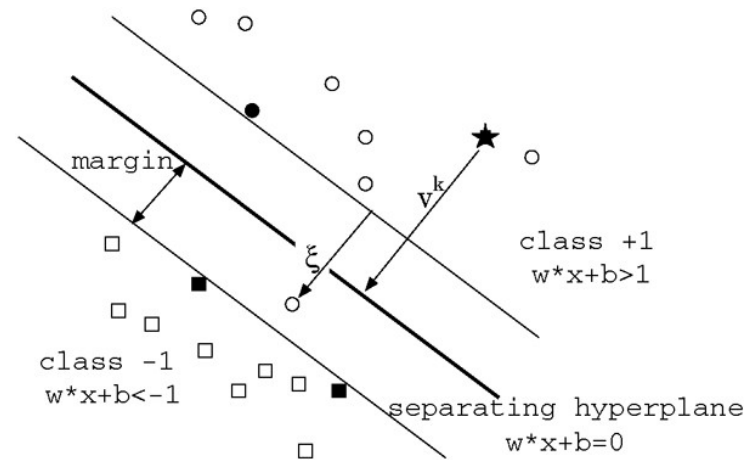
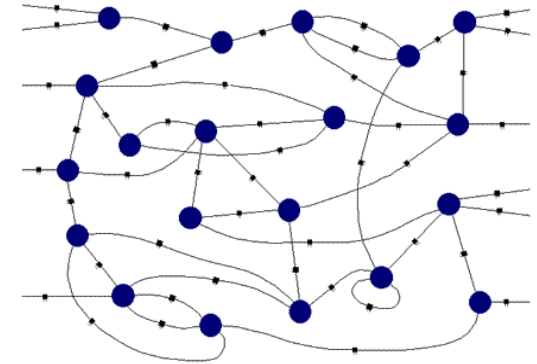
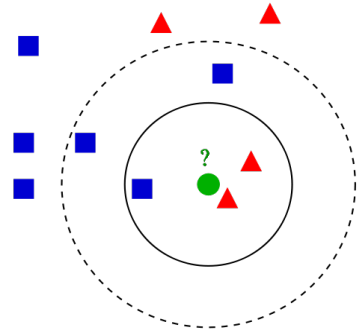


Non solo alberi di decisione ...

Alternative

- Classificatori a regole
- K-nearest neighbour
- Classificatori Bayesiani
- Reti neurali (esistono moltissimi modelli)
- Support Vector Machines
- Ensemble methods
 - Bagging
 - boosting
- Regressione



Thomas Bayes

Usare le regole per classificare

Si fa presto a dir “regole” ...

Esempi

If-then:

If (colore == arancione && forma == sfera) then arancia

ECA (event-condition-action):

Se (cade_bicchiere) qualora pieno(bicchiere) esegui pulisci

Clausole logiche:

bottiglia(X) :- contenitore(Y), tappo(Z), partOf(Y, X), partOf(Z, X),
avvitabile(Z, Y).

Grammatiche:

$E ::= E \text{ and } E \mid E \text{ or } E \mid (E) \mid \text{DNF}$

Funzioni:

$Y = X + 2$

... ecc ...

Si fa presto a dir “regole” ...

Fuoco:

If-then:

If (colore == arancione && forma == sfera) **then** *arancia*

ECA (event-condition-action):

Se (cade_bicchiere) qualora pieno(bicchiere) esegui pulisci

Clausole logiche:

bottiglia(X) :- contenitore(Y), tappo(Z), partOf(Y, X), partOf(Z, X),
avvitabile(Z, Y).

Grammatiche:

$E ::= E \text{ and } E \mid E \text{ or } E \mid (E) \mid \text{DNF}$

Funzioni:

$Y = X + 2$

... ecc ...

Antecedenti e conseguenti

Antecedente:

$$(att_1 \text{ op } val_1) \wedge (att_2 \text{ op } val_2) \wedge \dots \wedge (att_k \text{ op } val_k)$$
$$op \in \{ =, \neq, <, >, \leq, \geq \}$$

Consequente:

Classe predetta

Una regola si attiva quando la descrizione di un'istanza soddisfa la condizione posta ad antecedente

