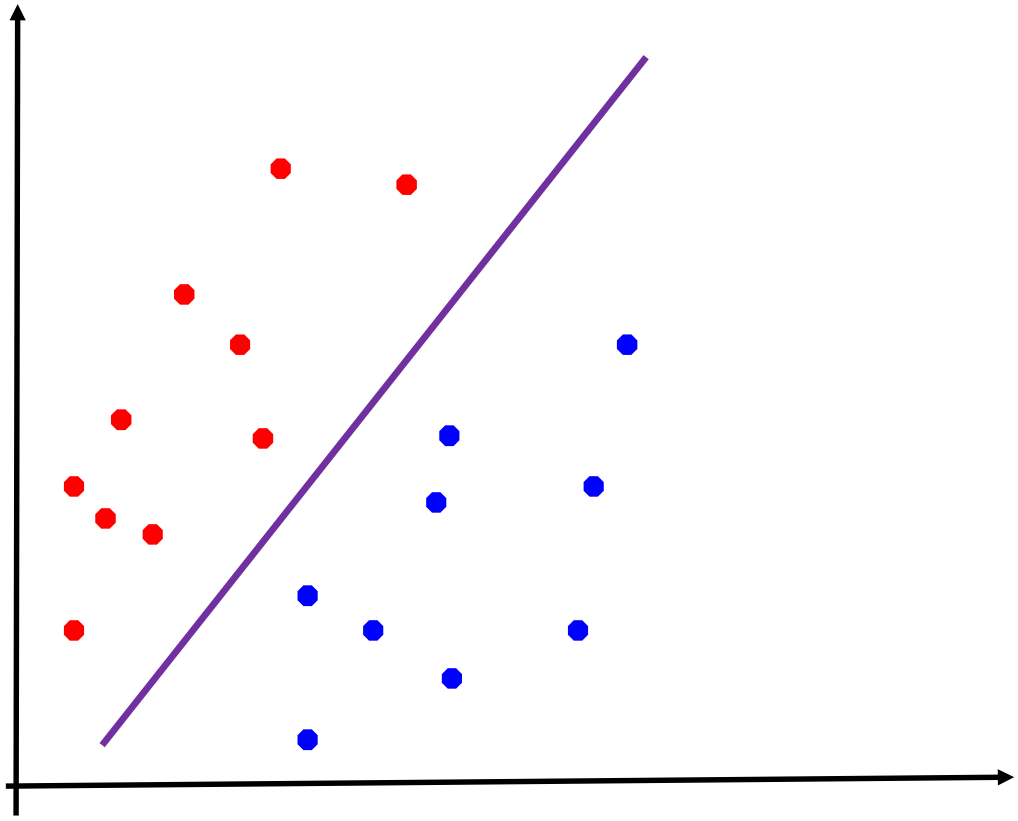
2.2 Support Vector Machine

- Classification is the supervised learning task for modeling and predicting categorical variables.
- SVM algorithms finds a boundary that maximizes the distance between the closest members of separate classes.

