به نام خدا

کارگاه علم داده با پایتون پیشرفته

جلسه هفتم: ماشین بردار پشتیبان (SVM)

مدرس :

مهرناز جلیلی

دانشجو کارشناسی ارشد علم داده ها

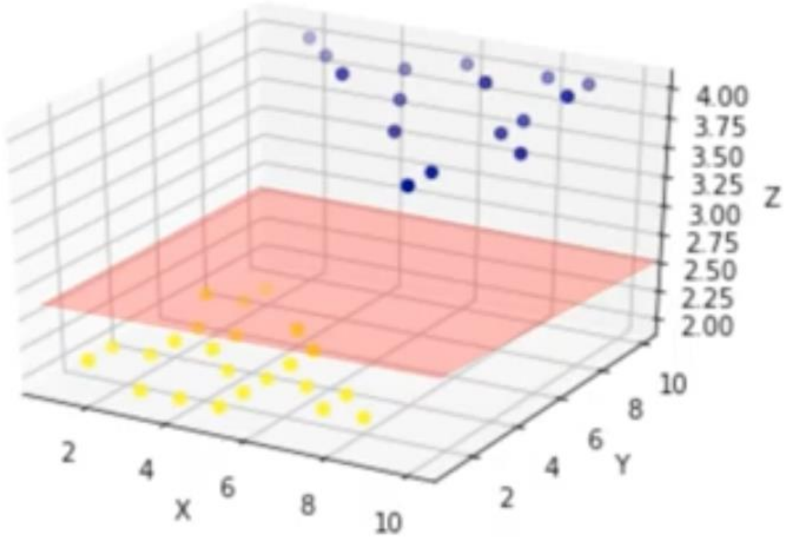
دانشگاه شهید بهشتی

Classification

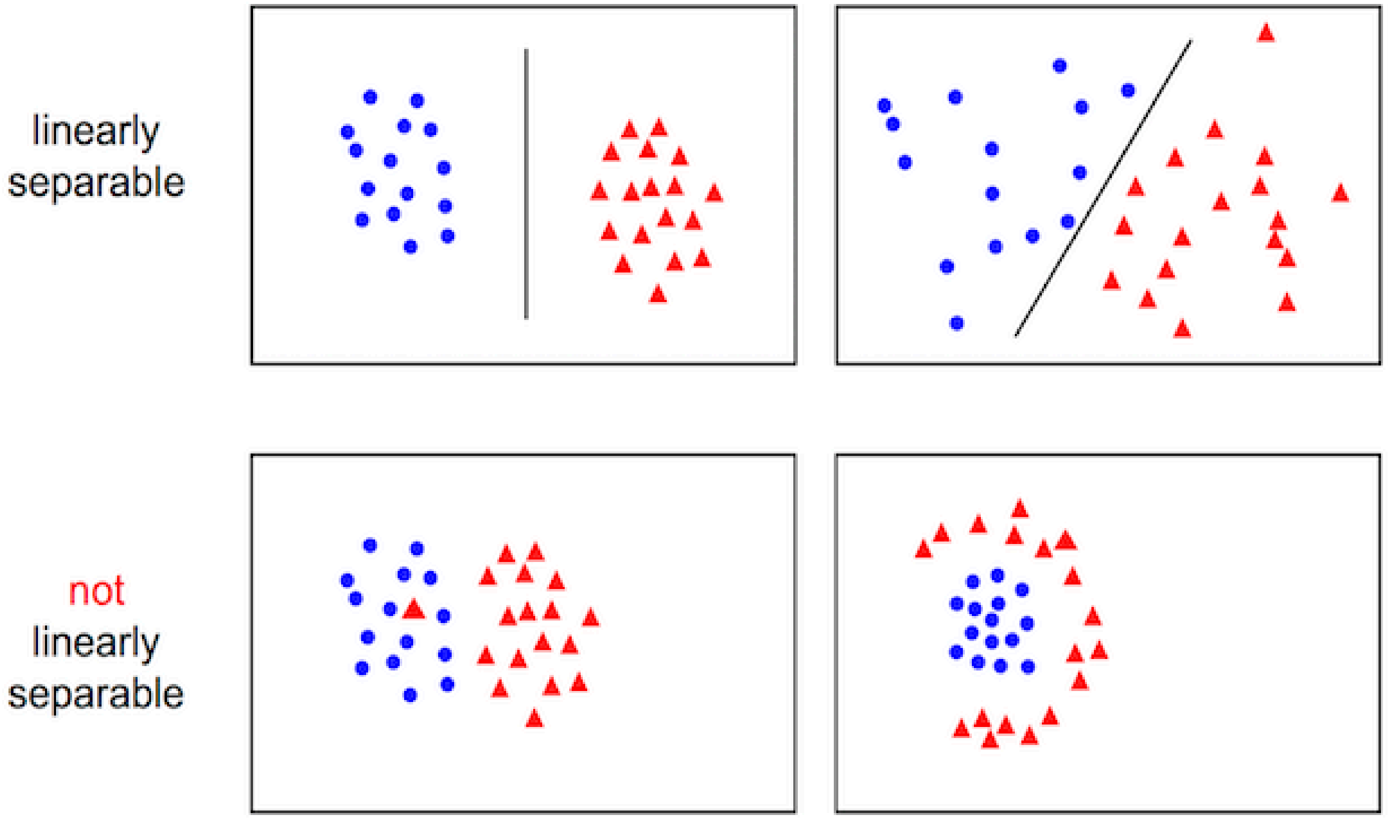
Support Vector Machines

- supervised
- classifier based on separator
- mapping data to high-dimensional so a hyperplane separator can be drawn
- Lots of real world datas are Linearly non separable , but what if we go to a higher dimension? ;)

Clump	UnifSize	UnifShape	MargAdh	SingEpiSize	BareNuc	BlandChrom	NormNucl	Mit	Class
5	1	1	1	2	1	3	1	1	benign
5	4	4	5	7	10	3	2	1	benign
3	1	1	1	2	2	3	1	1	malignant
6	8	8	1	3	4	3	7	1	benign
4	1	1	3	2	1	3	1	1	benign
8	10	10	8	7	10		7	1	malignant
1	1	1	1	2	10	3	1	1	benign
2	1	2	H	2	1	3	1	1	benign
2	1	1	1	2	1	1	1	5	benign
4	2	1	1	2	1	2	1	1	benign