Qualità di una regola

Copertura

$$\frac{|A|}{|D|}$$

$|D|$ = cardinalità del dataset

$|A|$ = numero di istanze che soddisfano l'antecedente della regola

$|A \cap y|$ = numero delle istanze che soddisfano A e che sono di classe y

Numero di istanze
coperte dalla regola

Accuratezza

$$\frac{|A \cap y|}{|A|}$$

Numero di istanze
correttamente classificate
dalla regola

Esempio

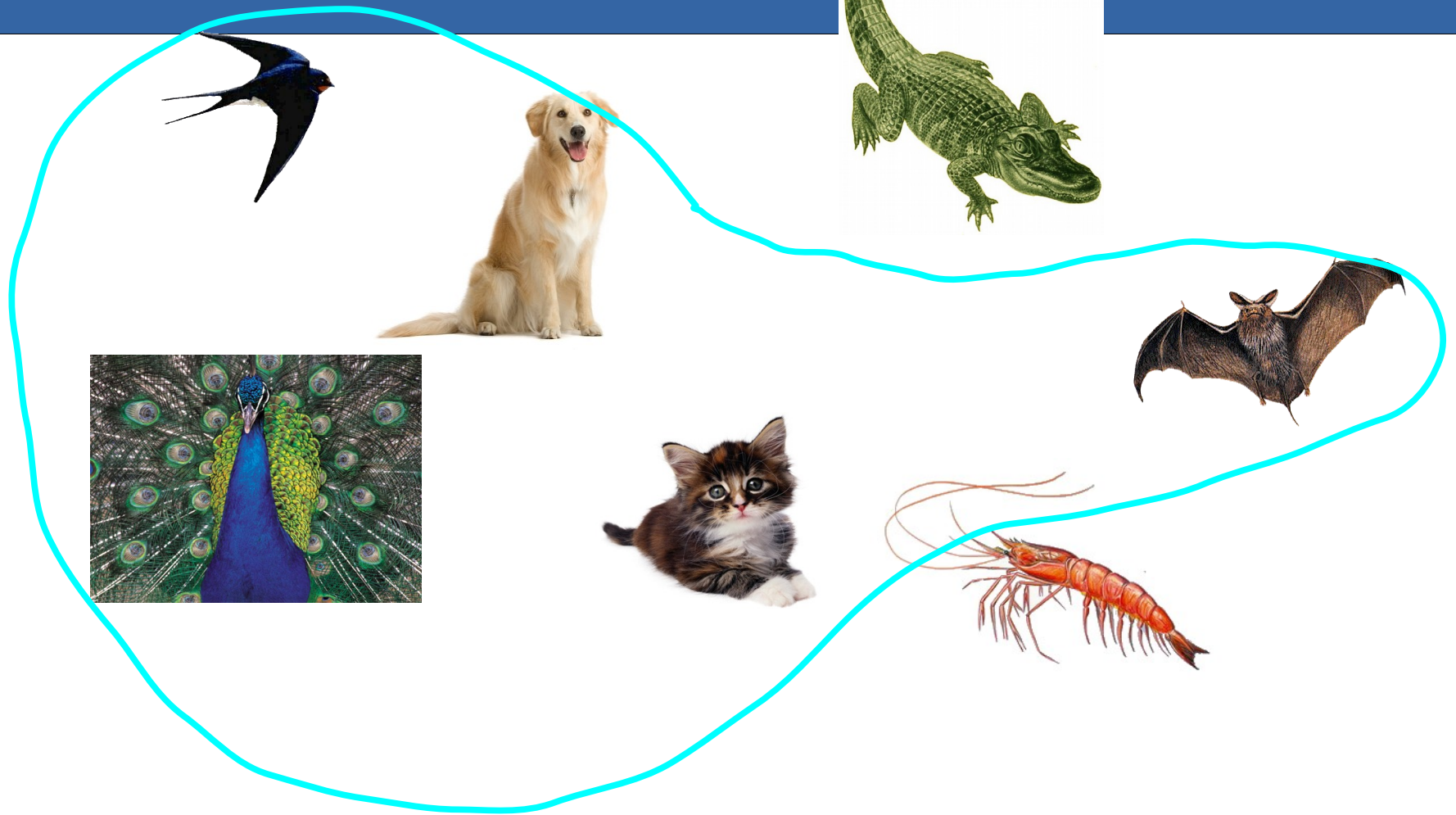


If (sangue = caldo and vola = no) then mammifero

Copertura = 2/7

Accuratezza = 2/2

Esempio



If (sangue = caldo) then mammifero

Copertura = 5/7

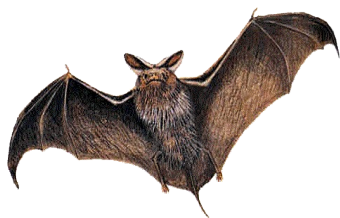
Accuratezza = 3/5

Proprietà degli insiemi di regole

- **Regole mutuamente esclusive:** ogni regola è attivata da un insieme di esempi disgiunto dagli insiemi di esempi che attivano le altre regole;
- **Regole esaustive:** le regole sono esaustive se ogni possibile combinazione di valori degli attributi è catturata (tutte le istanze saranno catturate da qualche regola);
- Regole esaustive e mutuamente esclusive garantiscono che ogni istanza sia catturata da una sola regola;

Proprietà degli insiemi di regole

- Regole mutuamente esclusive: ogni regola è attivata da un insieme di esempi disgiunto dagli insiemi di esempi che attivano le altre regole;
- Regole esaustive: le regole sono esaustive se ogni possibile combinazione di valori degli attributi è catturata (tutte le istanze saranno catturate da qualche regola);
- Regole esaustive e mutuamente esclusive garantiscono che ogni istanza sia catturata da una sola regola;
- Il non avere la mutua esclusione implica che un'istanza possa far scattare tante regole, che possono anche predire classi diverse!!



If (sangue = caldo) then mammifero

If (vola) then uccello

Proprietà degli insiemi di regole

- Regole mutuamente esclusive: ogni regola è attivata da un insieme di esempi disgiunto dagli insiemi di esempi che attivano le altre regole;
- Regole esaustive: le regole sono esaustive se ogni possibile combinazione di valori degli attributi è catturata (tutte le istanze saranno catturate da qualche regola);
- Regole esaustive e mutuamente esclusive garantiscono che ogni istanza sia catturata da una sola regola;
- Il non avere la mutua esclusione implica che un'istanza possa far scattare tante regole, che possono anche predire classi diverse!!
- Il non avere l'esaustività implica che alcuni casi non saranno classificabili. In questa circostanza si può introdurre una regola di default, che specifica l'assunzione da fare in mancanza di altre ipotesi:

if (true) then classe_default

Se le regole non sono mutuamente esclusive ...

- **Liste di decisione:**

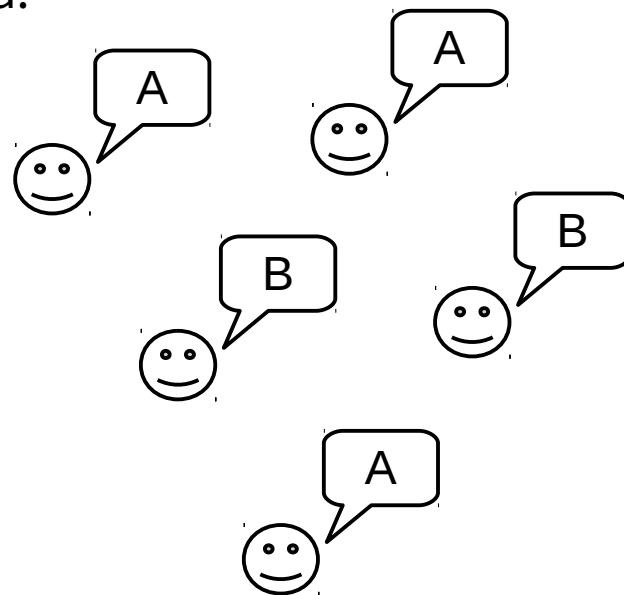
le regole vengono ordinate in ordine di priorità decrescente;

un'istanza sarà classificata secondo l'ipotesi formulata dalla regola che ha priorità più alta fra quelle che si attivano.

1. if (cond1) then ipotesi 1
2. if (cond2) then ipotesi 2
- **3. if (cond3) then ipotesi 3**
4. if (cond4) then ipotesi 4
- 5. if (cond5) then ipotesi 5**
- ...

- **Insiemi non ordinati:**

se un'istanza fa scattare tante regole diverse, che producono ipotesi diverse, si esegue un conteggio e si attribuisce all'istanza la classe più votata.



Ordinamento delle regole



- **Basato su regole:**

l'ordinamento dipenda da una qualche *misura della qualità* delle regole stesse

l'idea è che ogni istanza venga sempre coperta dalla *regola “migliore”* fra quelle che si attivano

nota: una regola di rank basso viene considerata solo se tutte quelle superiori non si attivano (se non è questo caso e neppure questo e neanche questo ...)