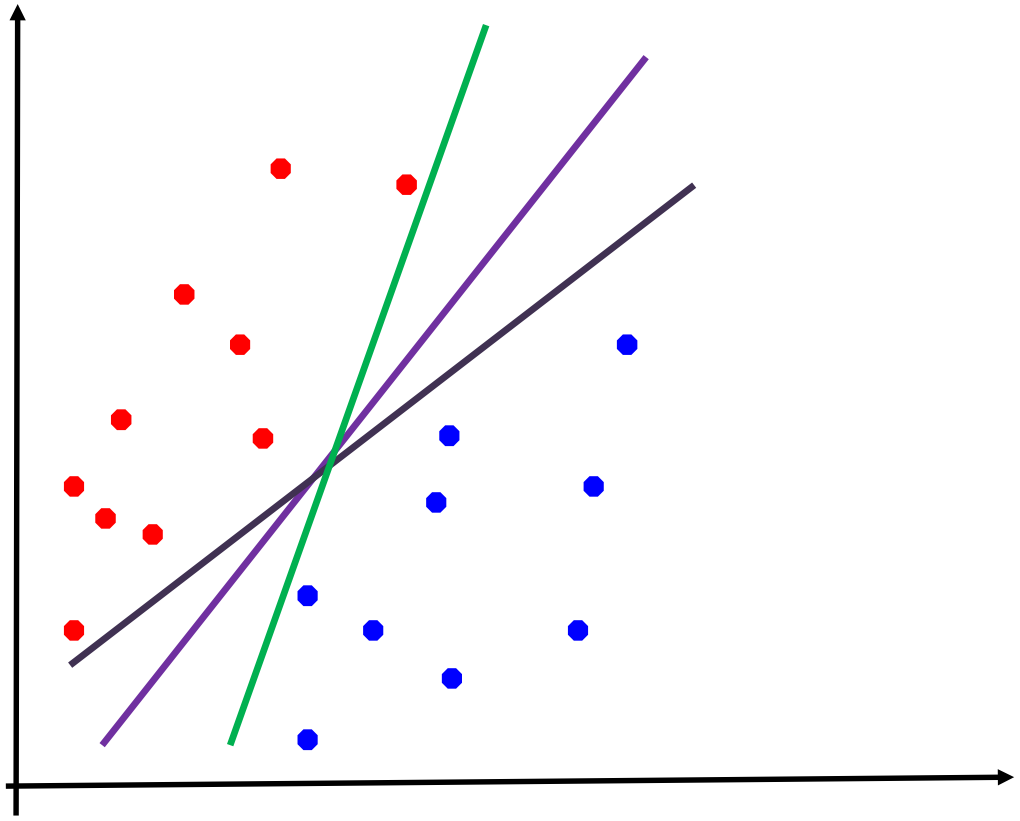
Two classes of data points



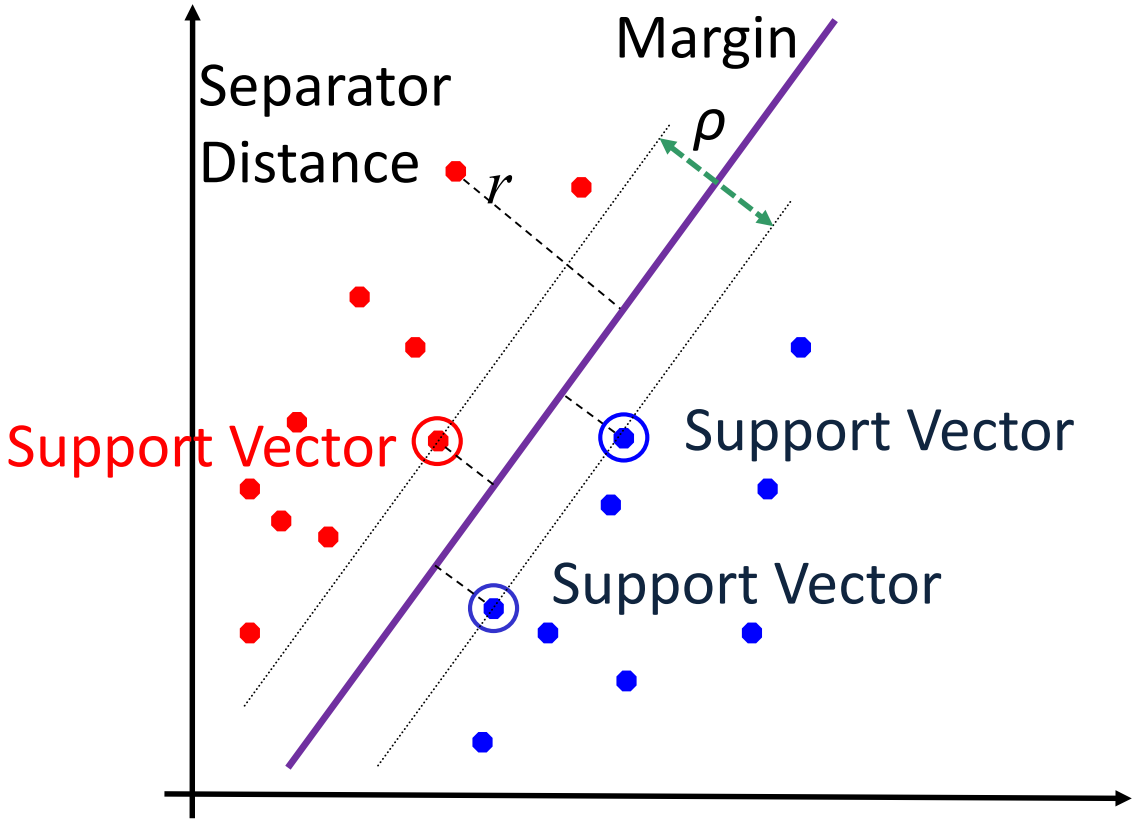
Linear Separator



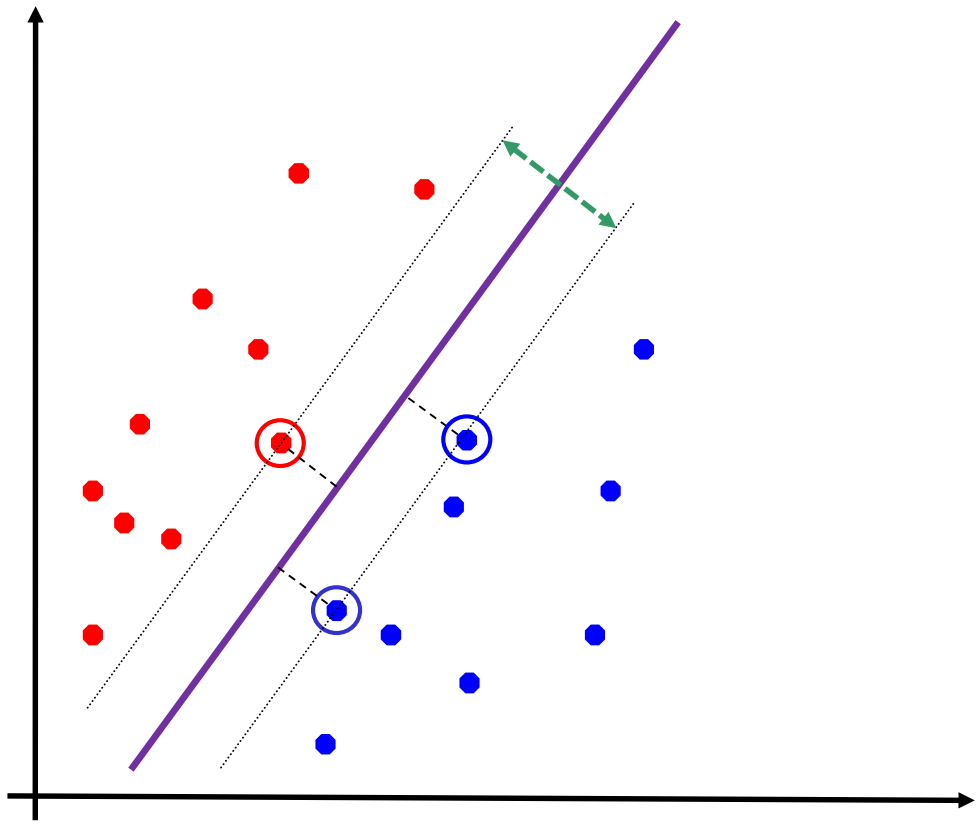
Linear Separator



Optimal Separator

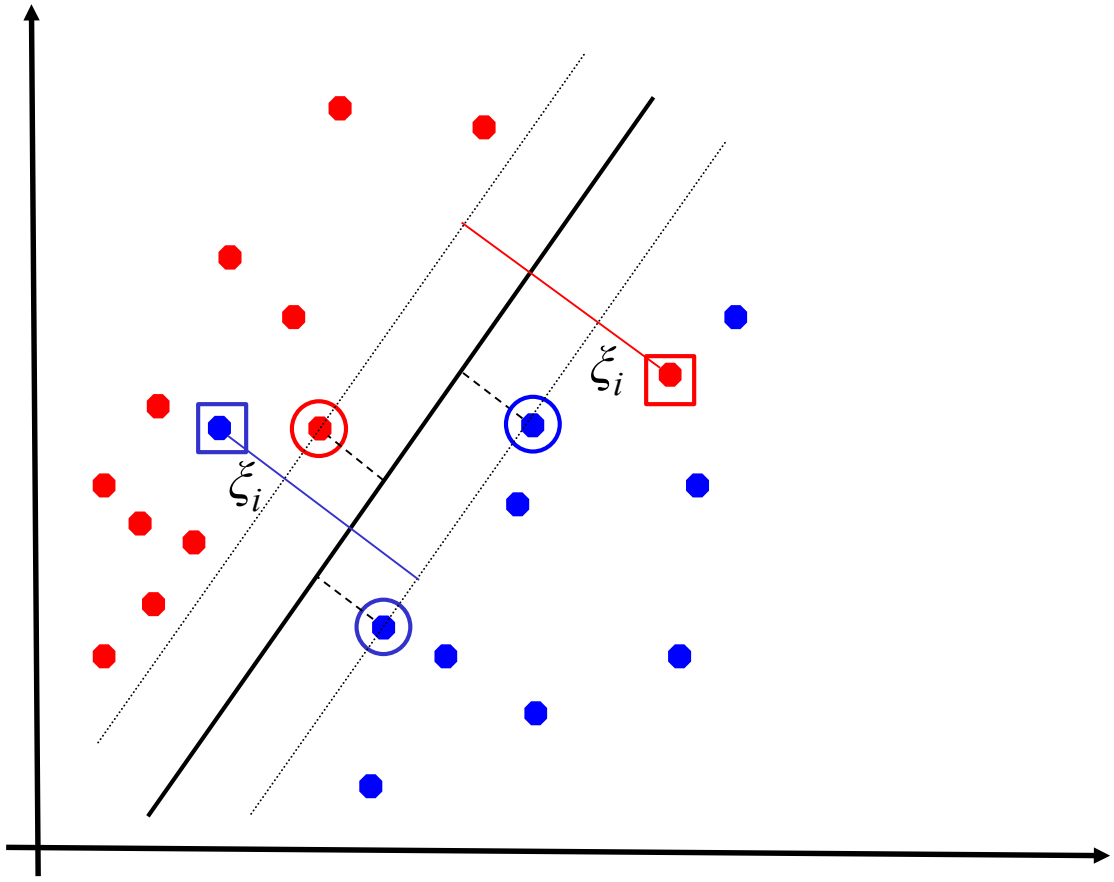


Optimal Separator



- Maximizing the margin is good intuitively.
- Only support vectors matters

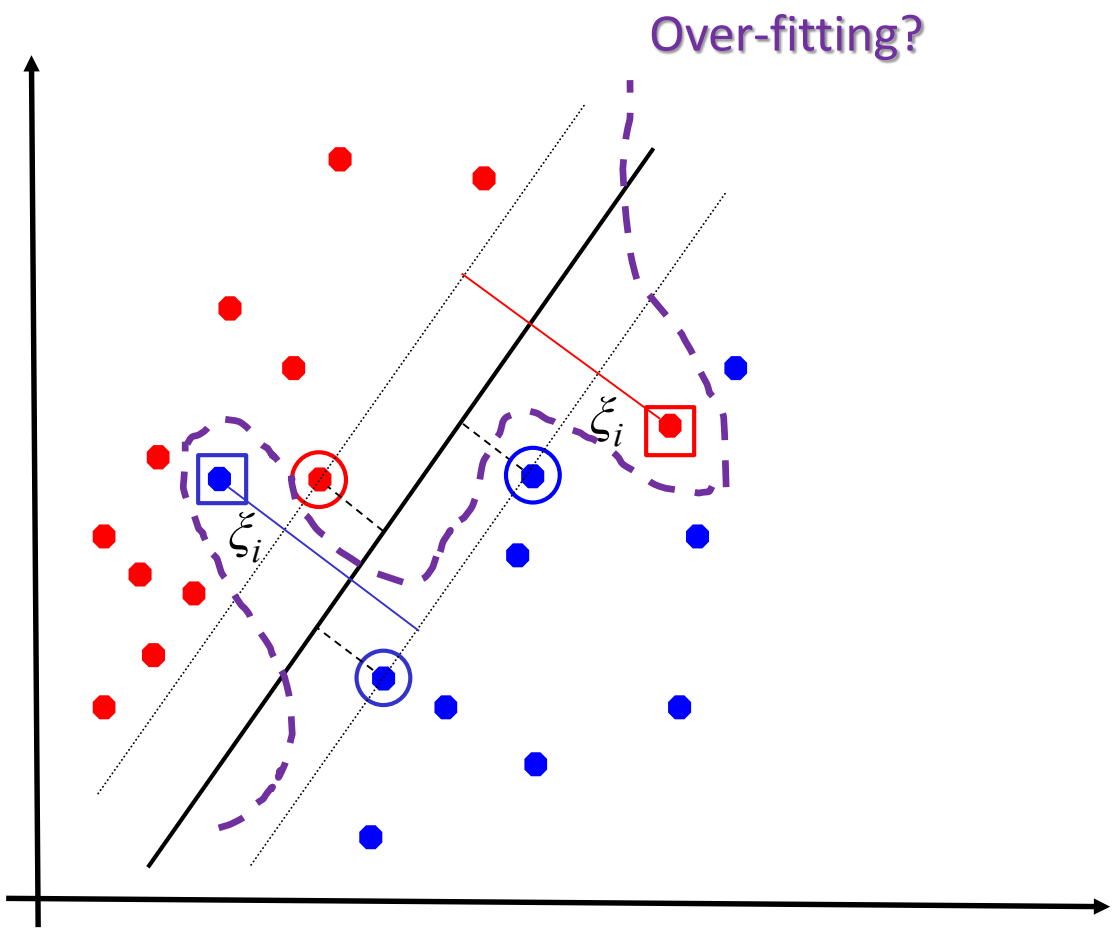
Soft Margin



Solutions:

