

[AIMLBD] MACHINE LEARNING, BIG DATA, ARTIFICIAL INTELLIGENCE per medicina e chirurgia high tech

L05: Classification

Dott. Giorgio De Magistris

demagistris@diag.uniroma1.it

CORSO DI LAUREA IN MEDICINA E CHIRURGIA HIGH TECH



SAPIENZA
UNIVERSITÀ DI ROMA

I3S

FACOLTÀ DI INGEGNERIA DELL'INFORMAZIONE, INFORMATICA E STATISTICA

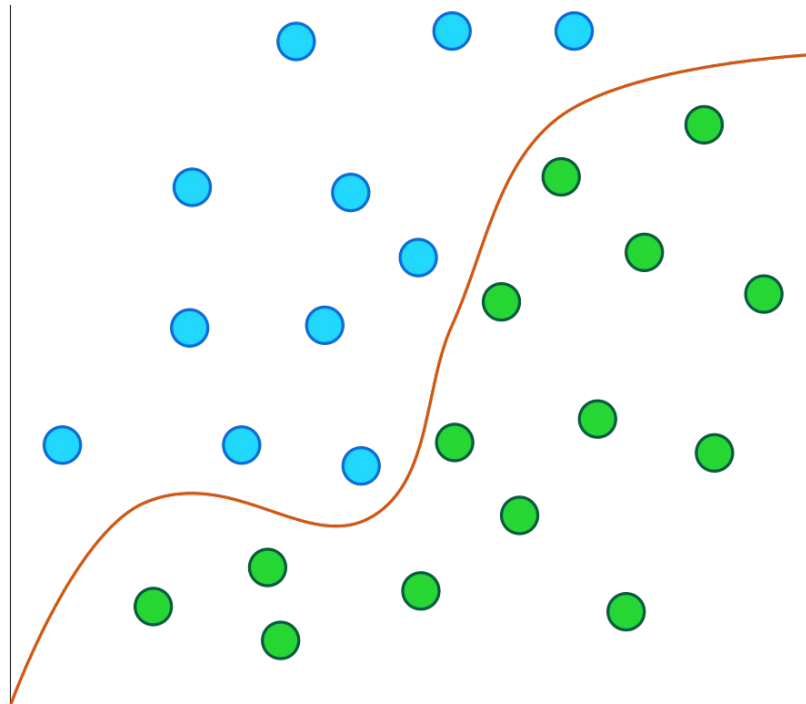
DIAG

DIPARTIMENTO DI INGEGNERIA INFORMATICA, AUTOMATICA E GESTIONALE

TUTTI I DIRITTI RELATIVI AL PRESENTE MATERIALE DIDATTICO ED AL SUO CONTENUTO SONO RISERVATI A SAPIENZA E AI SUOI AUTORI (O DOCENTI CHE LO HANNO PRODOTTO).
È CONSENTITO L'USO PERSONALE DELLO STESSO DA PARTE DELLO STUDENTE A FINI DI STUDIO. NE È VIETATA NEL MODO PIÙ ASSOLUTO LA DIFFUSIONE, DUPLICAZIONE,
CESSIONE, TRASMISSIONE, DISTRIBUZIONE A TERZI O AL PUBBLICO PENA LE SANZIONI APPLICABILI PER LEGGE