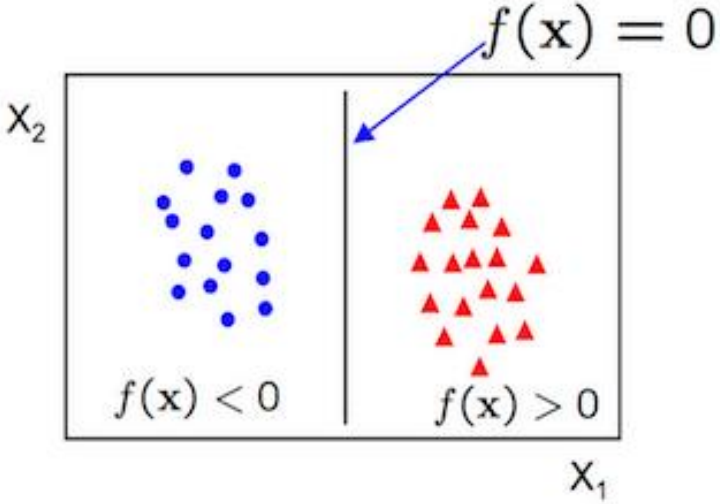
Linear separability



Linear classifiers

A linear classifier has the form

$$f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$$

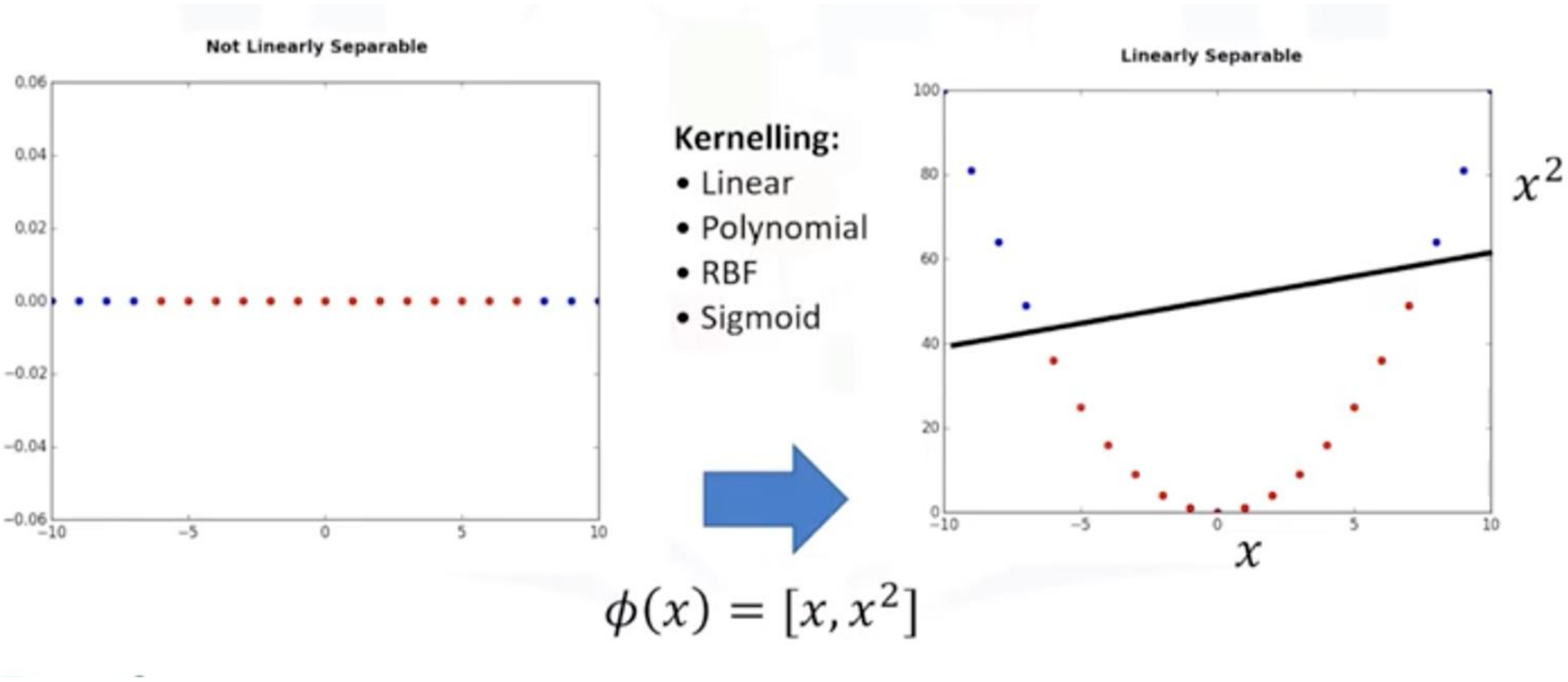