- Linear solution with penalty
- Non-linear solutions
- Higher dimensions

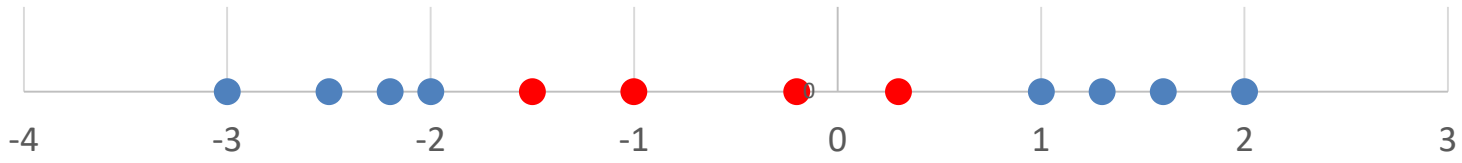
Soft Margin



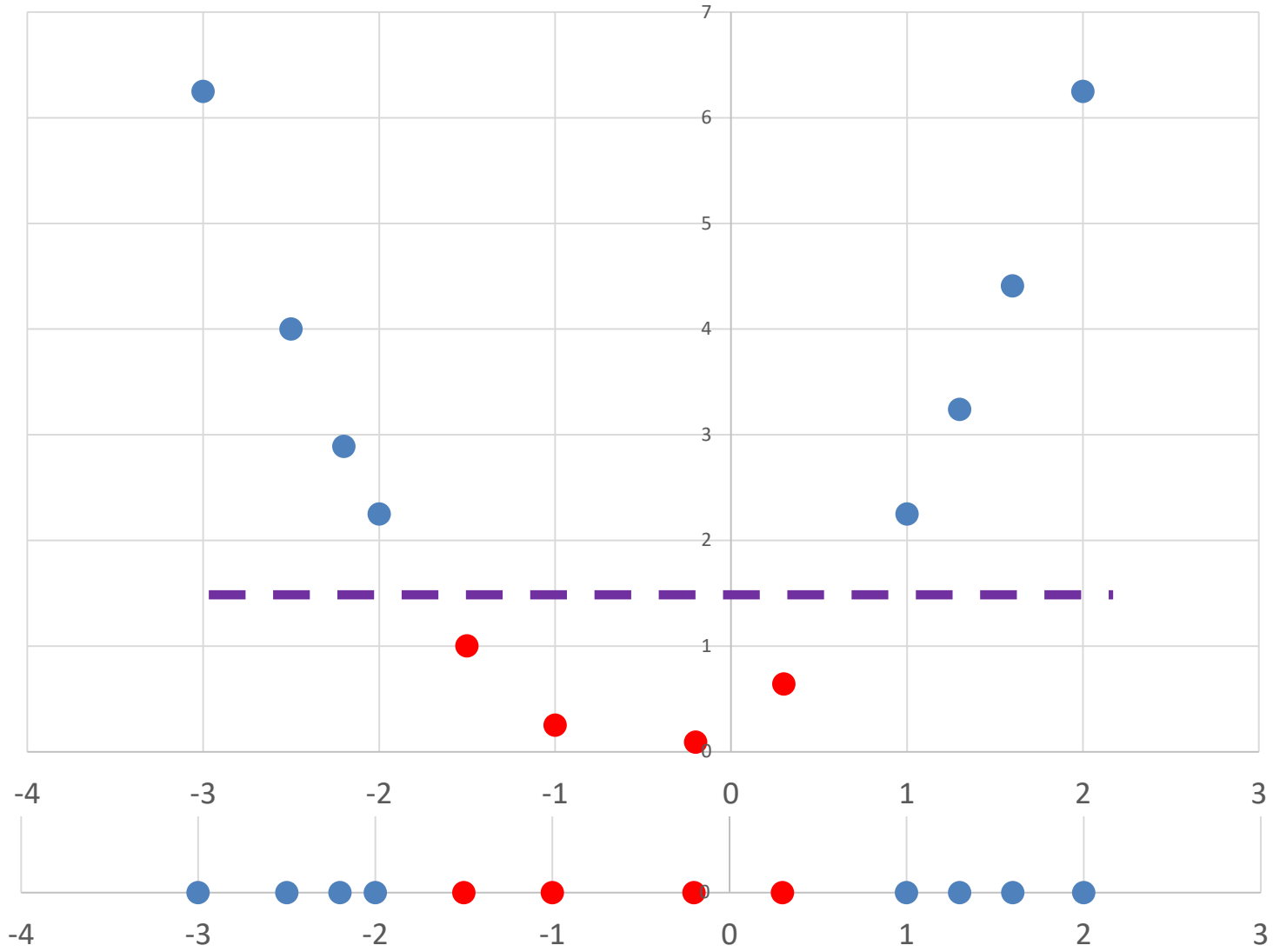
Solutions:

- Linear solution with penalty
- Non-linear solutions
- Higher dimensions

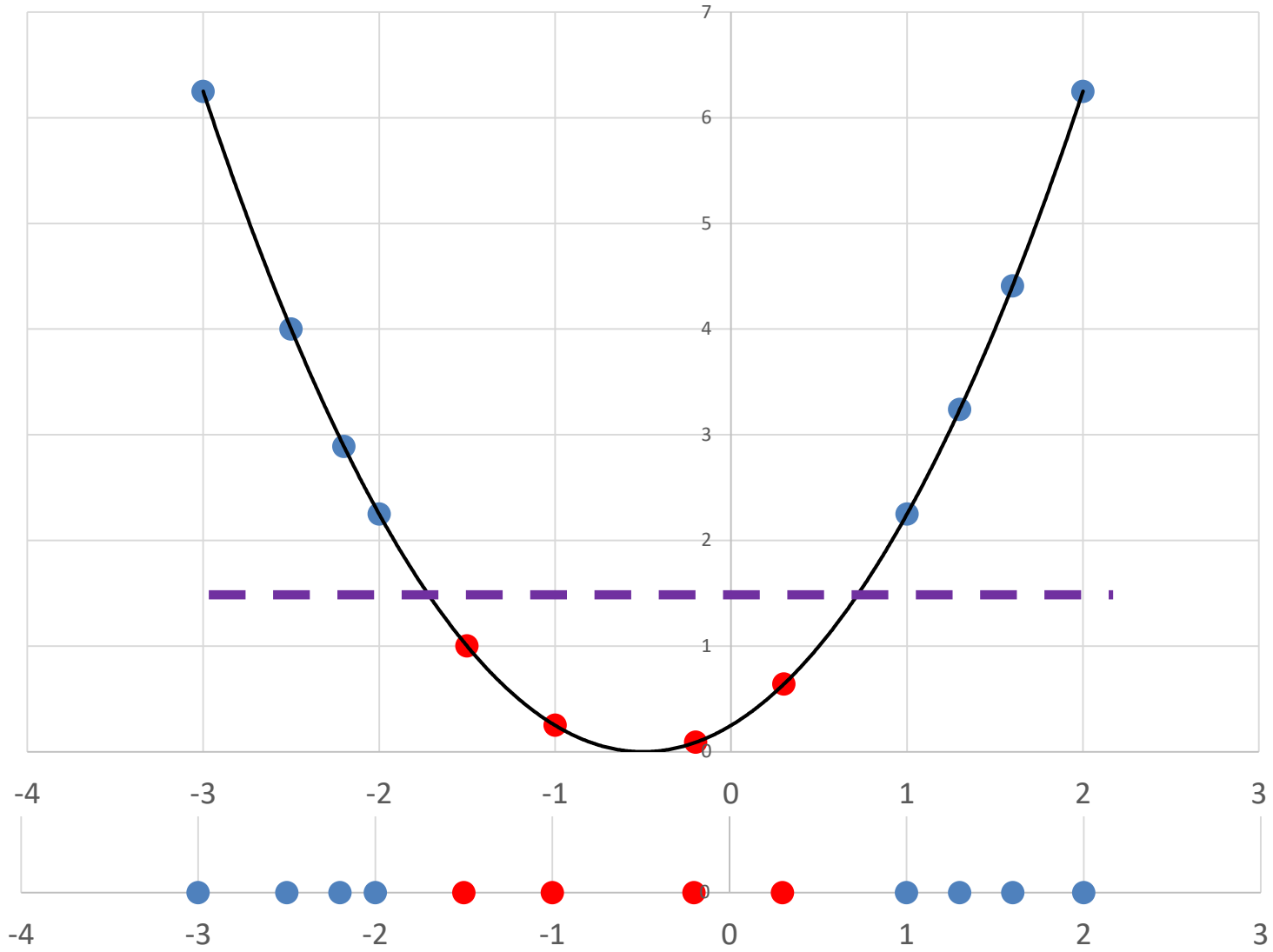
Feature Space of higher dimension



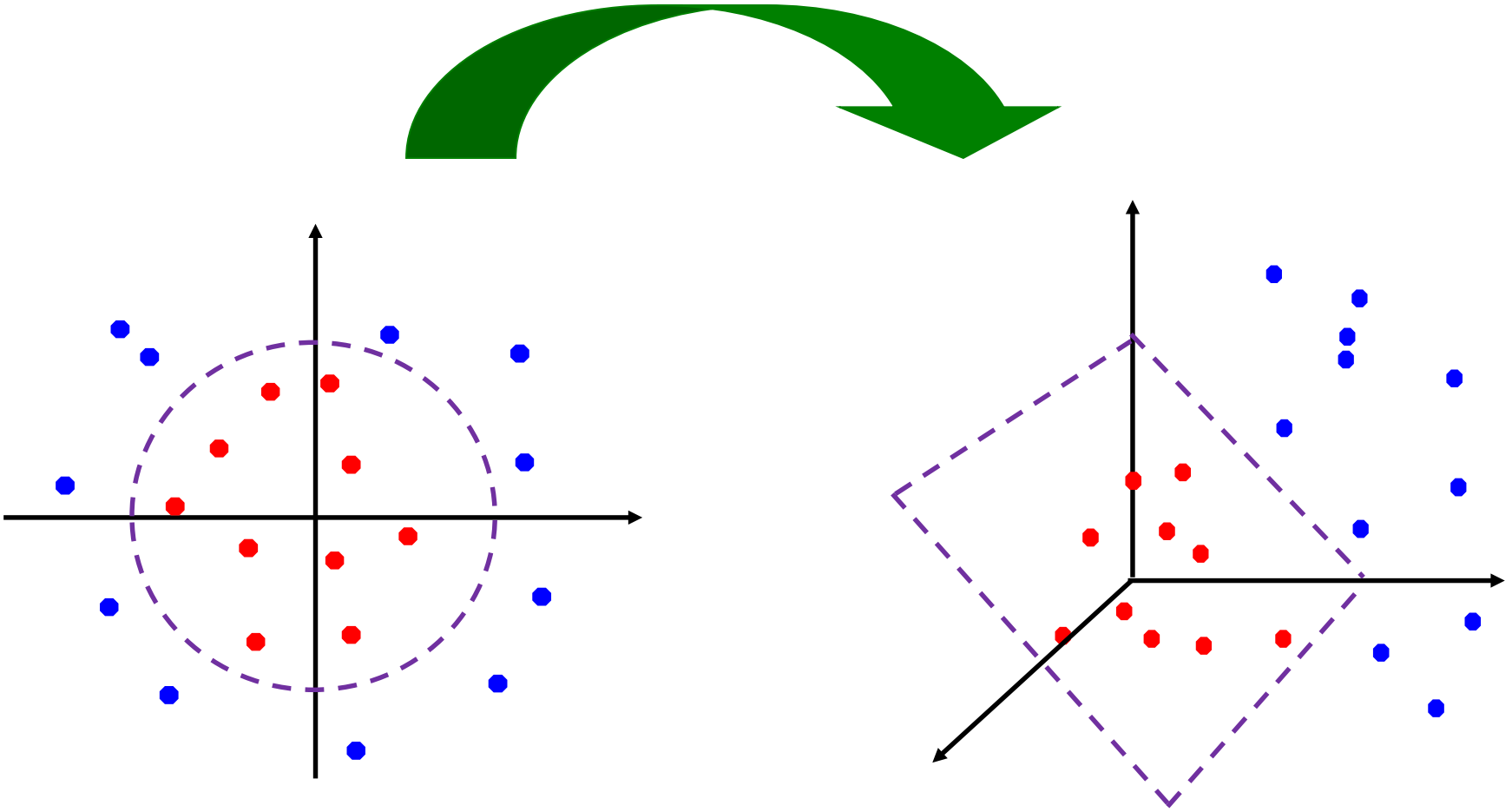
Feature Space of higher dimension



Feature Space of higher dimension



Feature Space of higher dimension



Example

Cupcakes




Muffins




Collecting Data

Recipe	Flour	Sugar	Butter	Milk	Egg	Baking Powder	Salt	Vanilla
Muffin1	2 cups	1/2 cup	1/4 cup	1 cup	2 eggs	1 tbsp	1/2 tsp	1 tsp
Cupcake1	2 cups	3/4 cup	1/2 cup	1 cup	2 eggs	2 tsp	1/2 tsp	1 tsp
…	…	…	…	…	…	…	…	…

Cleaning the Data

Recipe	Flour	Sugar	Other
Muffin1	2 cups	1/2 cup	...
Cupcake1	2 cups	3/4 cup	...

