

Lezione 4

▼ Corso	Riconoscimento e recupero dell'informazione per Bioinformatica
📅 Data	@October 25, 2021 1:30 PM
☑ Rifacimento	<input type="checkbox"/>
▼ Status	In progress
▼ Tipo	Lezione

Una volta costruita una rappresentazione, dobbiamo costruire un modello.

Classificazione

Abbiamo un insieme di addestramento → come ho rappresentato gli oggetti e un insieme di etichette: ho un insieme di oggetti del problema per cui conosco la classe vera, cosa voglio fare? **Costruire un modello**: in grado di discriminare tra le due classi.

Nello spazio delle feature visualizzo le dimensioni e gli oggetti nello spazio.

Uso i modelli per classificare un oggetto: il modello può essere assegnato ad una classe, una categoria.

1. Come costruire il modello?
2. Come classifico gli oggetti in base al mio modello?

1. Troviamo un sistema di classificazione

Ho qualcosa che viene applicato alla rappresentazione dell'oggetto in ingresso e mi ritorna la classe: l'output è discreto, è la classe.

Ho una funzione $f(x) = y$, dove y è l'output discreto.

Voglio trovare una funzione che sbagli il meno possibile (errore di generalizzazione) → ricordiamo che il modello deve essere in grado di generalizzare.

Non devo avere errore di classificazione (un oggetto di una classe 1 viene classificato come di classe 2)

Costruire la funzione $f \rightarrow$ **TEORIE DELLA DECISIONE**

è un insieme di regole, approcci e metodi che ci permettono di costruire un classificatore:

- come vengono caratterizzate le entità
- come viene determinata la regola di decisione
- come possono essere interpretate le soluzioni

TEORIE PIÙ IMPORTANTI

- Teoria di Bayes
- Statistical Learning Theory - più complessa

Non c'è una chiara separazione tra le teorie

Teorema di Bayes

- Approccio molto utilizzato e storicamente provato (ci sono molti risultati teorici: importante nel contesto della qualità dell'algoritmo).
 - Il problema è posto in termini probabilistici (probabilità di una classe, di fare un errore, ecc.)
 - Tutte le probabilità sono conosciute

Come funziona?

Uso le probabilità per costruire il classificatore.

Def: ω è la classe dell'oggetto

ω_1 = prima classe

La regola di classificazione mi permette di rispo



Dato un oggetto x , lo assegno a ω_1 , a ω_2 o a ω_3 ? In altre parole: a quale classe deve essere assegnato x ?

Ci sono diverse probabilità importanti all'interno della teoria della decisione di Bayes:

- **Probabilità a priori**

codifica l'informazione a priori, capacità descrittiva molto bassa

es. rappresenta la probabilità dello stato che è nota a priori (non ho osservato nulla)

- **Probabilità condizionale**

mi dice la probabilità all'interno di una determinata classe

es. se x è lo stipendio, la distribuzione dello stipendio è fatta in un certo modo

quindi se sto considerando solo i calciatori, è come si distribuiscono gli stipendi, quindi distribuisco gli oggetti all'interno di una classe

è molto più descrittiva di quella a priori, ma non tiene conto di quest'ultima → non abbastanza descrittiva

- **Probabilità a posteriori**

Ho un modo di descrivere il problema che tiene conto di tutte le probabilità.

$$P(A|B) = \frac{P(A,B)}{P(B)} \rightarrow P(B)P(A|B) = P(A, B)$$

$$P(B|A) = \frac{P(B,A)}{P(A)} \rightarrow P(A)P(B|A) = P(B, A)$$

Dato che $P(A, B)$ e $P(B, A)$

$$P(B)P(A|B) = P(A)P(B|A)$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

»

La regola di decisione di Bayes si basa sulla probabilità a posteriori, e permettere di classificare un oggetto x : si può dimostrare che la regola di decisione di Bayes minimizza la probabilità di errore.

Regola di decisione di Bayes:



Dato x , decisi ω_1 se $p(\omega_1|x) < p(\omega_2|x)$, ω_2 altrimenti.

Equivalentemente:

- l'evidenza rappresenta un fattore di scala che descrive quanto frequentemente si osserva un pattern x
- non dipende da ω_1 o da ω_2 , quindi è influente per la regola di decisione

$$posterior = \frac{likelihood \times prior}{evidence}$$

La regola utilizza le posterior (non note), per stimare le posterior si possono usare le likelihood e le prior.

Approcci Generativi

Un modello per ogni classe

Parentesi: la densità normale

è analiticamente trattabile, fornisce la migliore modellazione di problemi sia teorici che pratici. Il teorema del Limite centrale asserisce che "sotto varie condizioni, la distribuzione della somma di d variabili aleatorie indipendenti tende ad un limite"

/formule

La varianza misura la larghezza della campana, posso avere una gaussiana anche multidimensionale (formula simile -ma con matrici e vettori)

Densità normale multivariata

Approccio generativo → voglio stimare le proprietà di una classe: es- la mia classe segue una distribuzione gaussiana, stimo i parametri (media e deviazione standard): modello le classi quindi generativo, e approccio parametrico

Nel caso bidimensionale i due parametri sono media e varianza.

Matrice di covarianza

Una matrice 2×2 nel caso bidimensionale; misura quanto varia una feature rispetto a quanto varia un'altra feature - cattura la relazione che c'è tra i due parametri e mi dà un'idea della forma.

Nel caso in cui i parametri siano scorrelati, la covarianza è 0.

$$M = \begin{bmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{bmatrix}$$

Ho zero perché non ho relazioni tra le due covarianze.

I numeri che trovo sulla diagonale sono la varianza di x e quella di y. In un caso generico ho:

$$M = \begin{bmatrix} \sigma_x & \sigma_{x,y} \\ \sigma_{x,y} & \sigma_y \end{bmatrix}$$

Quando assumo che una classe ha distribuzione gaussiana posso avere un'idea di come sono correlate le feature. Quando so la forma diminuiscono anche i parametri che devo stimare.

-
- Stimo le probabilità a priori e le probabilità condizionali
 - Stime parametriche:
 - stimo i parametri data la forma della pdf
 - es. nella gaussiana stimo la media
 - Stimo la probabilità a priori
 - Si combinano a formare le probabilità a posteriori
 - Rappresenta il vero problema, poiché voglio stimare i parametri sconosciuti della funzione conosciuta
1. Stimola forma
 2. Ne stimo i parametri

Costruire un classificatore bayesiano con stima parametrica

Si stima dal training set la probabilità a priori per ogni classe

Per la probabilità condizionale: si decide o si stima, la forma per ogni classe (ad esempio gaussiana) & si stimano i parametri a partire dai dati di training (un insieme di parametri per ogni classe)

- Si classifica con la regola di Bayes

Stima non parametrica

Con la stima parametrica assumo che la forma delle densità di probabilità sia nota, ma questa assunzione non può sempre essere fatta.

Se la scelta della forma è sbagliata, la stima sarà povera.

Soluzione: metodi non parametrici

Approcci discriminativi

Modello direttamente alla fine