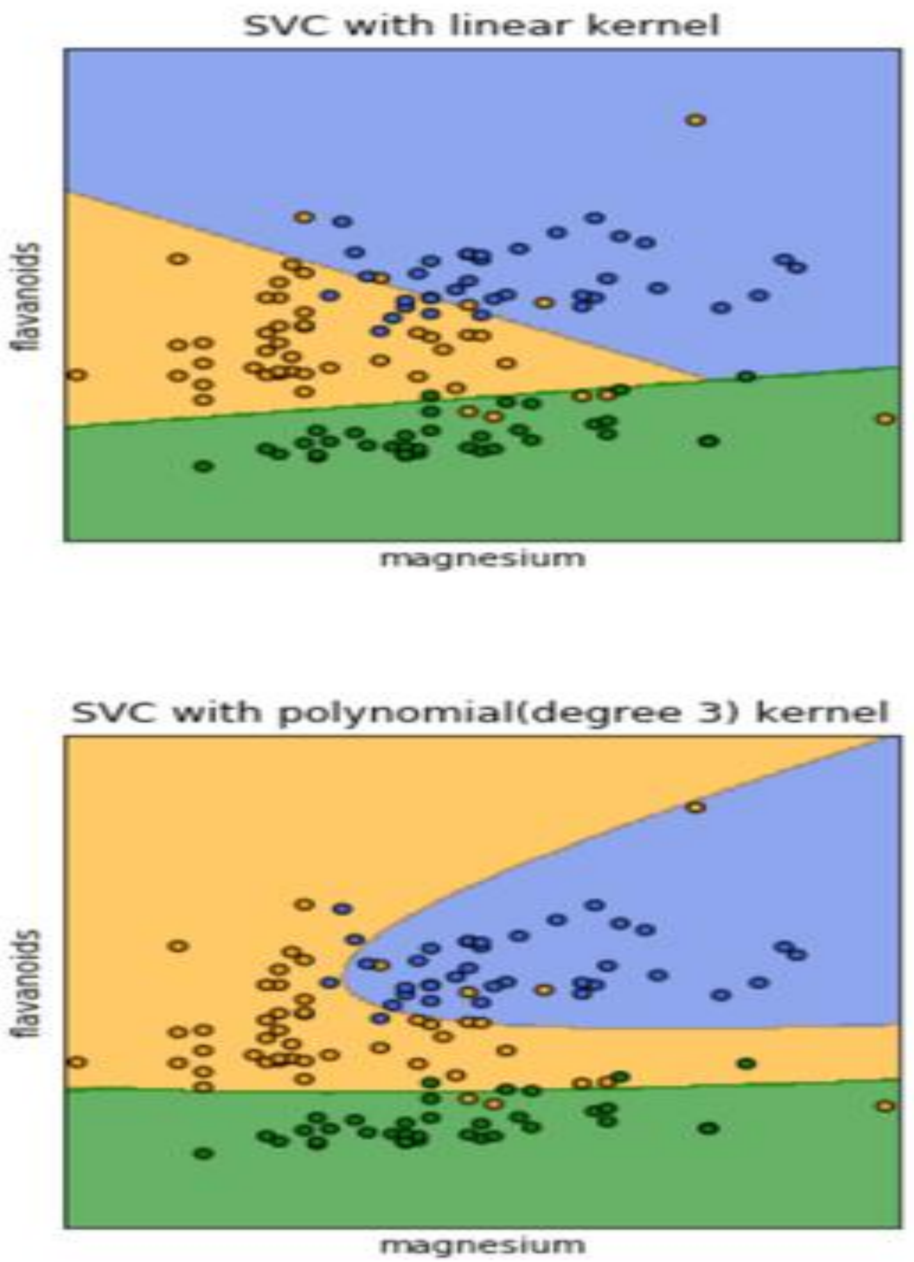
- in 2D the discriminant is a line
- $\mathbf{w}$ is the normal to the line, and b the bias
- $\mathbf{w}$ is known as the weight vector

