

CORRIERE DELLA SERA



UNIVERSITÀ
DI SIENA
1240



CINECA

DISCLAIMER

ULTIME AVVERTENZE PRIMA DELLA RIVOLUZIONE

BIOMIND

INTELLIGENZA ARTIFICIALE PER L'INNOVAZIONE
IN FARMACIA E BIOINFORMATICA

KNOWLEDGE PARTNER

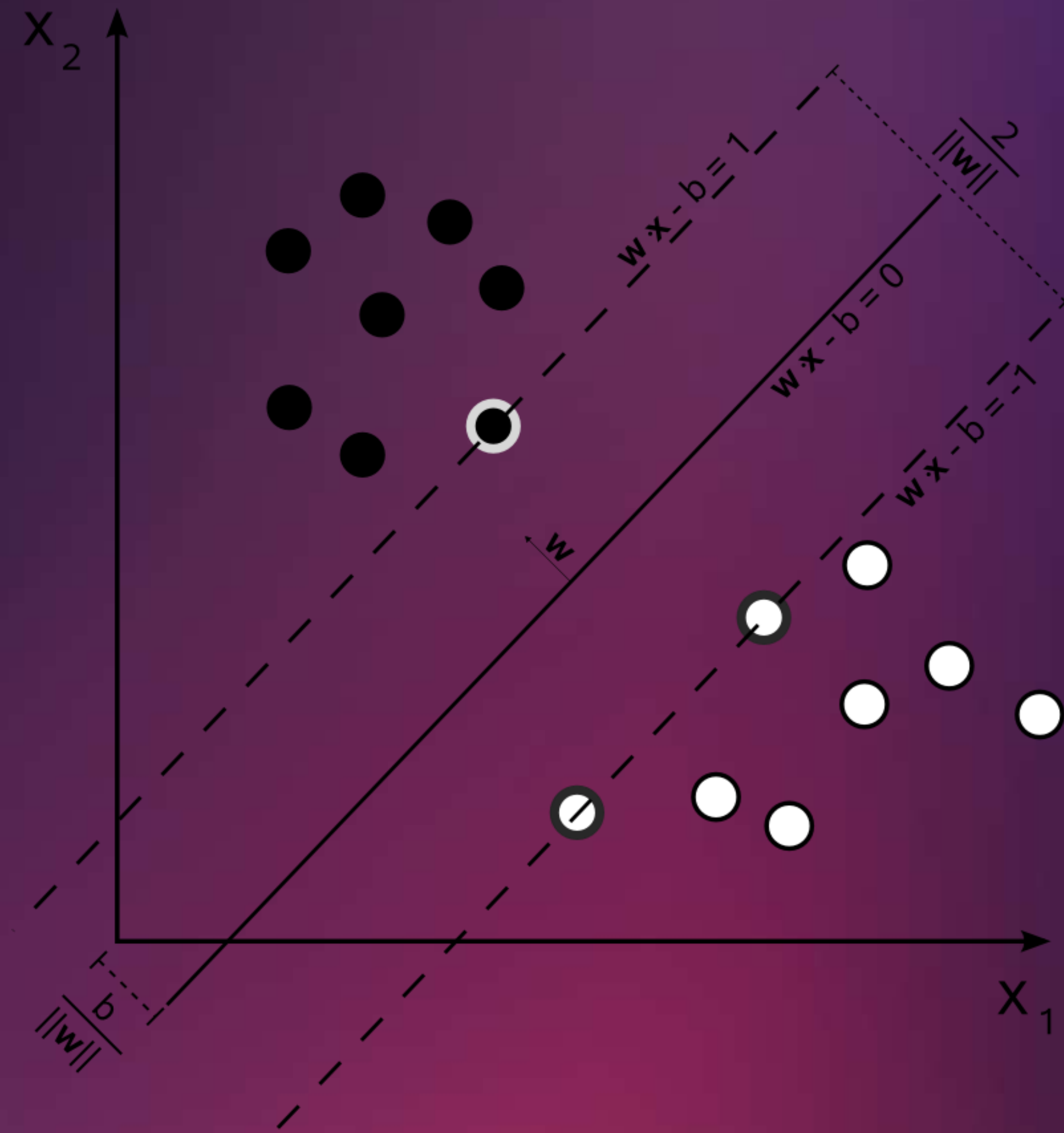


INTRODUZIONE

- Cos'è l'algoritmo
 - Le Support Vector Machines (SVM) sono un potente algoritmo di apprendimento supervisionato utilizzato principalmente per compiti di classificazione, ma applicabile anche alla regressione. L'idea centrale è trovare un iperpiano che separi i dati di classi diverse massimizzando il margine tra esse.
- A cosa serve
 - Un modello SVM cerca l'iperpiano ottimale che:
 - Massimizza la distanza (margine) tra i punti dati più vicini delle diverse classi, detti support vectors.
 - Riduce il rischio di errore di classificazione su nuovi dati.
- Contesto applicativo
 - Le SVM vengono ampiamente utilizzate in applicazioni pratiche come il riconoscimento di immagini, la classificazione di testi e la bioinformatica.

CONCETTI CHIAVE

- Struttura/architettura
- Elementi principali
 - training data
 - hyperplane
 - margin
 - support vectors
 - objective function
 - kernel trick



HOW DOES IT WORK?

● TRAINING PROCESS

- We dropped the columns (binding_affinity, active) from the original dataset and stored it in X.
- We then stored the target column(active) in Y.
- We split the dataset in 80:20 as train:test.
- Data Preprocessing : Min-Max Scaling

Our known data is split into training and validation sets. The former is used to teach the model while the latter is reserved for evaluating its performance on unseen labeled data

ADVANTAGES AND LIMITS

- ✓ Advantages:

- - Robustness against overfitting
- - Accuracy also with relatively small dataset

- ✗ Limits:

- - Requires an accurate choice of the kernel
- - less understandable compare to simpler linear models
- - real world data is not very linearly classified which makes the SVM hard margin use bit difficult

RESULT AND TUNING

We can modify the behavior of the model tweaking kernel and polynomial degree:

- linear
- RBF
- Sigmoid
- **Polynomial**

The best degree for the shape of our data is 3.

Kernel	Accuracy (%)	Polynomial degree
Linear	~90%	/
RBF	~85%	/
Sigmoid	~81%	/
Polynomial	~88%	degree 4
Polynomial	~91%	degree 3

CORRIERE DELLA SERA



UNIVERSITÀ
DI SIENA
1240



CINECA

GRAZIE DELL'ATTENZIONE!

DISCLAIMER

ULTIME AVVERTENZE PRIMA DELLA RIVOLUZIONE

BIOMIND

INTELLIGENZA ARTIFICIALE PER L'INNOVAZIONE
IN FARMACIA E BIOINFORMATICA

KNOWLEDGE PARTNER