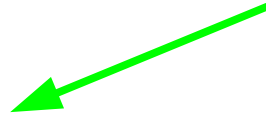
- **Basato su classi:**

le regole sono raggruppate *classe per classe*

le classi sono ordinate

classificazione: all'istanza viene attribuita la classe di rank più elevato, una delle cui regole si attiva

Ordinamento delle regole: esempi



- **Basato su regole:**

1. piumaggio and creatura_aerea -> uccello
2. sangue caldo and viviparo -> mammifero
3. sangue caldo and !viviparo -> uccello
4. semi-acquatico -> anfibio
5. scaglie and !acquatico -> rettile
6. scaglie and acquatico -> pesce

Ordinamento delle regole: esempi



- **Basato su regole:**

1. piumaggio and creatura_aerea -> uccello
2. sangue caldo and viviparo -> mammifero
3. sangue caldo and !viviparo -> uccello
4. semi-acquatico -> anfibio
5. scaglie and !acquatico -> rettile
6. scaglie and acquatico -> pesce

- **Basato su classi:**

- A. piumaggio and creatura_aerea -> uccello

A. sangue caldo and !viviparo -> uccello
- B. sangue caldo and viviparo -> mammifero
- C. semi-acquatico -> anfibio
- D. scaglie and !acquatico -> rettile
- E. scaglie and acquatico -> pesce

Produzione delle regole

- **Metodi diretti:**

estraggono le regole dai dati

Sequential covering

- Learn-one-rule function
 - General-to-specific
 - Specific-to-general

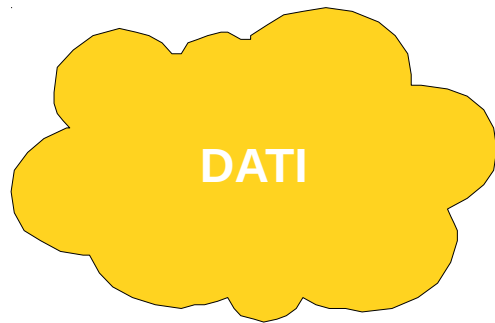
- **Metodi indiretti:**

estraggono le regole da altri classificatori (per esempio dagli alberi di decisione)

Problemi:

- Produzione delle regole
- Ordinamento delle regole

Metodi diretti: sequential covering

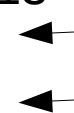


1. si decide un **ordinamento delle classi** con un qualche criterio:

C1
C2
C3
...

2. seguendo l'ordinamento ci si focalizza su **una classe alla volta** e si creano le regole di classificazione:

→ C1: r11, r12, r13
→ C2: r21, r22
→ C3: ...



Algoritmo

E = training set

$A = \{ (A_j, v_j) \}$

$Y_0 = \{ y_1, y_2, \dots, y_k \}$

$R = \{ \}$

coppie attributo-valore

insieme ordinato di classi

regole risultanti (insieme inizialmente vuoto)

Foreach $y \in Y_0 - \{ y_k \}$ **do**

// per ogni classe tranne l'ultima

While (!finito) **do**

// ripeti

$R \leftarrow \text{learn-one-rule}(E, A, y)$

// produci una regola

$E \leftarrow E - \{ \text{istanze coperte da } r \}$

// rimuovi le istanze classificate

$R \leftarrow R \vee r$

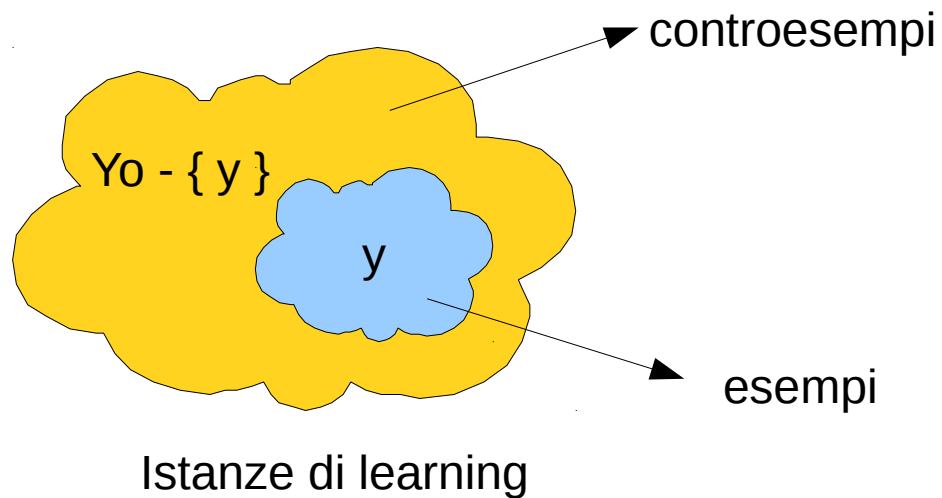
// aggiorna il risultato

End

End

Aggiungi la regola di default in ultima posizione

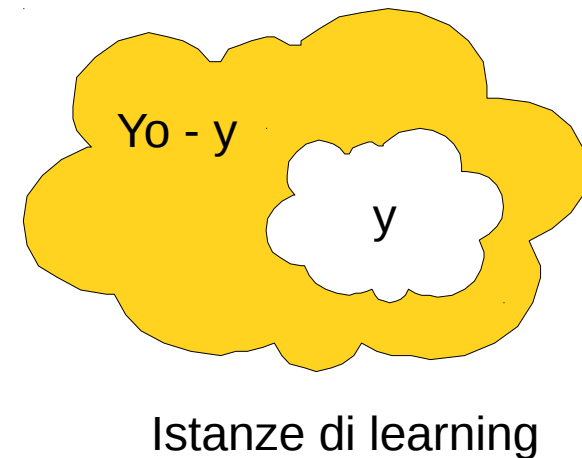
Iterazione relativa alla classe y



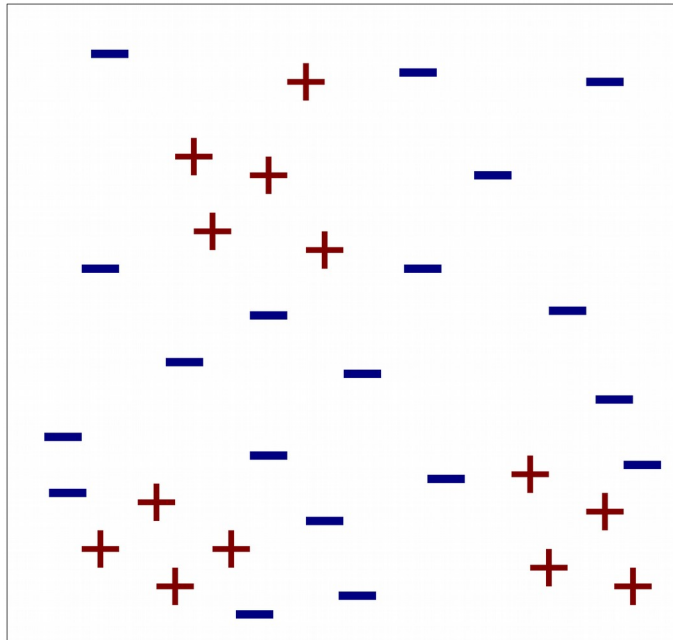
Da questa base si costruiscono le regole per classificare le istanze di classe y , usando tutte le altre come controesempi

Una volta terminato si rimuovono dal training set le istanze di y .