Classification

Support Vector Machines

- but… how to move to n-dimention?
- there are different kernel functions
- our libraries will do, we will just compare
- How to find the hyperplane?




SVC with linear kernel
flavanoids
magnesium
SVC with RBF kernel
flavanoids
magnesium
SVC with polynomial(degree 3) kernel
flavanoids
magnesium
SVC with sigmoid kernel
flavanoids
magnesium

Classification

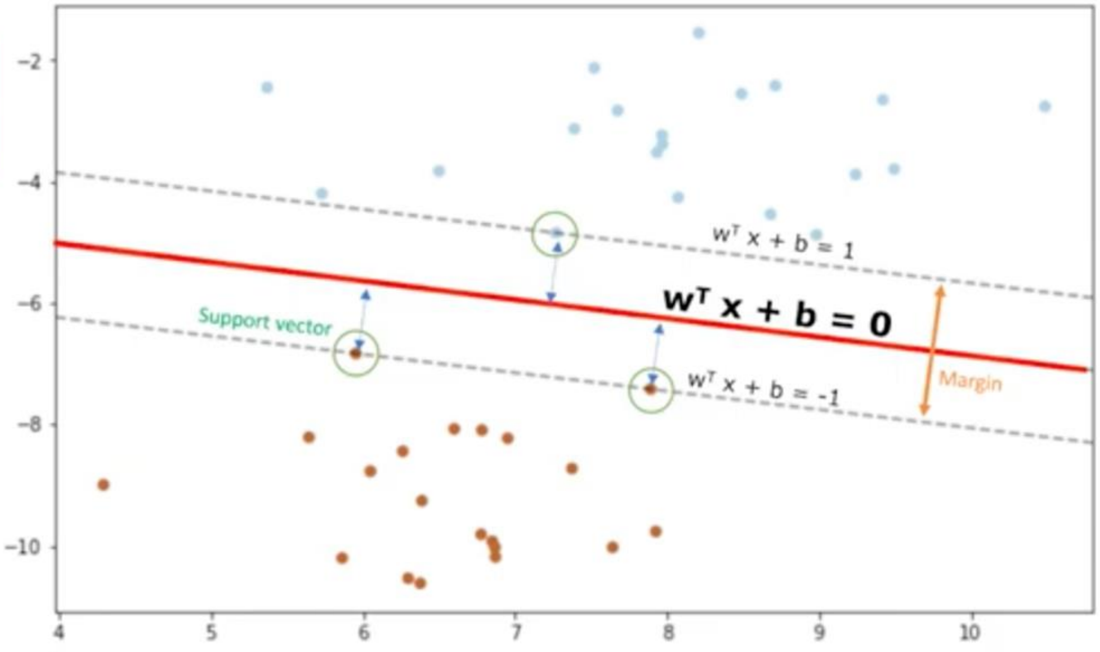
Support Vector Machines

- to find the hyperplane, we are looking for largest margins from support vectors