Recipe	Flour	Sugar	Other	Total Volume
Muffin1	47%	24%	...	100%
Cupcake1	42%	21%	...	100%

The Data

Label

Type	Flour	Milk	Sugar	Butter	Egg	Baking Powder	Vanilla	Salt
Muffin	55	28	3	7	5	2	0	0
Muffin	47	24	12	6	9	1	0	0
Muffin	47	23	18	6	4	1	0	0
Muffin	50	25	12	6	5	2	1	0
Muffin	55	27	3	7	5	2	1	0
Muffin	54	27	7	5	5	2	0	0
Muffin	47	26	10	10	4	1	0	0
Muffin	50	17	17	8	6	1	0	0
Muffin	50	17	17	11	4	1	0	0
Cupcake	39	0	26	19	14	1	1	0
Cupcake	34	17	20	20	5	2	1	0
Cupcake	39	13	17	19	10	1	1	0
Cupcake	38	15	23	15	8	0	1	0
Cupcake	42	18	25	9	5	1	0	0
Cupcake	36	14	21	14	11	2	1	0
Cupcake	38	15	31	8	6	1	1	0
Cupcake	36	16	24	12	9	1	1	0
Cupcake	34	17	23	11	13	0	1	0

The 8 Dimensional Data

Label

Not Useful

Type	Flour	Milk	Sugar	Butter	Egg	Baking Powder	Vanilla	Salt
Muffin	55	28	3	7	5	2	0	0
Muffin	47	24	12	6	9	1	0	0
Muffin	47	23	18	6	4	1	0	0
Muffin	50	25	12	6	5	2	1	0
Muffin	55	27	3	7	5	2	1	0
Muffin	54	27	7	5	5	2	0	0
Muffin	47	26	10	10	4	1	0	0
Muffin	50	17	17	8	6	1	0	0
Muffin	50	17	17	11	4	1	0	0
Cupcake	39	0	26	19	14	1	1	0
Cupcake	34	17	20	20	5	2	1	0
Cupcake	39	13	17	19	10	1	1	0
Cupcake	38	15	23	15	8	0	1	0
Cupcake	42	18	25	9	5	1	0	0
Cupcake	36	14	21	14	11	2	1	0
Cupcake	38	15	31	8	6	1	1	0
Cupcake	36	16	24	12	9	1	1	0
Cupcake	34	17	23	11	13	0	1	0

The 8 Dimensional Data

Label

Maybe

Type	Flour	Milk	Sugar	Butter	Egg	Baking Powder	Vanilla	Salt
Muffin	55	28	3	7	5	2	0	0
Muffin	47	24	12	6	9	1	0	0
Muffin	47	23	18	6	4	1	0	0
Muffin	50	25	12	6	5	2	1	0
Muffin	55	27	3	7	5	2	1	0
Muffin	54	27	7	5	5	2	0	0
Muffin	47	26	10	10	4	1	0	0
Muffin	50	17	17	8	6	1	0	0
Muffin	50	17	17	11	4	1	0	0
Cupcake	39	0	26	19	14	1	1	0
Cupcake	34	17	20	20	5	2	1	0
Cupcake	39	13	17	19	10	1	1	0
Cupcake	38	15	23	15	8	0	1	0
Cupcake	42	18	25	9	5	1	0	0
Cupcake	36	14	21	14	11	2	1	0
Cupcake	38	15	31	8	6	1	1	0
Cupcake	36	16	24	12	9	1	1	0
Cupcake	34	17	23	11	13	0	1	0

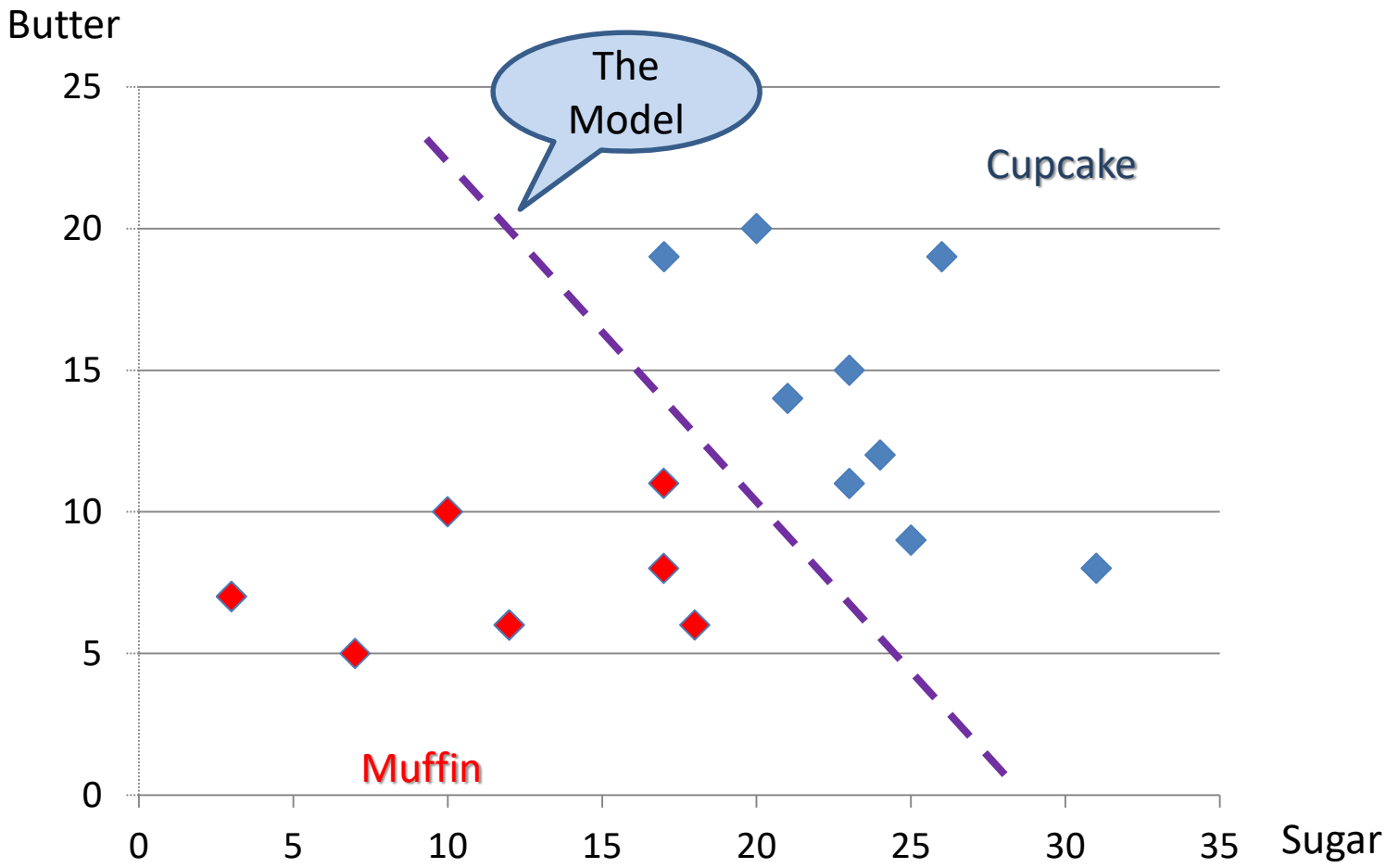
The 8 Dimensional Data

Label

Our Selection

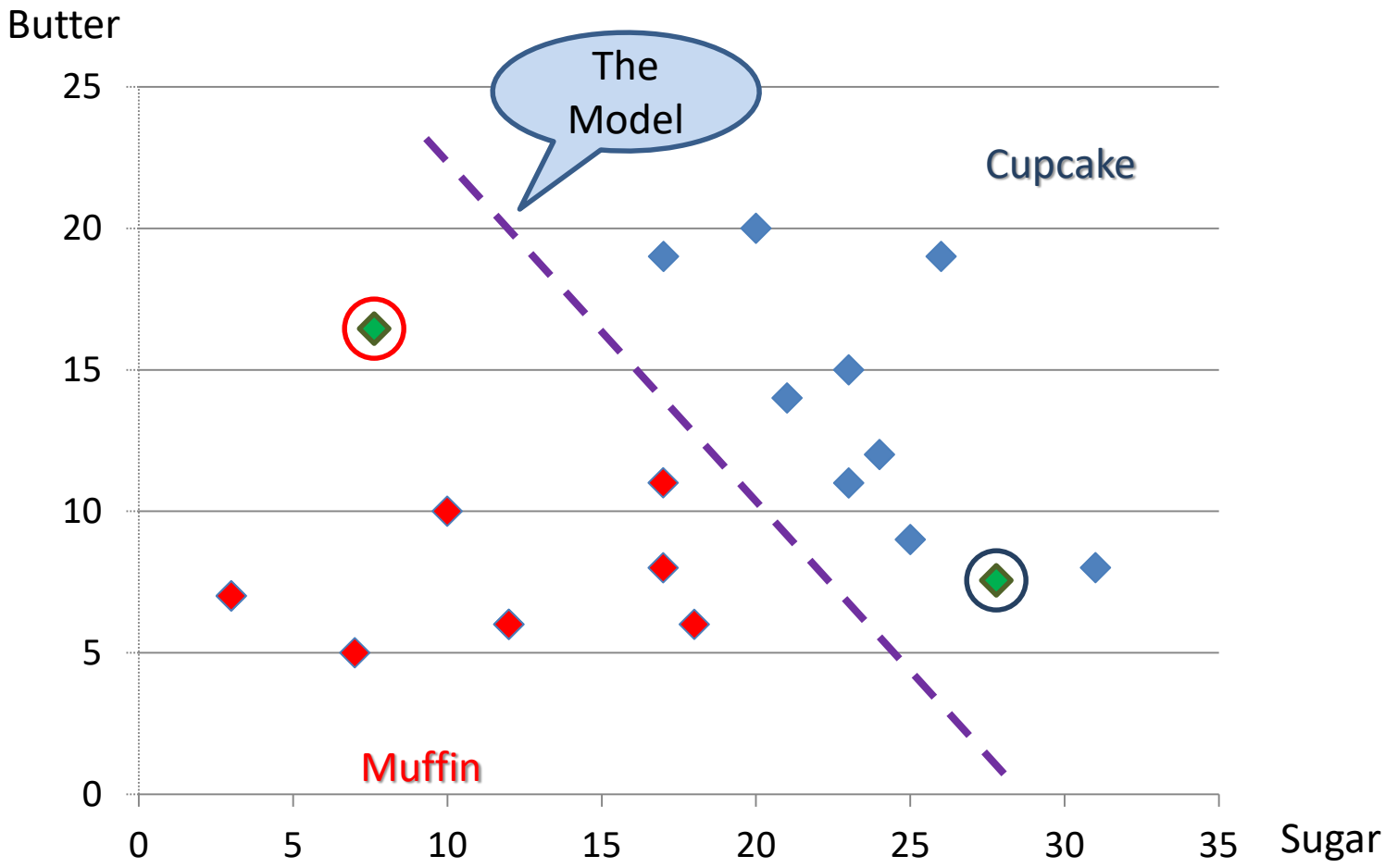
Type	Flour	Milk	Sugar	Butter	Egg	Baking Powder	Vanilla	Salt
Muffin	55	28	3	7	5	2	0	0
Muffin	47	24	12	6	9	1	0	0
Muffin	47	23	18	6	4	1	0	0
Muffin	50	25	12	6	5	2	1	0
Muffin	55	27	3	7	5	2	1	0
Muffin	54	27	7	5	5	2	0	0
Muffin	47	26	10	10	4	1	0	0
Muffin	50	17	17	8	6	1	0	0
Muffin	50	17	17	11	4	1	0	0
Cupcake	39	0	26	19	14	1	1	0
Cupcake	34	17	20	20	5	2	1	0
Cupcake	39	13	17	19	10	1	1	0
Cupcake	38	15	23	15	8	0	1	0
Cupcake	42	18	25	9	5	1	0	0
Cupcake	36	14	21	14	11	2	1	0
Cupcake	38	15	31	8	6	1	1	0
Cupcake	36	16	24	12	9	1	1	0
Cupcake	34	17	23	11	13	0	1	0