Alla prossima iterazione si lavorerà su un insieme ridotto di istanze



Più nello specifico ...



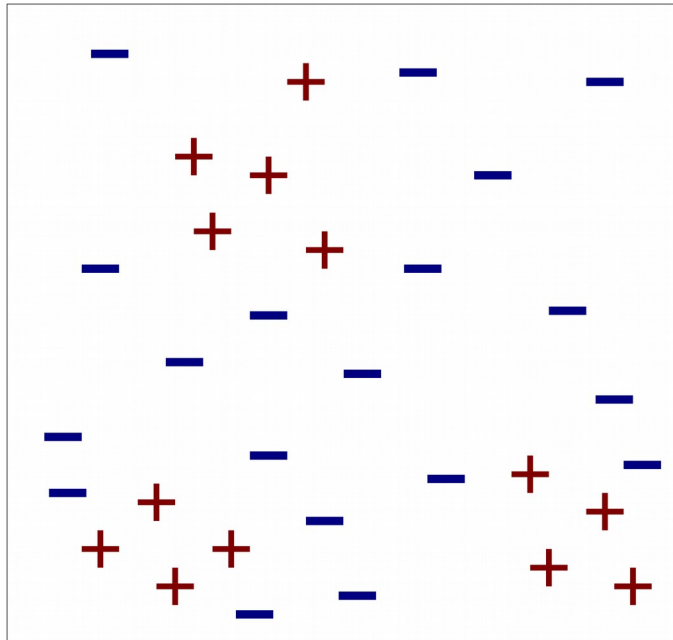
(i) Original Data

Dataset sulla base del quale vogliamo costruire le regole di classificazione

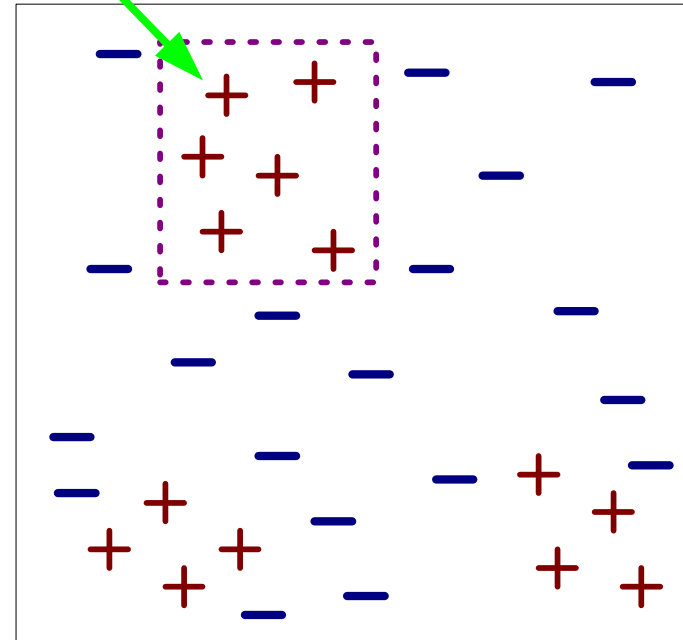
Sono rappresentate, in forma di punti sul piano, due classi: la classe da apprendere e le istanze delle altre classi genericamente viste come controesempi

Più nello specifico ...

Identifico un primo sottoinsieme di esempi catturabili da una regola



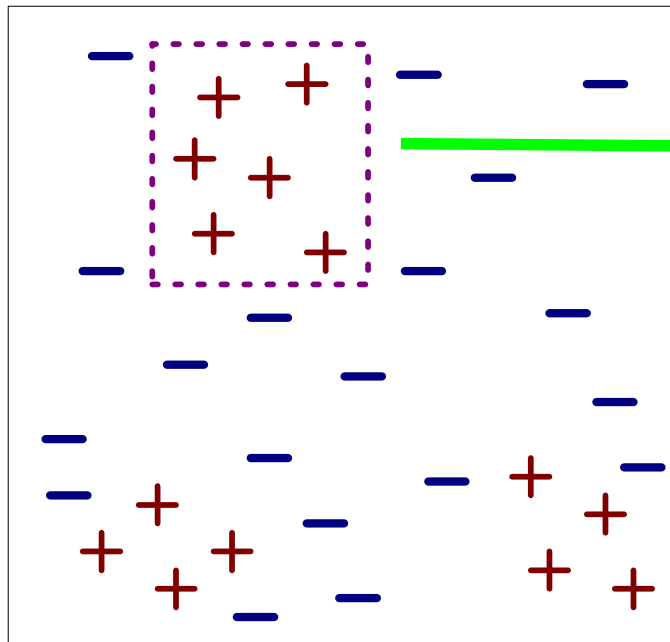
(i) Original Data



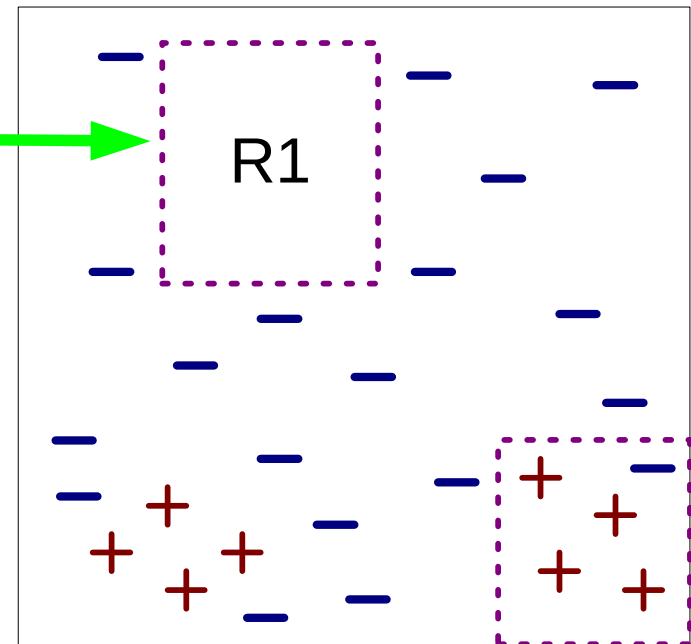
(ii) Step 1

Più nello specifico ...