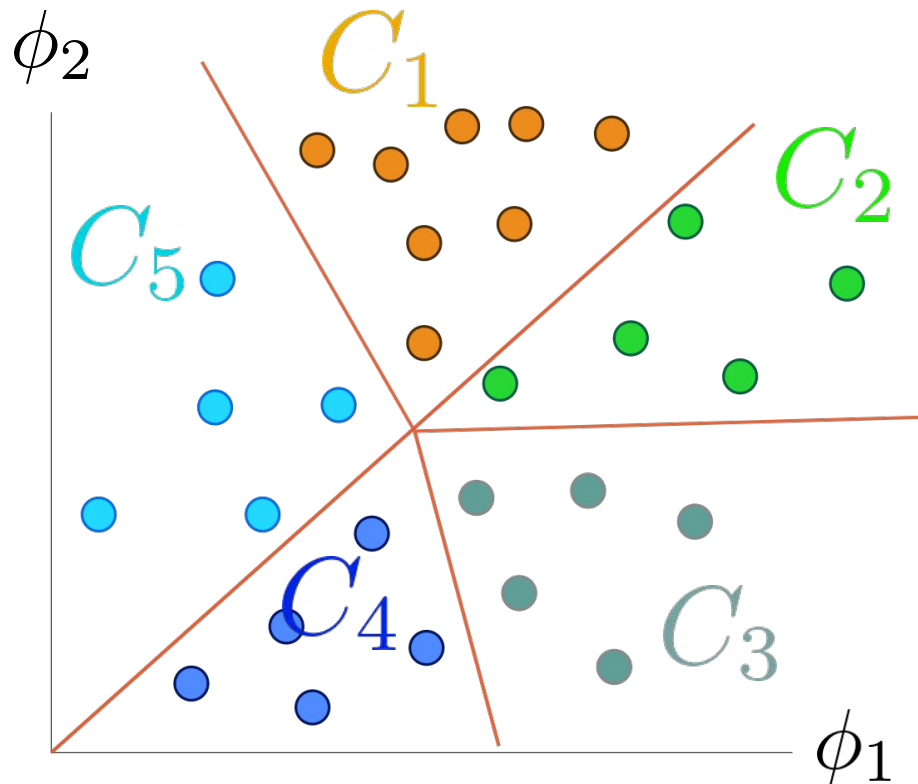
Classification

- Given some data points $X = \{x_1, \dots, x_n\}$
- Assign to each data point a label C_i from a set of labels $C = \{C_1, \dots, C_k\}$



Decision Boundaries

- Assume that data points are represented by vectors in an m-dimensional feature space
- The classification task consists in partitioning the feature space into k partitions (where k is the number of categories)



Linear models for classification

- A classification model is linear if the decision boundary is a linear function in the feature space, i.e. a line in 2D or a $m-1$ dimensional hyperplane in an m dimensional feature space
- Note that being linear in the feature space does not imply that the model is linear with respect to the original data space
- We could apply a nonlinear transformation of the input ϕ that transforms the data space into a feature space
- For example we have seen the polynomial expansion of the input, or we can use other transformations like Radial Basis Functions
- In general we will write a data sample as its expansion into the feature space

$$\phi_{\mathbf{i}} = \left(\phi_0(\mathbf{x}_{\mathbf{i}}) \quad \phi_1(\mathbf{x}_{\mathbf{i}}) \quad \dots \quad \phi_m(\mathbf{x}_{\mathbf{i}}) \right)^T$$

- If no transformation is applied then

$$\phi_{\mathbf{i}} = \left(\phi_0(\mathbf{x}_{\mathbf{i}}) = 1 \quad \phi_1(\mathbf{x}_{\mathbf{i}}) = x_1 \quad \dots \quad \phi_m(\mathbf{x}_{\mathbf{i}}) = x_m \right)^T = \mathbf{x}_{\mathbf{i}}$$

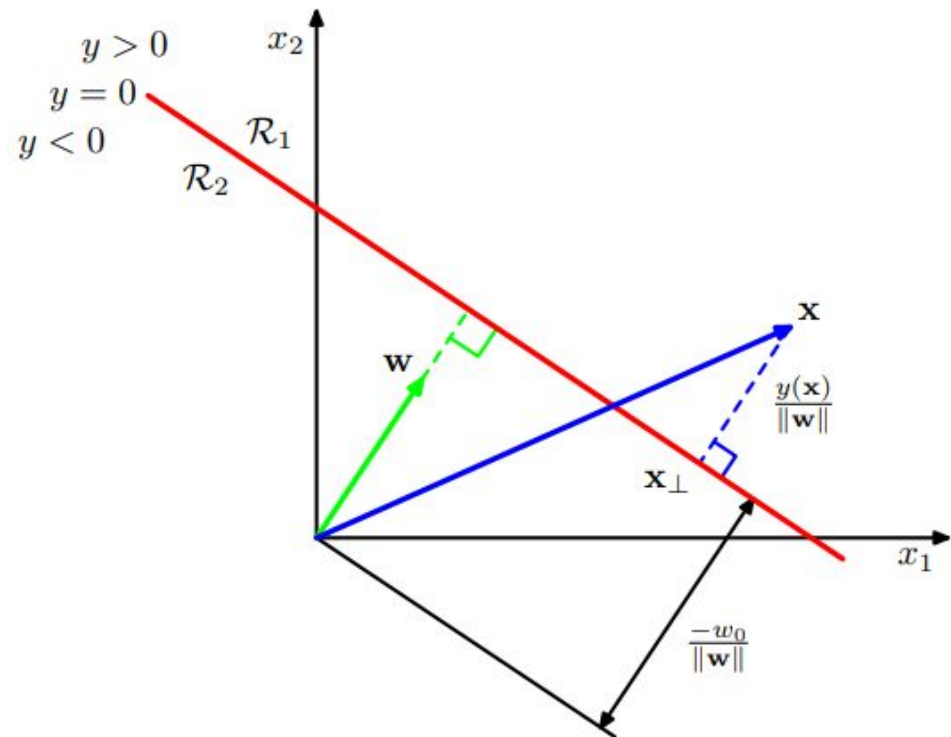
Binary Classification

- The decision boundary is a linear function of the features

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0 \quad \text{or} \quad y = \mathbf{w}^T \phi$$