SVM Training



Cupcake or Muffin?

SVM Testing

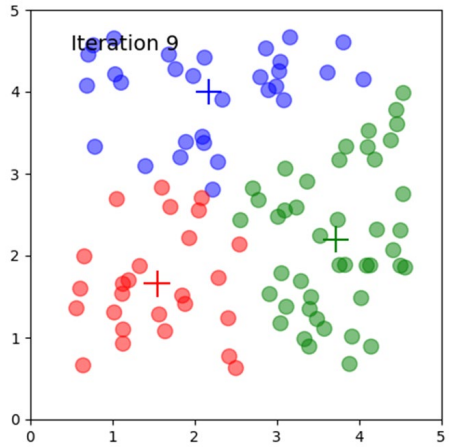

Butter
25
20
15
10
5
0
The Model
Cupcake
Muffin
0
5
10
15
20
25
30
35
Sugar

3. Unsupervised Learning

- Clustering groups a set of objects in such a way that objects in the same group are more similar to each other than to those in other groups.
- *K*-means clustering aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean.

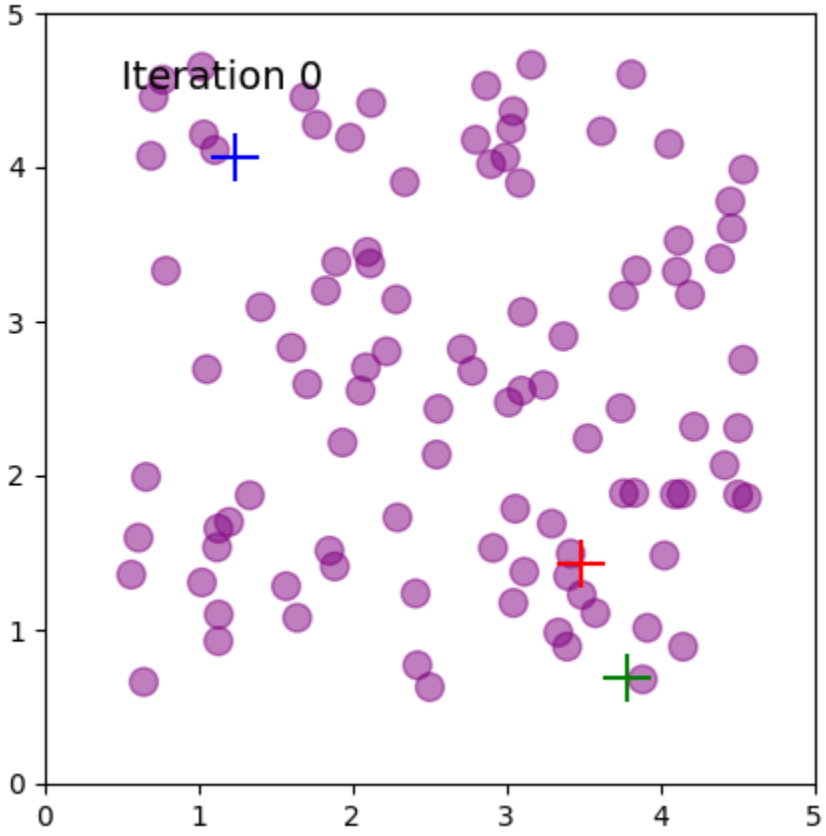
K-means algorithm

1) Pick *k* cluster centers randomly.
2) Assign every data point to its nearest cluster center.
3) Move each cluster center to the mean of its assigned data points.
4) Repeat Steps 2-3 until it convergence.