Rimuovo gli esempi classificati correttamente



(ii) Step 1

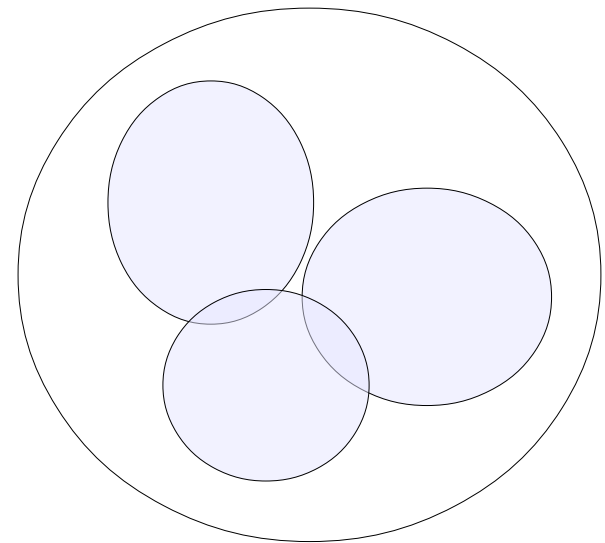
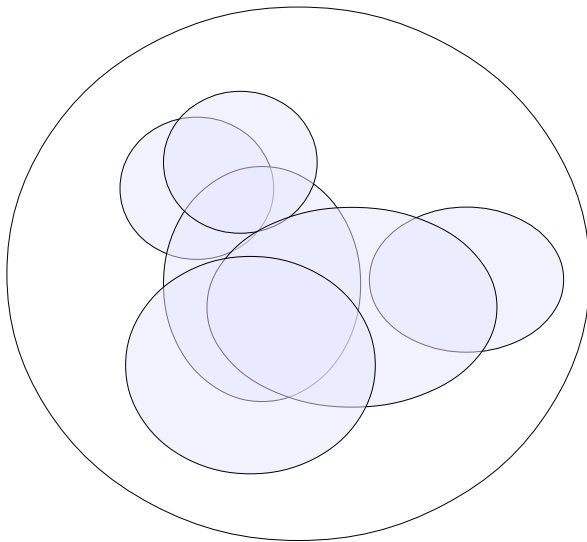


(iii) Step 2

Identifico un altro sottoinsieme di istanze della stessa classe catturabili da una regola

Perché rimuovere le istanze?

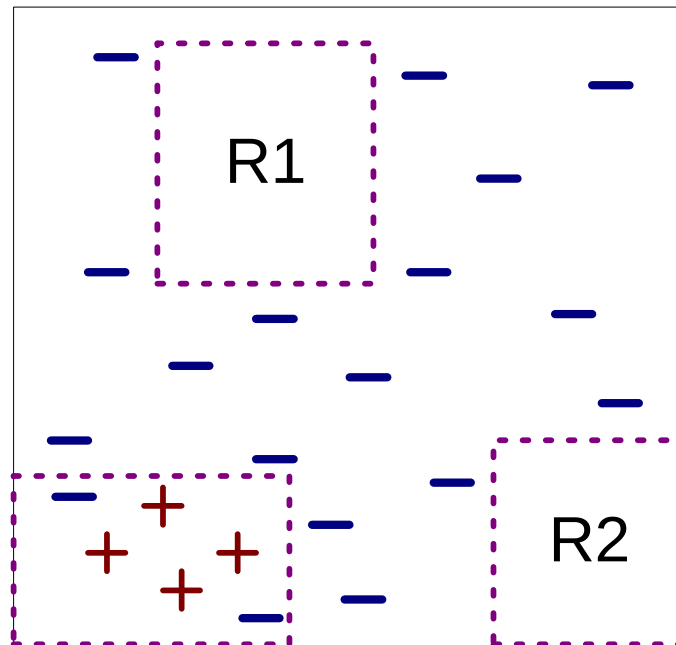
Meglio costruire regole con questo tipo di copertura ...



... oppure così?

Togliere gli esempi classificati da una regola aiuta a “distribuire” le regole generate in modo che coprano in maniera minimamente ridondante gli esempi e consentendo la copertura di aree in cui gli esempi sono meno densi

Più nello specifico ...



(iv) Step 3

Eccetera ...

Ricordate che le regole saranno ordinate nella stessa sequenza in cui sono state generate

R1
R2
R3

...

Learn one rule

È una funzione che permette di produrre una nuova regola per induzione da esempi.

Desiderata: la regola deve coprire il maggior numero di istanze positive e deve coprire il minor numero di controesempi possibile

Learn one rule

È una funzione che permette di produrre una nuova regola per induzione da esempi.

Desiderata: la regola deve coprire il maggior numero di istanze positive e deve coprire il minor numero di controesempi possibile

Meccanismo generale: produco una regola iniziale e poi la raffino per passi successivi

Strategie di raffinamento:

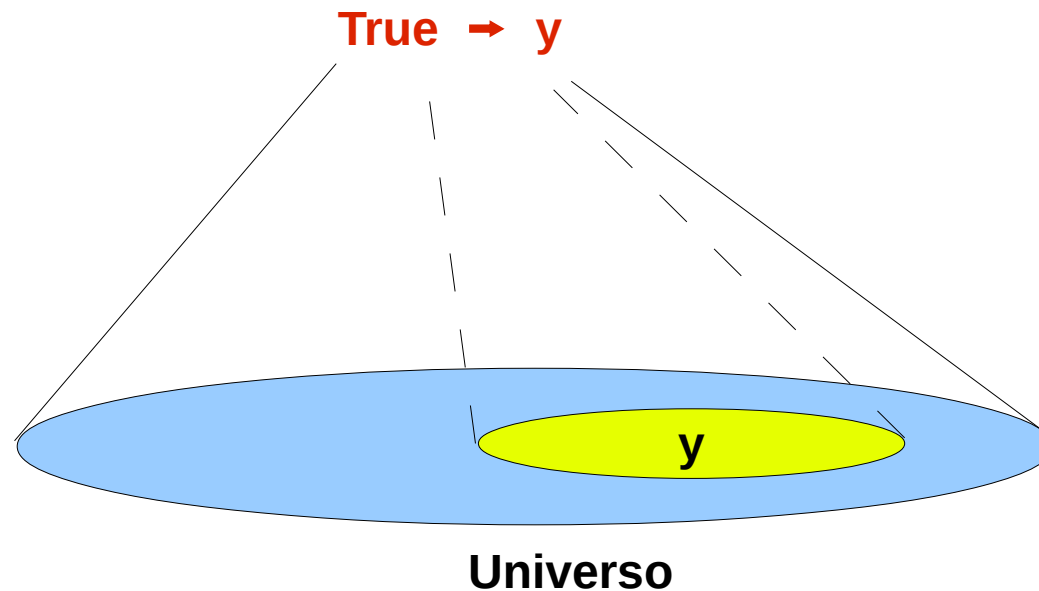
- (1) general to specific
- (2) specific to general

General-to-specific

Qual è **la regola più generale immaginabile** che riconosce sicuramente tutte le istanze di classe y ?