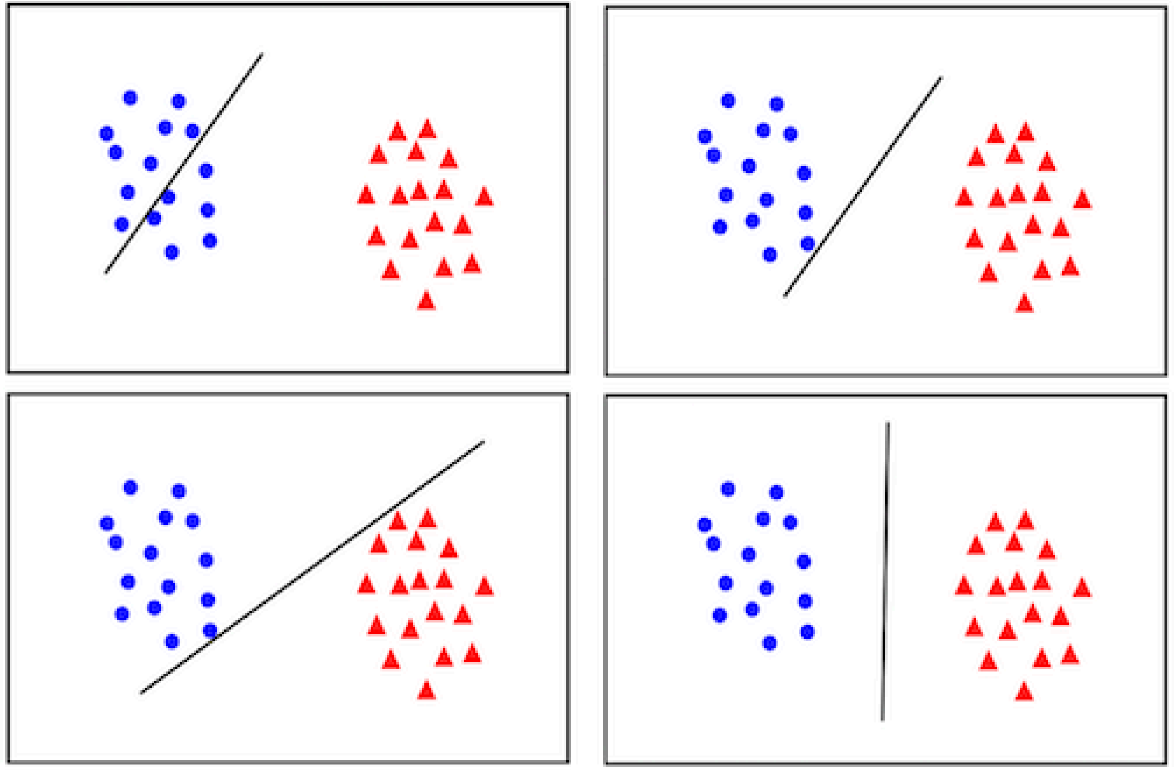
- can also be solved using gradient descent

- when learned, we can just check the data and see if i above the line or below it and decide

Find $\mathbf{w}$ and b such that
$\Phi(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w}$ is minimized;
and for all $\{(\mathbf{x}_i, y_i)\}$: $y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1$