- \mathbf{x} is assigned to C_1 if $y(\mathbf{x}) \geq 0$ and to class C_2 otherwise

- This is the geometrical interpretation for a 2D feature space



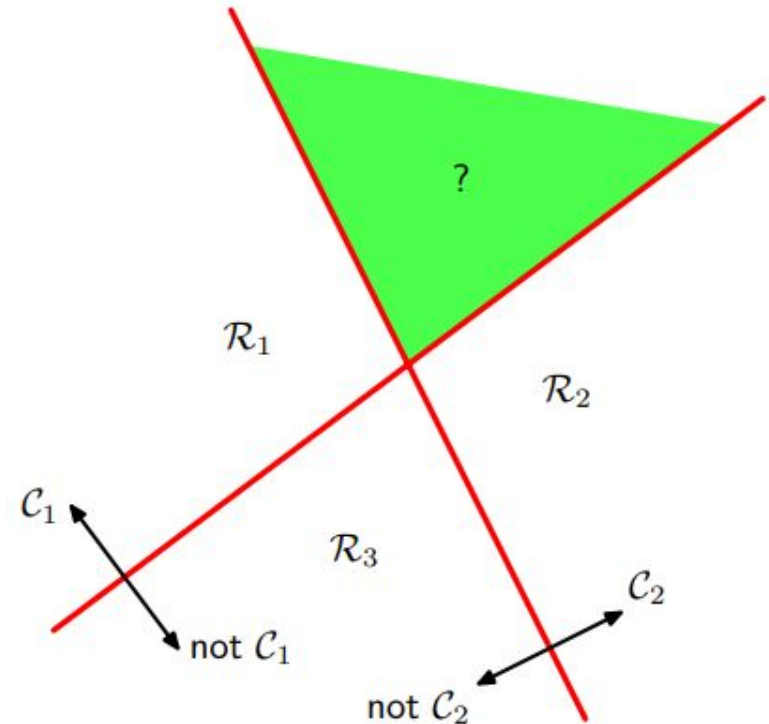
[Image Credit: Pattern Recognition and Machine Learning](#)

Multiclass Classification

- I can handle the case of multiple classes using different approaches:
 - *one-versus-the-rest*: for each class I train a classifier to separate points in that class from those that are not in that class
 - *one-versus-one*: for each pair of classes i train a classifier to distinguish the two classes and then apply a majority voting scheme
 - *k-class discriminant*: training k linear function and assign the point to the class that maximizes $y_k(x)$

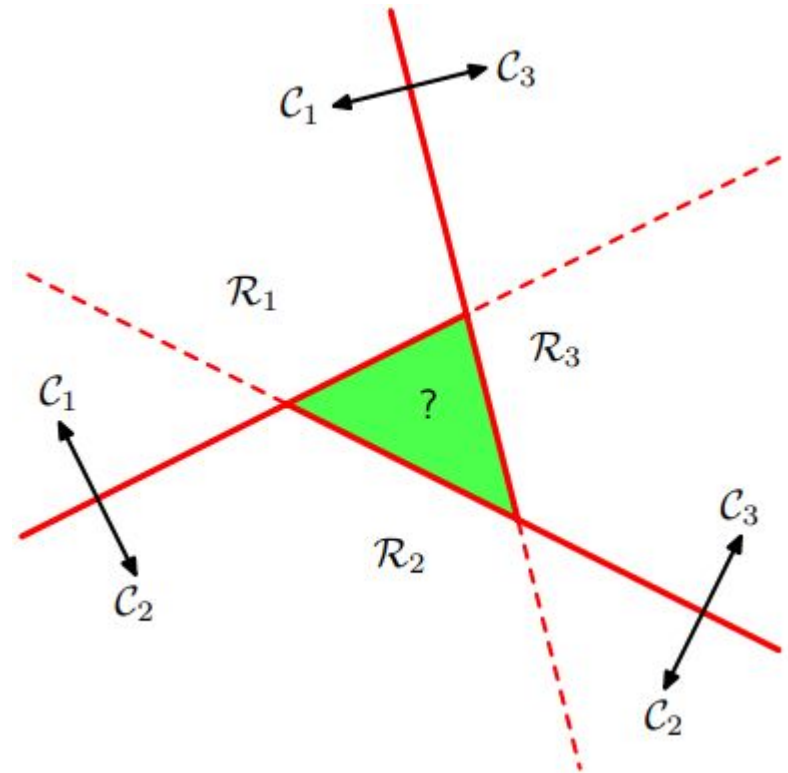
one-versus-the-rest

- Train k classifiers, one for each class, to classify samples belonging to that class versus samples not belonging to that class
- The classifier
$$y_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + w_{i0}$$
classifies class C_i versus all other classes
- This approach leads to ambiguous regions, as shown in the figure
- The training is negatively affected by the classes imbalance