General-to-specific

Qual è la regola più generale immaginabile che riconosce sicuramente tutte le istanze di classe y ?



Tutto è di classe y

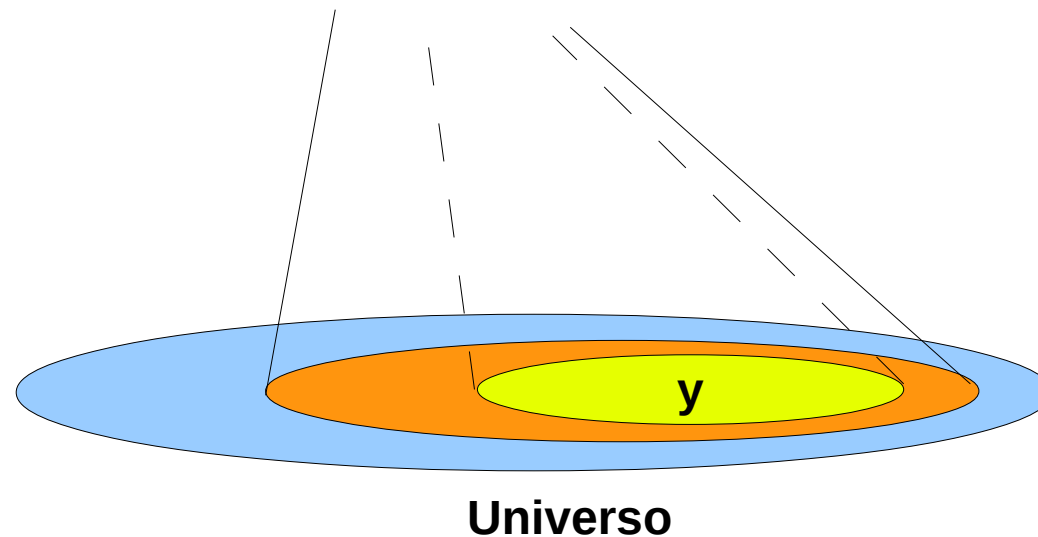
Problema: troppo generale
la qualità della regola è
bassa, non permette di di-
stinguere i non- y dagli y

General-to-specific

Specializzazione tramite aggiunta di congiunti nell'antecedente:

antecedente_{t+1} = antecedente_t and (attr op val)

True \wedge (att_i = val_i) \rightarrow y



Specific-to-general

Scelgo in modo casuale un esempio della classe da apprendere e definisco la regola:

$$(att_1 = val_{1e}) \wedge (att_2 = val_{2e}) \wedge \dots \wedge (att_k = val_{ke}) \rightarrow y$$

Congiunti: i valori corrispondono a quelli dell'esempio scelto

Numero dei congiunti: numero degli attributi con cui sono descritti gli esempi nel training set

Specific-to-general

Scelgo in modo casuale un esempio della classe da apprendere e definisco la regola:

$$(att_1 = val_{1e}) \wedge (att_2 = val_{2e}) \wedge \dots \wedge (att_k = val_{ke}) \rightarrow y$$

Congiunti: i valori corrispondono a quelli dell'esempio scelto

Numero dei congiunti: numero degli attributi con cui sono descritti gli esempi nel training set

Problema: troppo specifica!



If (viviparo) \wedge (sangue caldo) \wedge (volatile) \rightarrow mammifero

Specific-to-general

Scelgo in modo casuale un esempio della classe da apprendere e definisco la regola:

$$(att_1 = val_{1e}) \wedge (att_2 = val_{2e}) \wedge \dots \wedge (att_k = val_{ke}) \rightarrow y$$

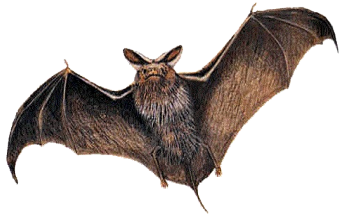
Congiunti: i valori corrispondono a quelli dell'esempio scelto

Numero dei congiunti: numero degli attributi con cui sono descritti gli esempi nel training set

Problema: troppo specifica!

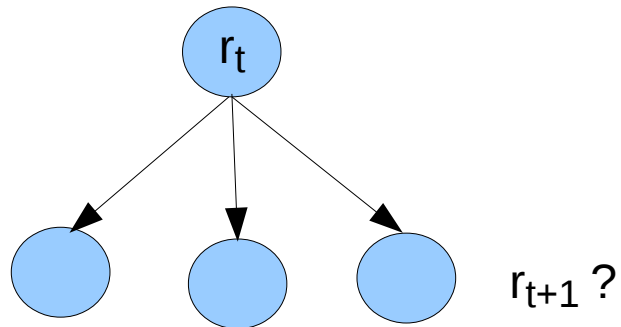
Meccanismo di generalizzazione:

rimozione di congiunti finalizzata a coprire un maggior numero di istanze positive



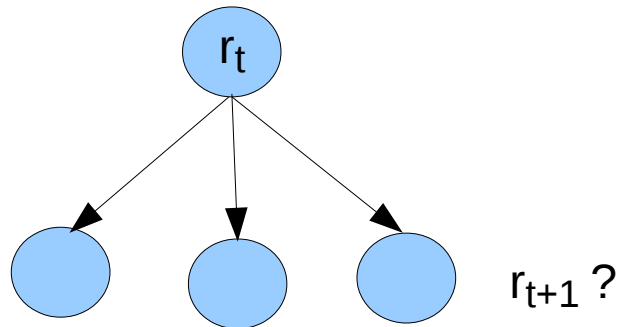
If (viviparo) \wedge (sangue caldo) \wedge (volatile) \rightarrow mammifero

Beam-search vs. greedy-search

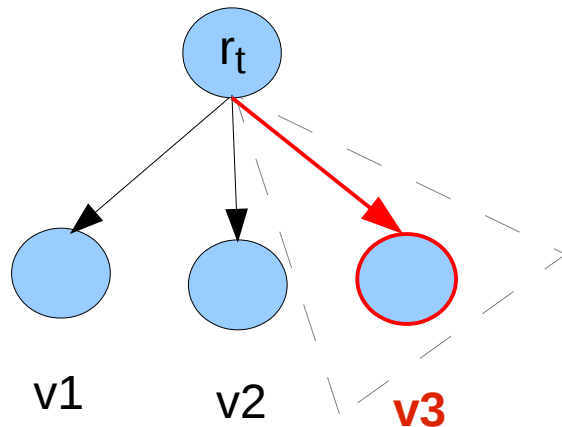


In generale data una regola esistono diversi modi per generalizzarla o per specializzarla. Quale sarà la migliore?

Beam-search vs. greedy-search



In generale data una regola esistono diversi modi per generalizzarla o per specializzarla. Quale sarà la migliore?