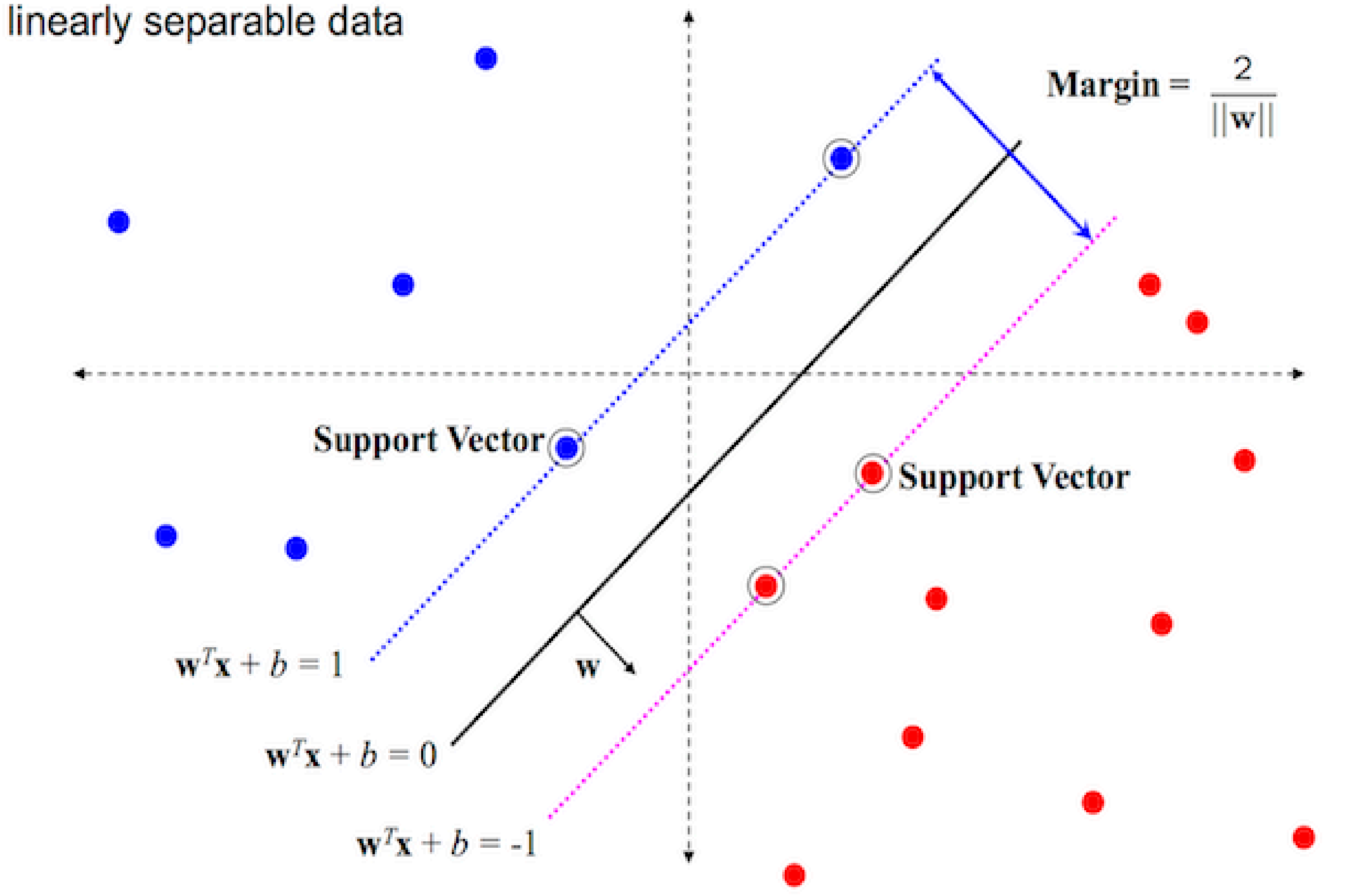


What is the best w?

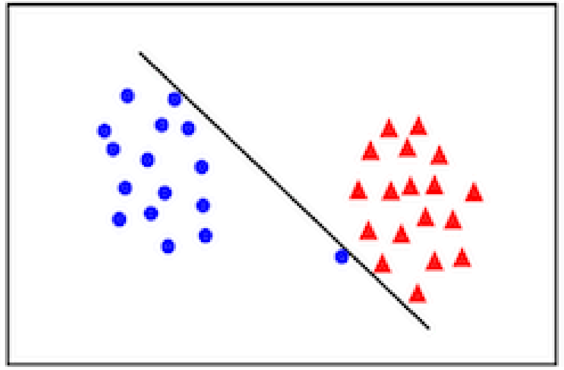


- maximum margin solution: most stable under perturbations of the inputs

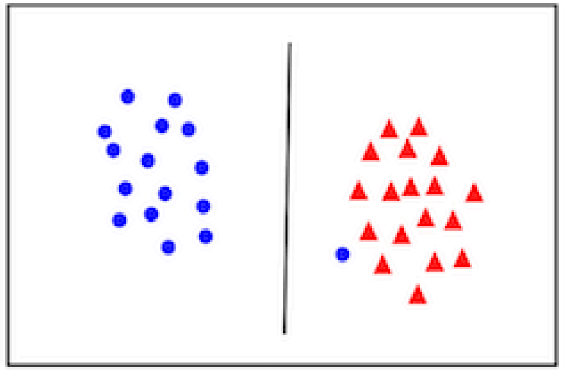
Support Vector Machine



Linear separability again: What is the best w?