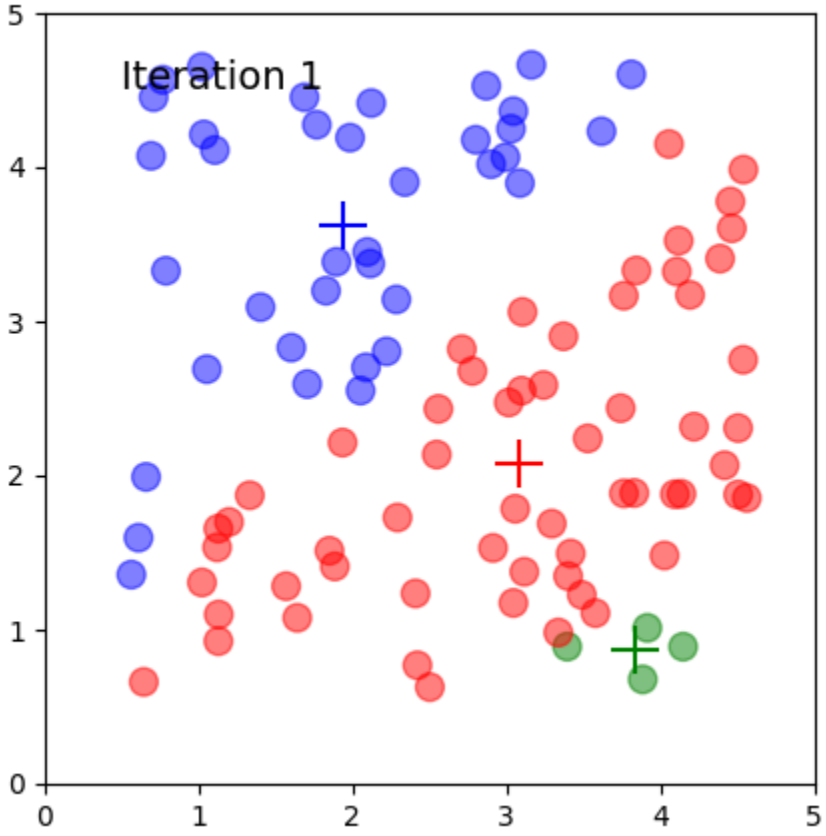


Slides from Wash Univ. BIO5488 lecture, 2004

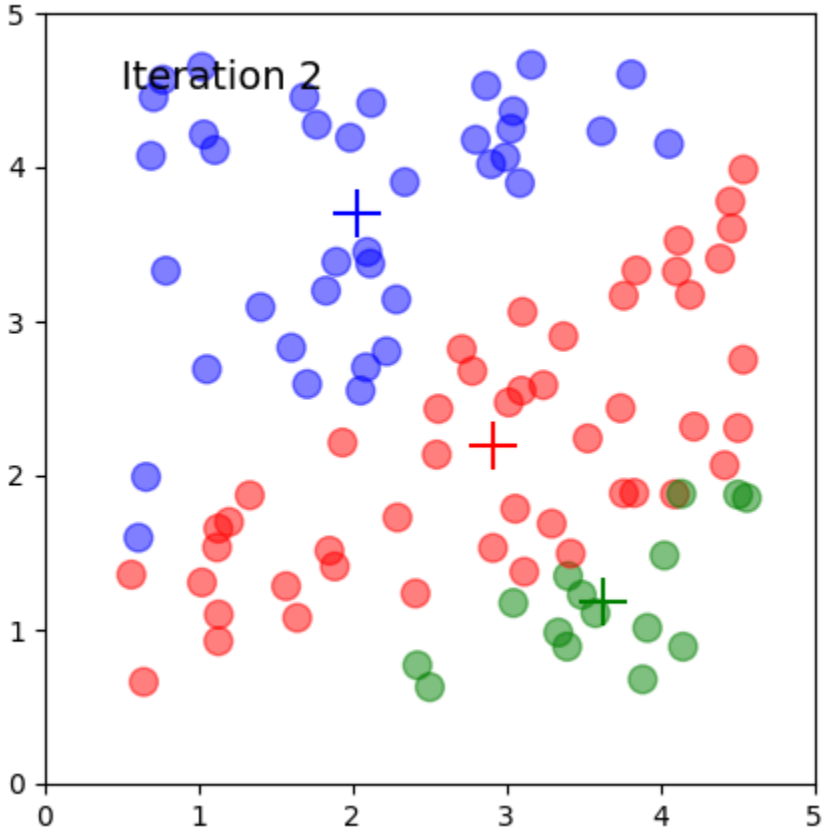
0



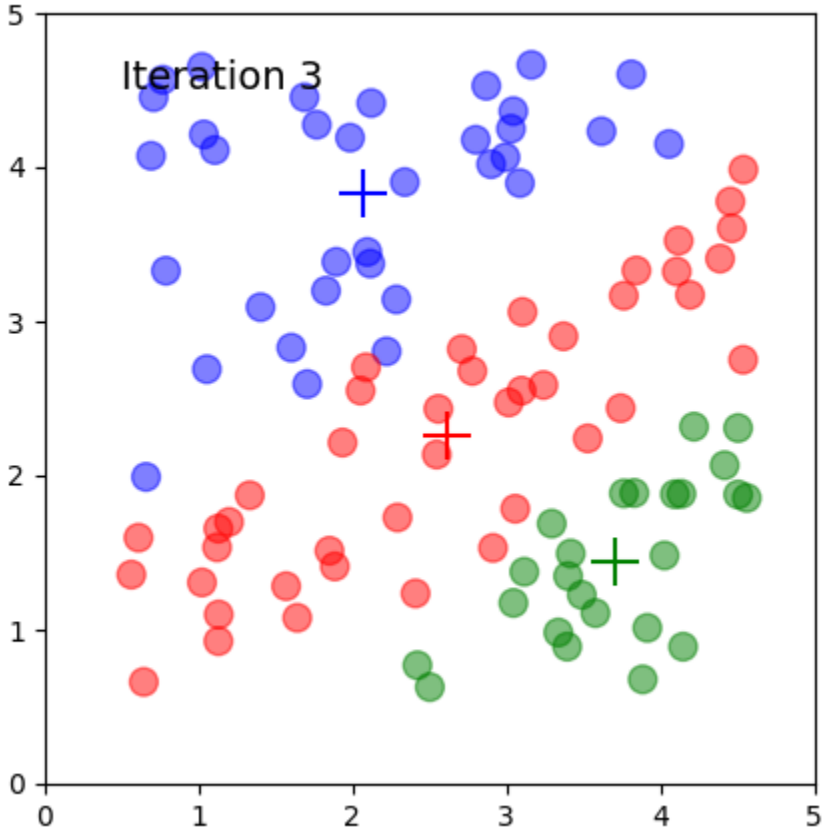
1



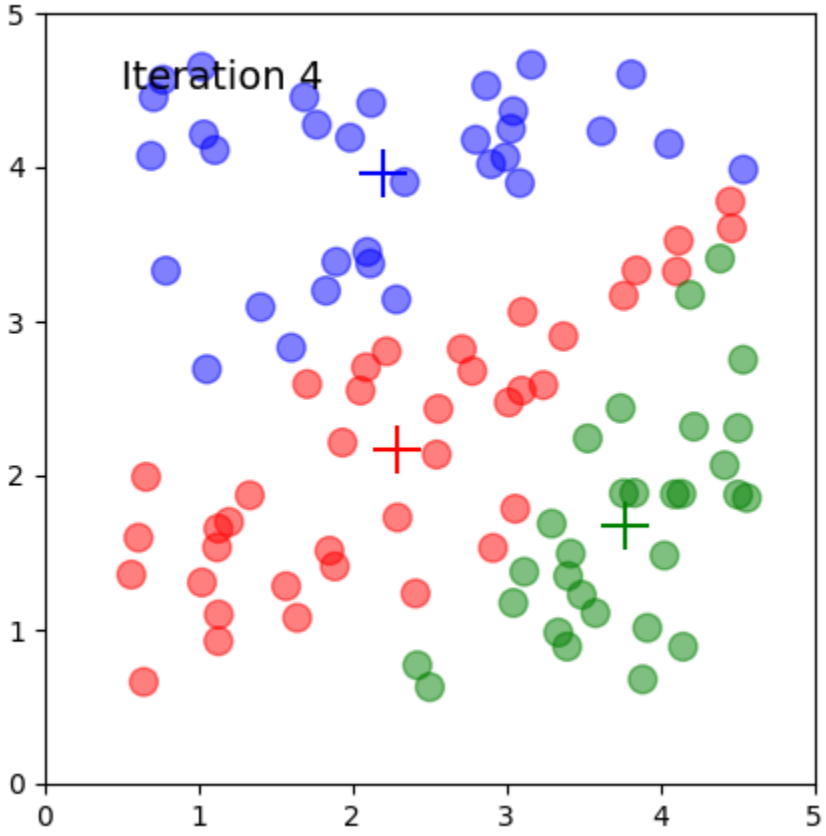
2



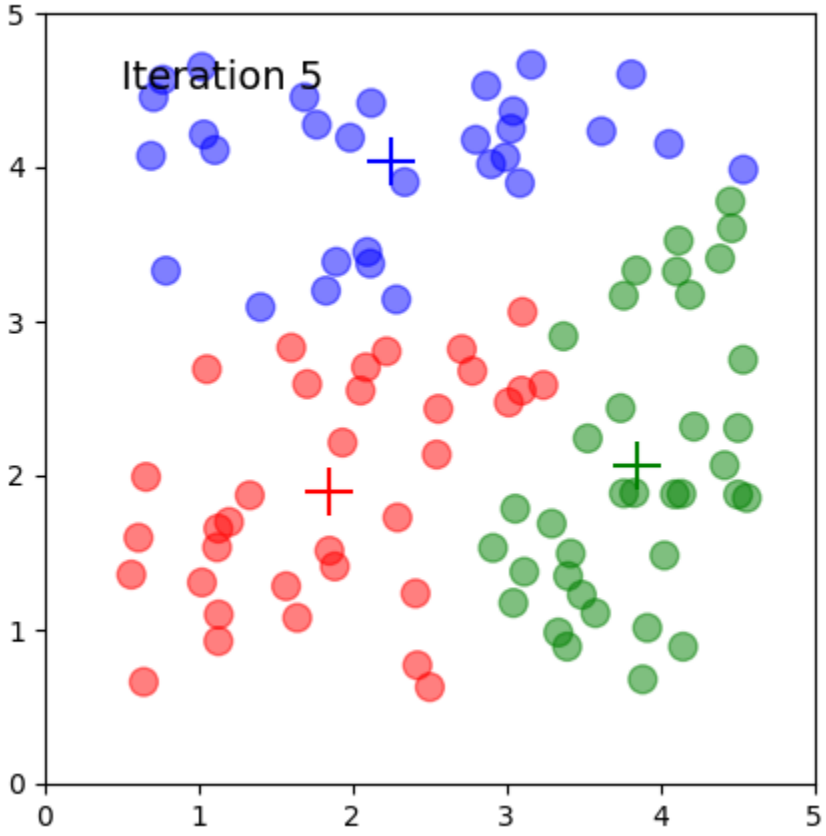
3



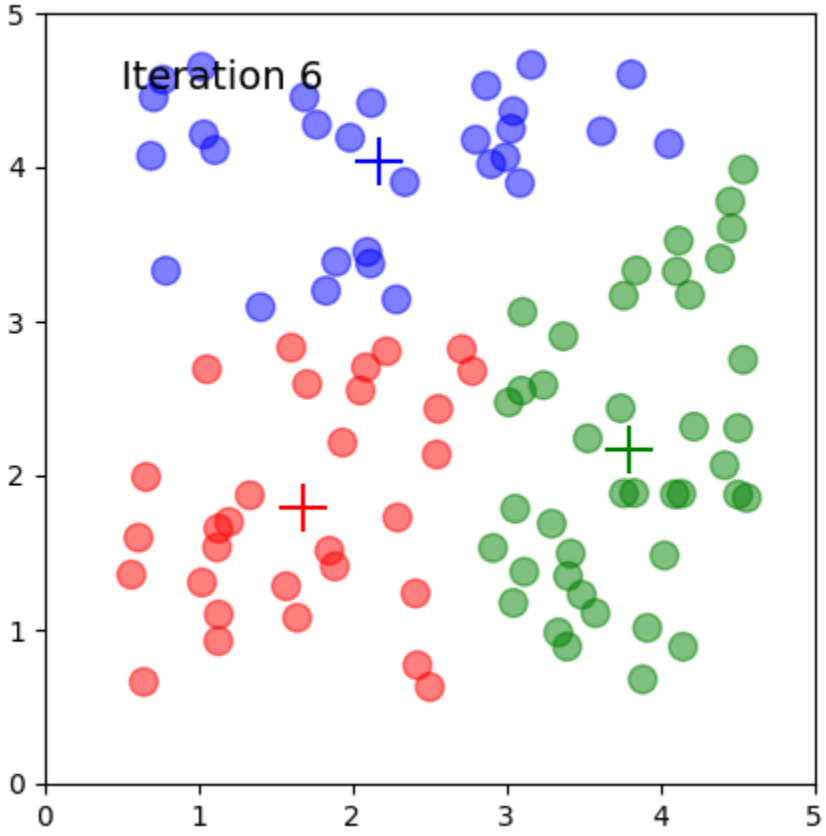
4



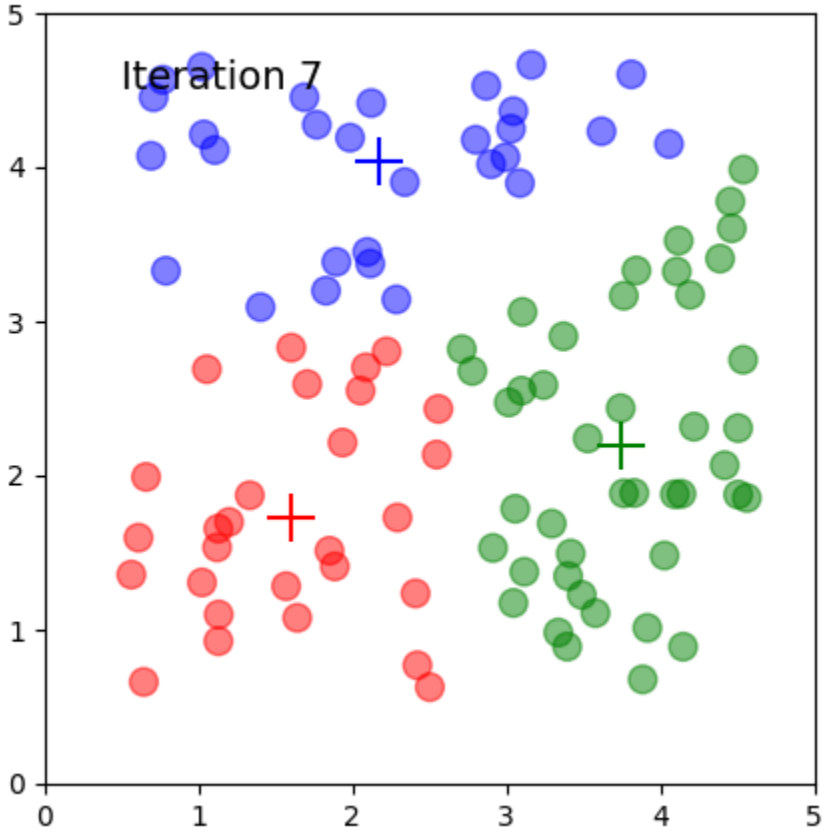
5



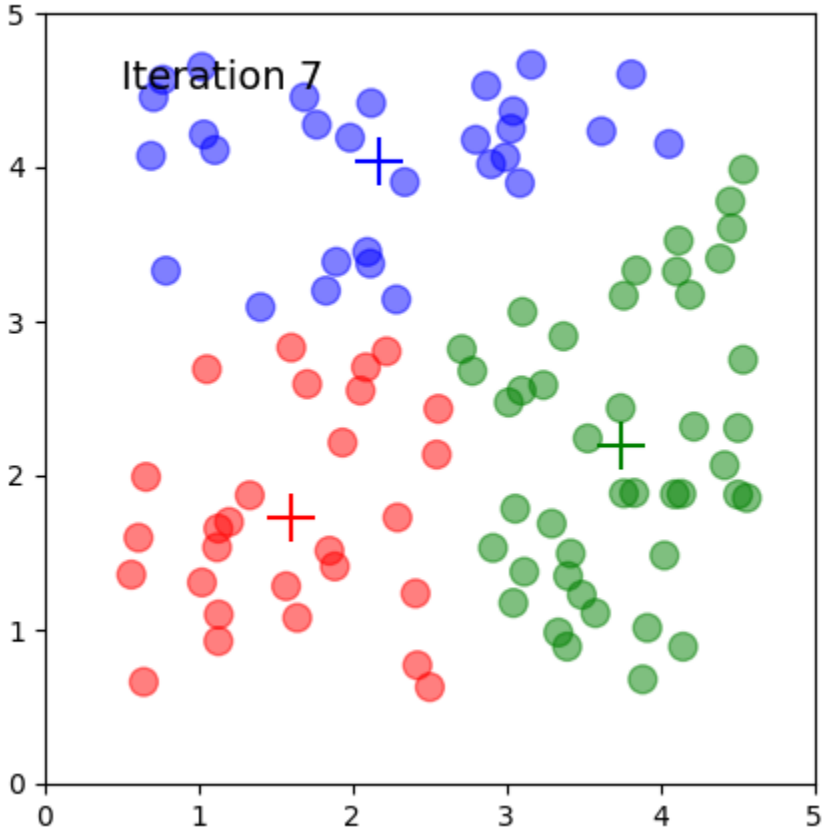
6



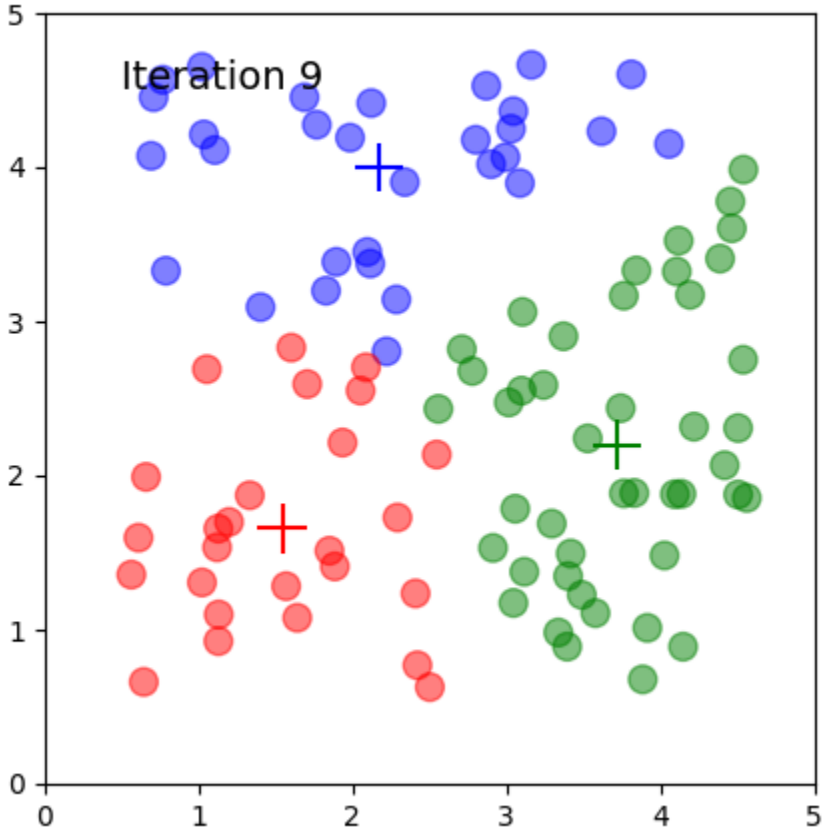
7



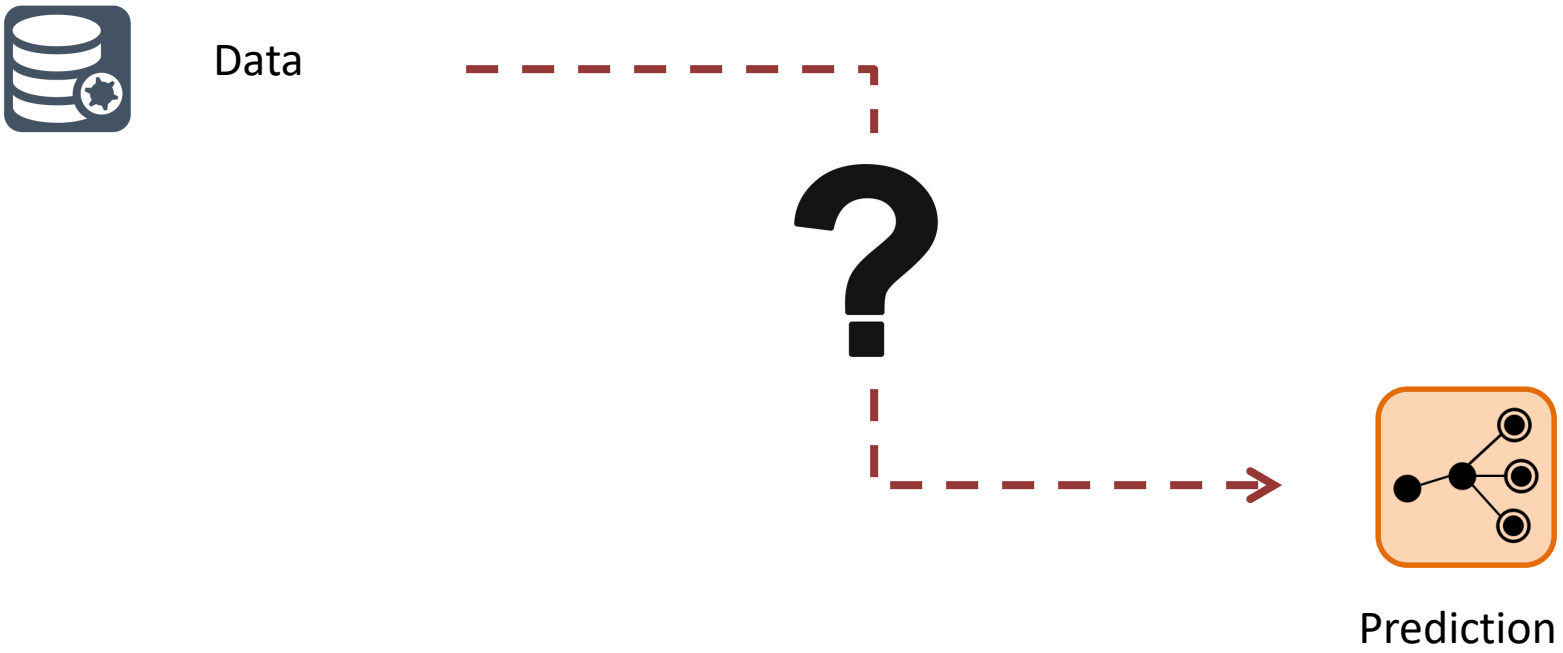
8



9



4. Machine Learning Process