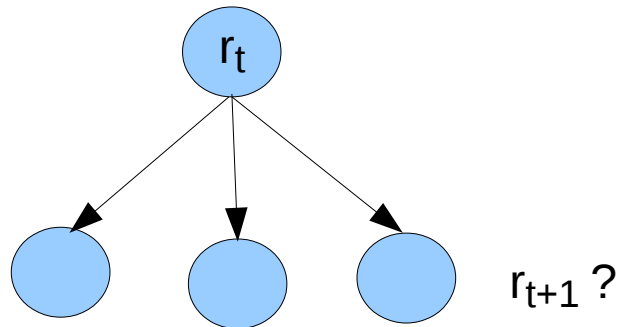


Greedy:

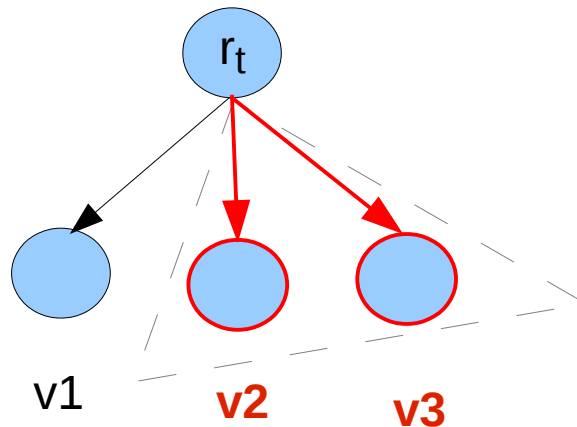
valuto tutte le alternative e espando quella con la valutazione più alta

$v3 > v2$
 $v3 > v1$

Beam-search vs. greedy-search



In generale data una regola esistono diversi modi per generalizzarla o per specializzarla. Quale sarà la migliore?



Beam search:

valuto tutte le alternative e espando
Quelle le cui valutazioni sono più
promettenti

Valutazione delle regole

Per generalizzare/specializzare come anche per ordinare le regole occorre poterle valutare. Si possono adottare molti criteri estremamente vari:

- (1) accuratezza (ma da sola non basta)
- (2) information gain
- (3) test statistici sulla copertura
- (4) metriche che tengono conto della copertura

Le regole possono essere sottoposte a pruning (come per gli alberi di decisione)
(dettagli sul libro)

Lazy learning

Eager vs Lazy learners



Eager learner

Alberi di decisione
Regole
RL

Dati gli esempi si mettono subito all'opera per costruire un modello che in futuro useranno per classificare nuove istanze



Lazy learner

Rote learner
Nearest Neighbour
(k-NN)

Dati gli esempi non se ne fanno nulla finché non viene il momento di classificare nuove istanze

Studiare a memoria



att1	att2	att3	att4	classe
a	b	c	d	A
a	f	b	b	B
b	b	c	d	A
c	c	d	f	C

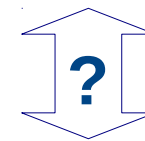
Il sistema memorizza gli esempi
senza effettuare alcuna genera-
lizzazione

Rote learner

Studiare a memoria



att1	att2	att3	att4	classe
a	b	c	d	A
a	f	b	b	B
b	b	c	d	A
c	c	d	f	C



a	f	b	b
---	---	---	---

Quando viene presentata una nuova istanza, questa viene confrontata con tutti gli esempi memorizzati

Rote learner

Studiare a memoria



att1	att2	att3	att4	classe
a	b	c	d	A
a	f	b	b	B
b	b	c	d	A
c	c	d	f	C

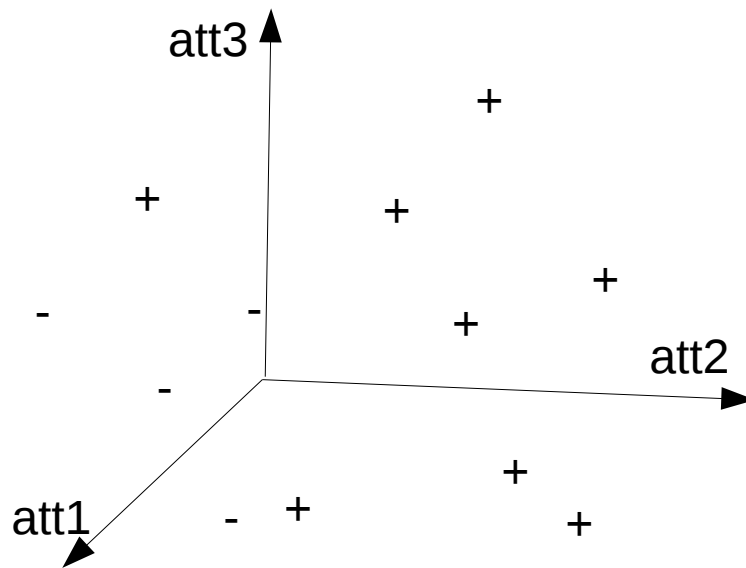


a	f	b	b
---	---	---	---

Se si trova un'istanza identica a quella in esame allora si restituisce la classe ad essa associata

K-nearest neighbour

"if it walks like a duck, quacks like a duck, and looks like a duck, then it is probably a duck"



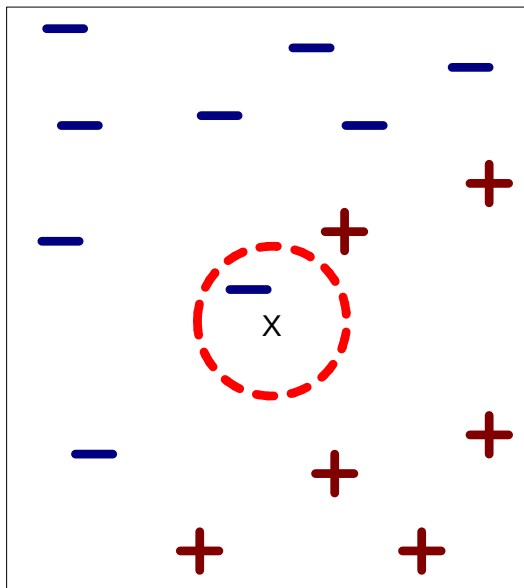
Learning set supervisionato

Gli esempi sono visti come **punti** di uno spazio N-dimensionale

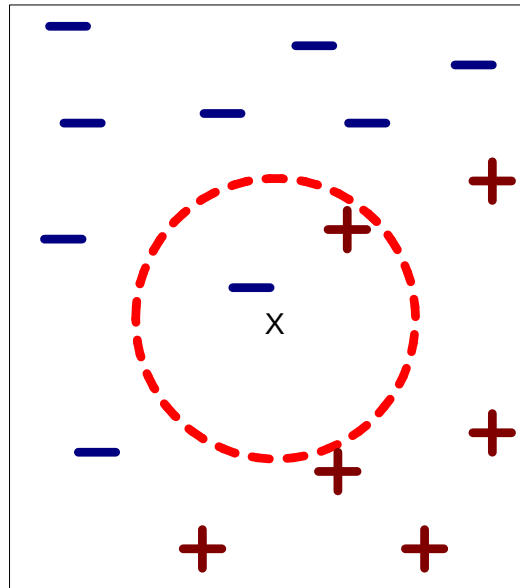
N = numero degli attributi descrittivi

Le nuove istanze sono classificate sulla base di un **criterio di vicinanza** a quelle memorizzate dal sistema