one-versus-one

- Train $k(k-1)/2$ classifiers, one for each pair of classes
- the classifier
$$y_{ij}(\mathbf{x}) = \mathbf{w}_{ij}^T \mathbf{x} + w_{ij0}$$
classifies class C_i versus class C_j
- Again some regions of the input space are not labelled correctly, as shown in the figure
- High cost, I've to train a huge number of classifiers



k class discriminant

- A K-class discriminant comprises k linear functions of the form

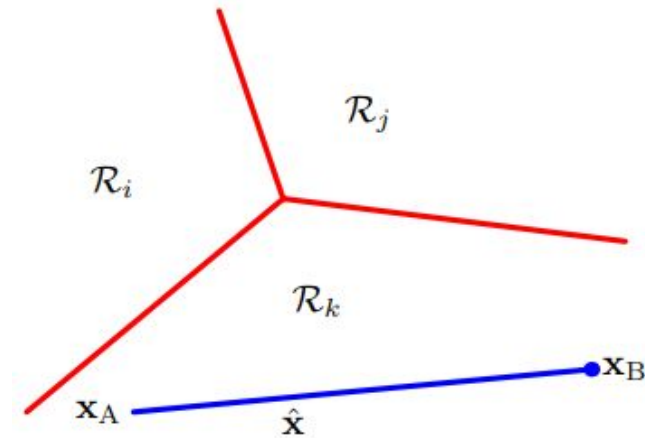
$$y_k(\mathbf{x}) = \mathbf{w}_k^T \mathbf{x} + w_{k0}$$

- a point \mathbf{x} is assigned to class C_k if $y_k(\mathbf{x}) > y_j(\mathbf{x})$ for all $j \neq k$
- the decision boundary between classes C_k and C_j is therefore given by the equation $y_k(\mathbf{x}) = y_j(\mathbf{x})$
- the decision boundary corresponds to a D-1 dimensional hyperplane with equation:

$$(\mathbf{w}_k - \mathbf{w}_j)^T \mathbf{x} + w_{k0} - w_{j0} = 0$$

k class discriminant

- With this approach there are no ambiguous regions in the input space
- To show this, consider two points \mathbf{x}_A and \mathbf{x}_B both inside region \mathcal{R}_k
- Then consider the line $\hat{\mathbf{x}}$ connecting them:
$$\hat{\mathbf{x}} = \lambda \mathbf{x}_A + (1 - \lambda) \mathbf{x}_B \quad \text{with } \lambda \in [0, 1]$$
- for the linearity of the discriminant functions we get
$$y_k(\hat{\mathbf{x}}) = \lambda y_k(\mathbf{x}_A) + (1 - \lambda) y_k(\mathbf{x}_B)$$
- now, knowing that $y_k(\mathbf{x}_A) > y_j(\mathbf{x}_A)$ and $y_k(\mathbf{x}_B) > y_j(\mathbf{x}_B) \quad \forall j \neq k$
we can conclude that $y_k(\hat{\mathbf{x}}) > y_j(\hat{\mathbf{x}}) \quad \forall j \neq k$
and so every point in $\hat{\mathbf{x}}$ also lies in \mathcal{R}_k



Slides distribuite con Licenza Creative Commons (CC BY-NC-ND 4.0) Attribuzione - Non commerciale - Non opere derivate 4.0 Internazionale

PUOI CONDIVIDERLE ALLE SEGUENTI CONDIZIONI

(riprodurre, distribuire, comunicare o esporre in pubblico, rappresentare, eseguire e recitare questo materiale con qualsiasi mezzo e formato)

Attribuzione*

Devi riconoscere una menzione di paternità adeguata, fornire un link alla licenza e indicare se sono state effettuate delle modifiche. Puoi fare ciò in qualsiasi maniera ragionevole possibile, ma non con modalità tali da suggerire che il licenziante avalli te o il tuo utilizzo del materiale.

Non Commerciale

Non puoi utilizzare il materiale per scopi commerciali.

Non opere derivate

Se remixi, trasformi il materiale o ti basi su di esso, non puoi distribuire il materiale così modificato.

Divieto di restrizioni aggiuntive

Non puoi applicare termini legali o misure tecnologiche che impongano ad altri soggetti dei vincoli giuridici a questa licenza