• the points can be linearly separated but there is a very narrow margin



• but possibly the large margin solution is better, even though one constraint is violated

In general there is a trade off between the margin and the number of mistakes on the training data

Given training vectors $x_i \in \mathbb{R}^p$, i=1,..., n, in two classes, and a vector $y \in \{1, -1\}^n$, our goal is to find $w \in \mathbb{R}^p$ and $b \in \mathbb{R}$ such that the prediction given by $\text{sign}(w^T\phi(x) + b)$ is correct for most samples.

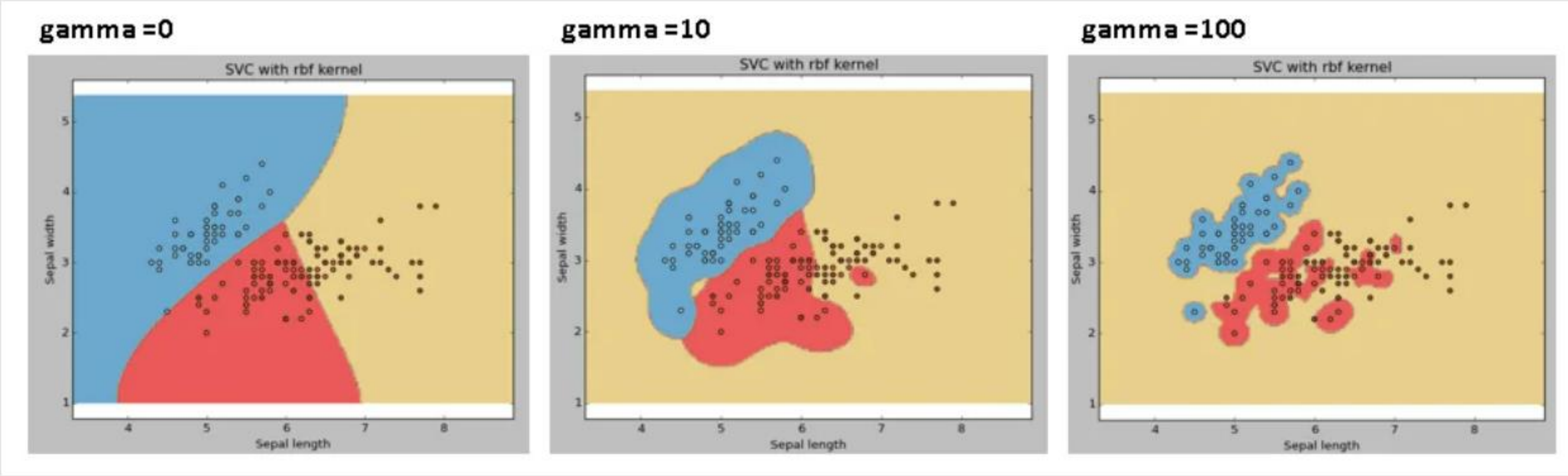
SVC solves the following primal problem:

$$\begin{aligned}\min_{w,b,\zeta} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \zeta_i \\ \text{subject to } y_i(w^T\phi(x_i) + b) \geq 1 - \zeta_i, \\ \zeta_i \geq 0, i = 1, \ldots, n\end{aligned}$$

Intuitively, we're trying to maximize the margin (by minimizing $||w||^2 = w^T w$), while incurring a penalty when a sample is misclassified or within the margin boundary. Ideally, the value $y_i(w^T\phi(x_i) + b)$ would be ≥ 1 for all samples, which indicates a perfect prediction. But problems are usually not always perfectly separable with a hyperplane, so we allow some samples to be at a distance ζ_i from their correct margin boundary. The penalty term `C` controls the strength of this penalty, and as a result, acts as an inverse regularization parameter (see note below).