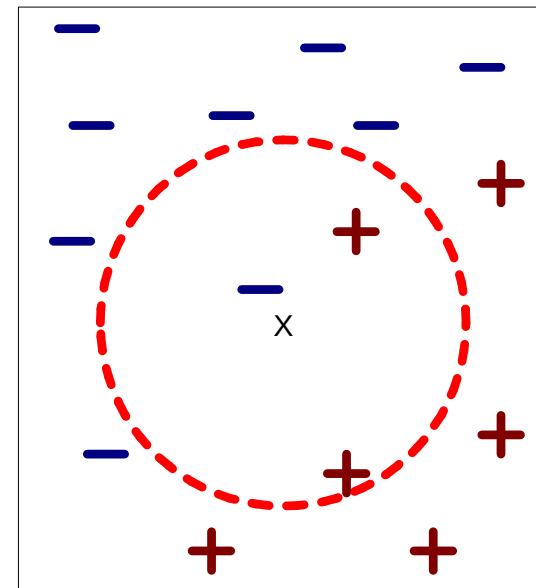
K-nearest neighbour



(a) 1-nearest neighbor



(b) 2-nearest neighbor

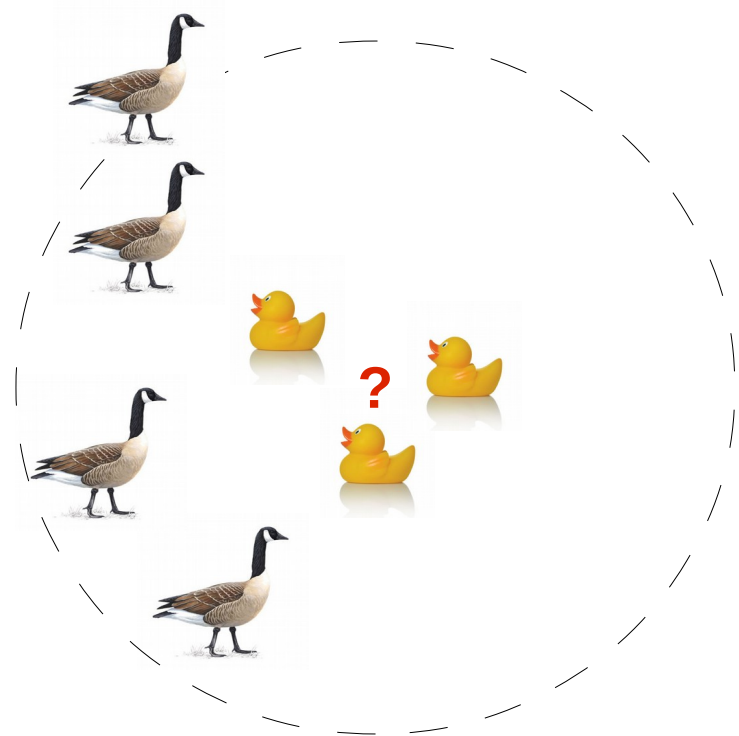
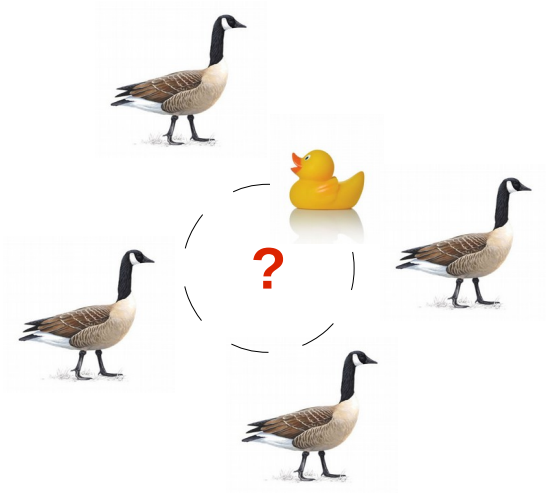


(c) 3-nearest neighbor

Definizione del k

Il problema è determinare il valore k più adeguato per il problema di classificazione di interesse:

- **k piccolo:** rischio di overfitting
- **k grande:** rischio di confusione dovuta al considerare esempi troppo distanti



Algoritmo: k-NN

Siano

K : numero di esempi di training da considerare per la classificazione

D : training set

x' : istanza da classificare

Foreach istanza $x \in D$ **do**

 Calcola la **distanza** d fra x' e x

 Aggiorna l'insieme $kset$ delle k istanze più vicine a x'

done

$y' = \text{calcolaClasse}(kset)$

Calcolo della classe

K-NN calcola la classe di appartenenza dell'esempio componendo i “suggerimenti” dei k esempi più vicini ad esso

Majority-voting:

$$y' = \underset{v}{\operatorname{argmax}} \sum_{(x_i, y_i) \in D} I(v = y_i)$$

y_i = classe del vicino v

$I(.)$ = fz che restituisce 1 se il suo argomento è vero

Risultato = Indice della classe più rappresentata

Distance-weighted Voting:

$$y' = \underset{v}{\operatorname{argmax}} \sum_{(x_i, y_i) \in D} \frac{I(v = y_i)}{d(x', x_i)^2}$$

Per rendere il risultato *meno dipendente dal numero k* si pesa ogni contributo dividendolo per la distanza fra il vicino e l'istanza da classificare:

Le ipotesi degli esempi più vicini contano di più

Pro e contro

K-NN e rote learning sono forme di “instance-based learning”, metodi di classificazione che non richiedono la costruzione di un modello:

- **pro:** viene evitata la complessità della costruzione del modello
- **pro:** la partizione dello spazio degli attributi non è rettangolare, quindi complessivamente il classificatore è più flessibile
- **contro:** ogni istanza da classificare deve essere confrontata con tutte le istanze memorizzate
- **attenzione:** le misure di distanza risentono della variabilità delle scale dei valori assunti dai vari attributi!

Att1 in [0.1, 0.9]

Att2 in [100, 1000]

Il secondo attributo è prevalente nel calcolo della distanza, anche qualora fosse meno rilevante al fine di distinguere le diverse classi