Machine Learning Process

Data

Machine Learning Process

Labels

+

Data

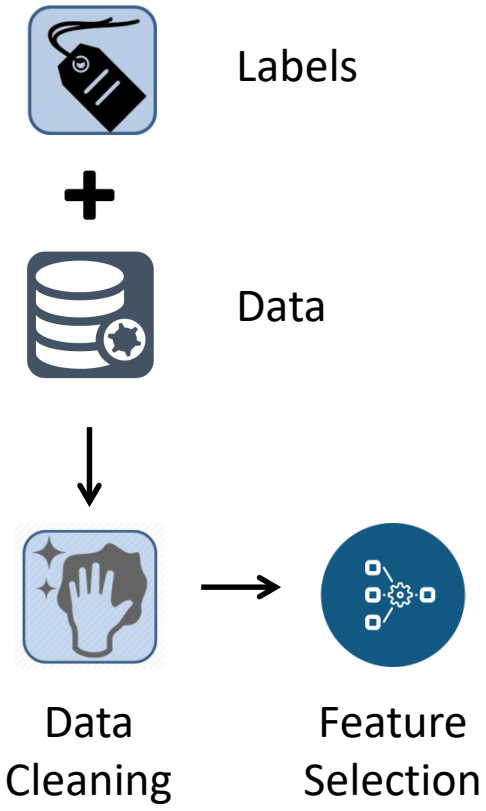
Machine Learning Process

Labels

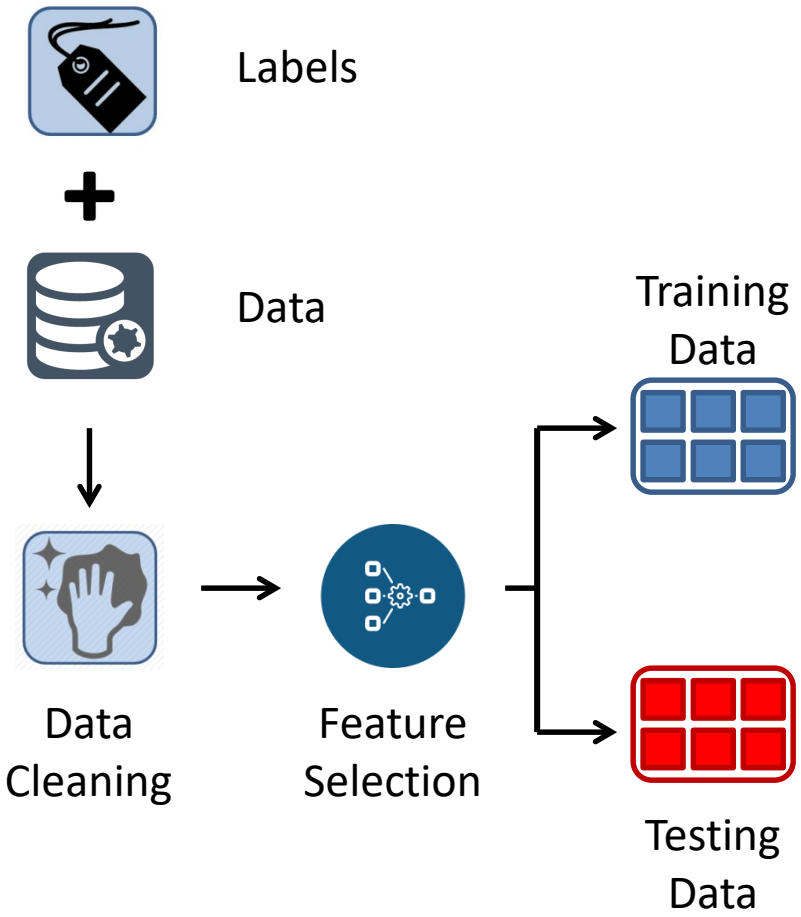
Data

Data
Cleaning

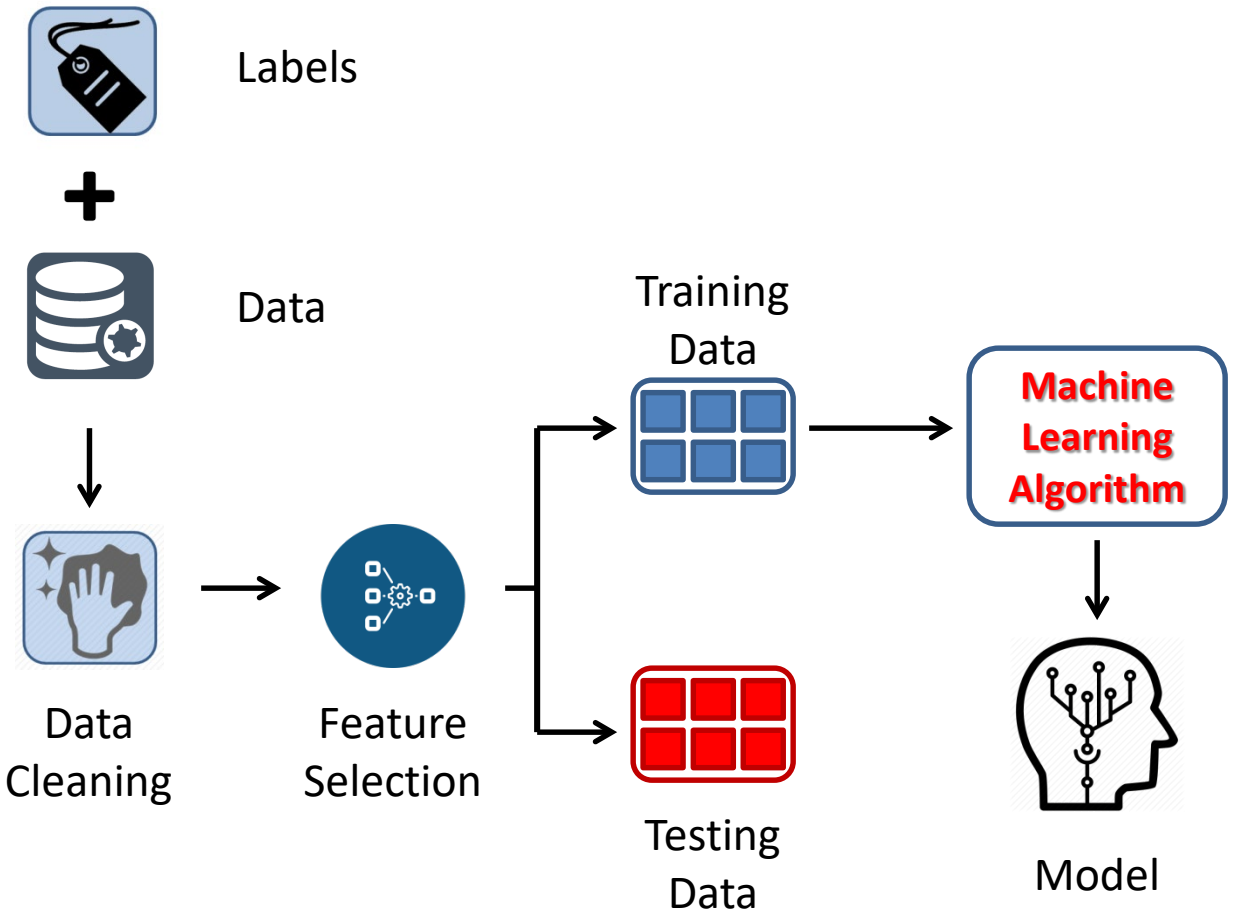
Machine Learning Process



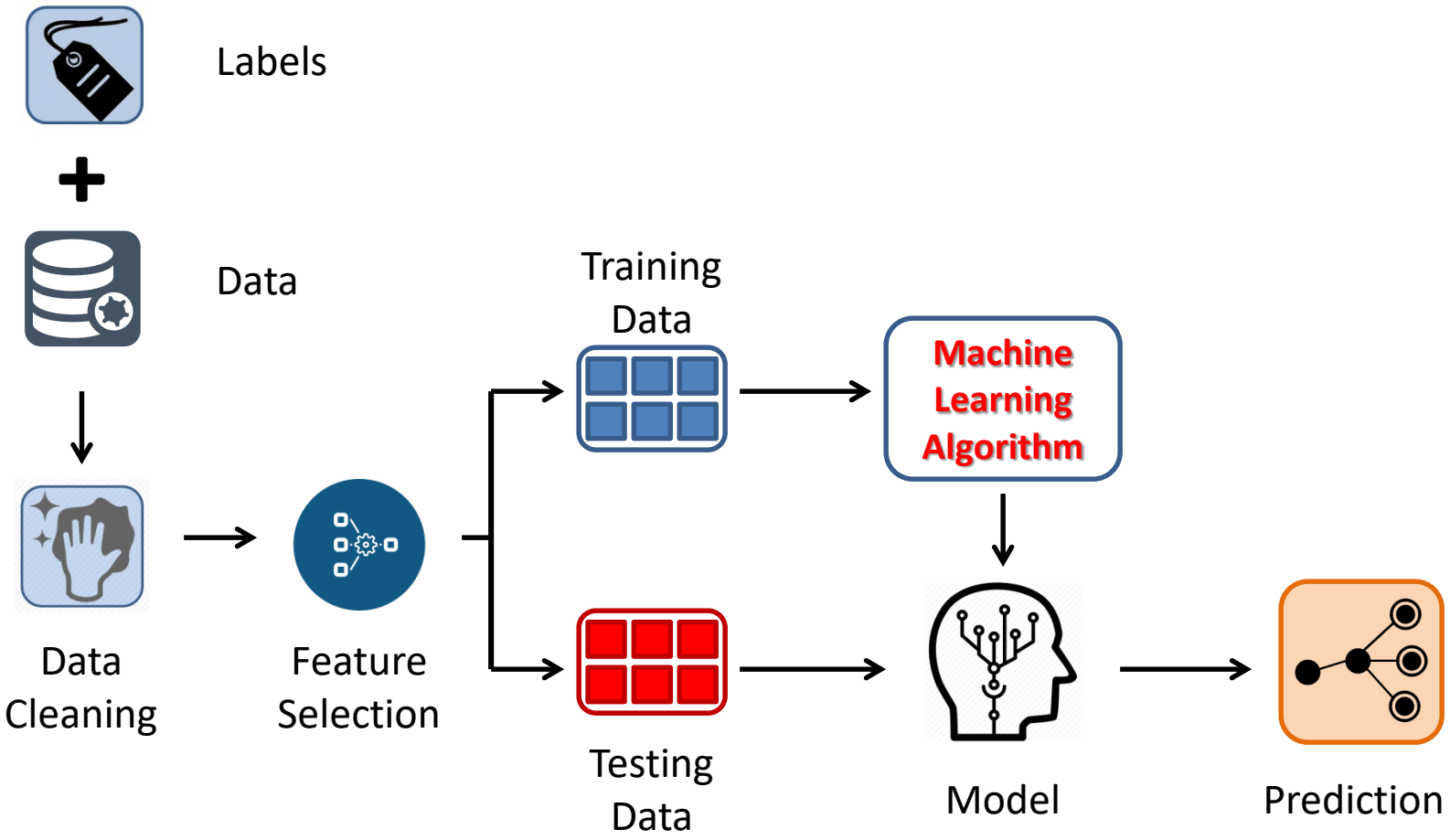
Machine Learning Process



Machine Learning Process



Machine Learning Process



5. Summary

- ML tends to work well when:
  - Learning a "simple" concept
  - There is lots of data available
  - There are attributes that distinguish the classes.
- ML tends to work poorly when:
  - Learning complex concepts from small amount of data
  - It is asked to perform on new types of data
- ML Sometimes requires tuning (parameters)

Credits

- AI for Everyone, Andrew Ng, https://www.coursera.org/learn/ai-for-everyone,
- Support Vector Machines: A Visual Explanation with Sample Python Code, Alice Zhao, https://www.youtube.com/watch?v=N1vOgolbjSc
- K-means illustrator, Wesam Elshamy, https://github.com/welshamy/k-means-illustrator