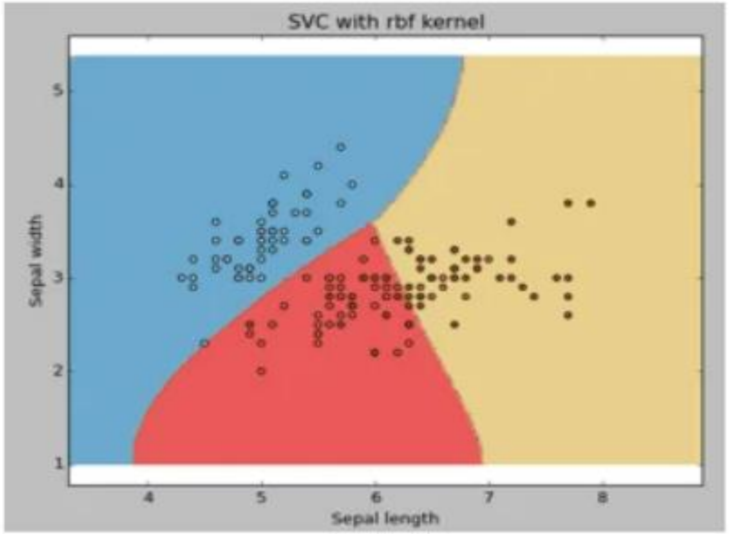
The dual problem to the primal is

$$\begin{aligned}\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - e^T \alpha \\ \text{subject to } y^T \alpha = 0 \\ 0 \leq \alpha_i \leq C, i = 1, \ldots, n\end{aligned}$$

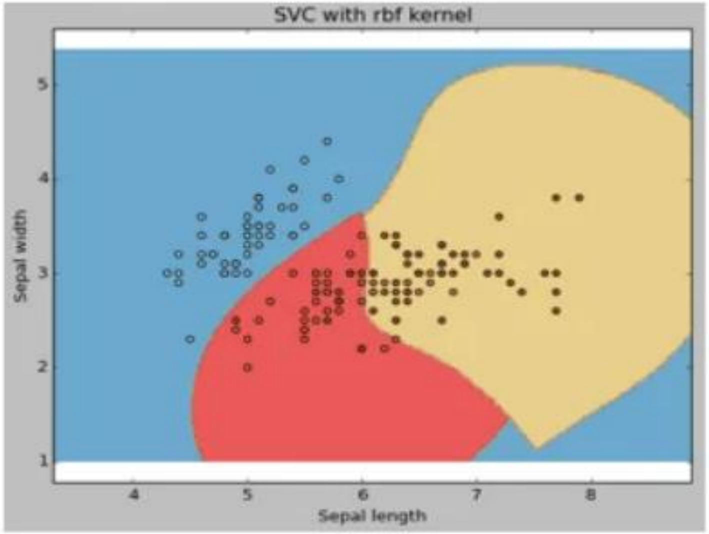


gamma =0
gamma =10
gamma =100
SVC with rbf kernel
Sepal width
Sepal length

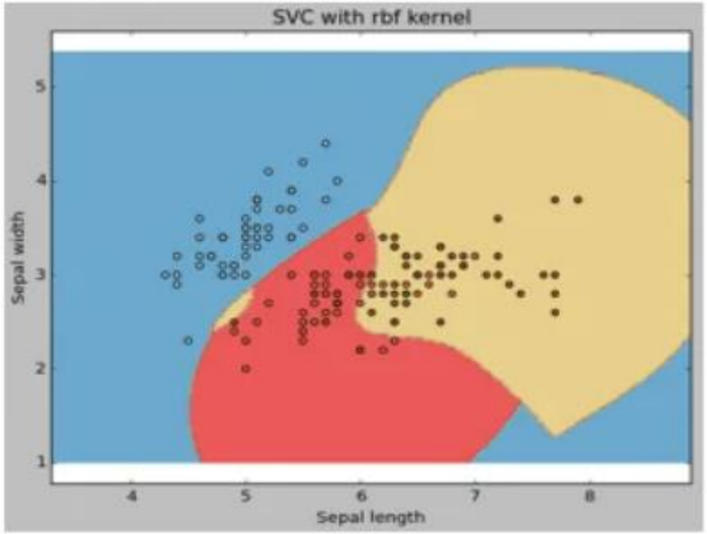
c =1



C =100



c =1000



Classification

Support Vector Machines

Pros

- accurate in high dimensional spaces
- memory efficient

Cons

- Prone to over-fitting if we have lots of features
- No probability estimation
- Not computationally efficient for large dataset (n>1000)

Classification

Support Vector Machines

Image recognition

Text Category Assignment

- spam
- category
- sentiment analysis

Gene Expression Classification

Outlier detection and clustering

Lab: SVM

- L07-MehrnazJalili