

Índice general

1. Panoramica del apprendimento supervisionato	2
1.0.1. Introduzione	2
1.0.2. Tipi di variabili e terminologia	2
1.0.3. Due	4

Capítulo 1

Panoramica del apprendimento supervisionato

1.1. Introduzione

I primi tre esempi descritti nel Capitolo 1 hanno parecchi componenti in comune. Per ogni c'è un insieme di variabili che potrebbe essere denotate come *input*, le quale sono misurate o predefinite. Quelle hanno alcuna influenza su uno o più *output*. In ogni esempio, il obiettivo è usare gli input per predire i valori degli output. Questo esercizio si chiama *apprendimento supervisionato*.

Abbiamo usato il linguaggio più moderno del apprendimento automatico. Nella letteratura statistica gli input sono spesso chiamati *predittori*, un termine che useremo in modo intercambiabile con *input*, e più classicamente si chiamano *variabili indipendenti*. Nella letteratura di riconoscimento di pattern, si preferisce il termine *caratteristiche*, il quale useremo anche noi. Gli output sono chiamati le *risposte*, o classicamente le *variabile dipendenti*.

1.2. Tipi di variabili e terminologia

Gli output variano in natura tra gli esempi. Nel esempio di predizione del glucosio, il output è una misura quantitativa, dove alcune misurazioni sono più grandi degli altri, e le misurazioni vicine in valore sono vicine in natura. Nel famoso esempio di discriminazione di Iris dovuto a R. A. Fisher, il output è qualitativo (la specie di Iris) e prende valori di un insieme finito $\mathcal{G} = \{Virginica, Setosa \text{ and } Versicolor\}$. Nel esempio delle cifre scritte a mano, il output è uno delle 10 classi diverse di cifre: $\mathcal{G} = \{0, 1, \dots, 9\}$. In entrambi non c'è un ordine esplicito nelle classi, e infatti si usano spesso etichette descrittive piuttosto che numeri per indicare le classi. Variabili qualitative sono anche indicate come variabili *categoriche* o *discrete* così come *fattori*.

Per entrambi i tipi di output, ha senso di pensare di usare gli input per predire il output. Date alcune misurazioni atmosferiche specifiche di oggi e ieri, vogliamo predire il livello di ozono di domani. Dati i valori in scala di grigi per i pixel di una immagine digitalizzata della cifra scritta

a mano, vogliamo predire la etichetta della sua classe.

Questa distinzione nel output ha portato a una convenzione di nome per i compiti di predizione: *regressione* quando prediciamo output quantitativi, e *classificazione* quando prediciamo output qualitativi. Vedremo che questi due compiti hanno molto in comune, e in particolare, entrambe possono essere visualizzati come un compito di approssimazione di funzione.

Gli input anche variano nel tipo di misurazione; possiamo avere alcuni variabili di input di ogni tipo qualitativo e quantitativo. Questi hanno anche portato a distinzioni nei tipi di metodi che sono usati per predire: alcuni metodi sono definiti più naturalmente per input quantitativi, alcuni più naturalmente per qualitativi, e alcuni per entrambi.

Un terzo tipo di variabile è la categorica ordinata, come *piccolo*, *mediano* e *grande*, dove c'è un ordine tra i valori, ma nessuna nozione metrica è adeguata (la differenza tra mediano e piccolo non necessariamente è la stessa di quella tra grande e mediano). Queste sono discusse ulteriormente nel Capitolo 4.

Le variabili qualitative tipicamente sono ripresentate numericamente con codici. Il caso più semplice è quando ci sono due classi, come “successo” e “fallimento”, “sopravvissuto” o “morto”. Queste spesso sono ripresentate con una singola cifra binaria o bit come 0 o 1, o invece come -1 e 1 . Per ragioni che diventeranno apparenti, tali codici numerici si denominano a volte come *target* o obiettivi. Quando ci sono più di due categorie, ci sono diverse alternative. La codifica più utile e più comunemente usata è tramite variabili di comodo (o variabili *dummy*). Qui, una variabile qualitativa di K livelli è ripresentata con un vettore di K variabili binari o bit, solo una delle quali è “accesa” in ogni momento. Sebbene sono possibili codifiche più compatte, le variabili di comodo sono simmetriche nei livelli del fattore.

Tipicamente denotiamo una variabile input con il simbolo X . Se X è un vettore, è possibile accedere i suoi componenti con pedici X_j . Output quantitativi saranno denotate come Y , e output qualitativi come G (di gruppo). Usiamo lettere maiuscole come X , Y o G quando intendiamo gli aspetti generici di una variabile. Valori osservati sono scritte in minuscola; quindi il i -esimo valore osservato di X è scritto come x_i (dove x_i è, di nuovo, uno scalare o un vettore). Le matrici sono rappresentate con lettere maiuscole in grassetto; per esempio, un insieme di N p -vettori di input x_i , $i = 1, \dots, N$ sarebbe rappresentato con la matrice \mathbf{X} di $N \times p$. In generale, i vettori non saranno scritti in grassetto, tranne quando hanno N componenti; questa convenzione distingue un p -vettore di input x_i per la i -esima osservazione dal N -vettore \mathbf{x}_j che consiste di tutti i osservazione della variabile X_j . Poiché si assumono tutti i vettori di essere vettori di colonna, la i -esima riga di \mathbf{X} è x_i^T , il vettore trasposto di x_i .

Per il momento, possiamo vagamente dichiarare il compito di apprendimento come segue: dato il valore di un vettore di input X , fa una buona predizione del output Y , indicata come \hat{Y} (pronunciato “y-cappuccio”). Se Y prende valori in \mathbb{R} allora così dovrebbe \hat{Y} ; allo stesso modo per output categorici, \hat{G} dovrebbe prendere valori dello stesso insieme \mathcal{G} associato a G .

Per una G di due classi, un approccio è indicare il target con codice binario come Y , e poi trattarlo come un output quantitativo. Le predizioni \hat{Y} tipicamente si troveranno in $[0, 1]$, e possiamo assegnare a \hat{G} la etichetta di classe a seconda che $\hat{y} > 0.5$. Questo approccio anche generalizza agli output qualitative di K livelli.

Abbiamo bisogno dei dati per costruire le rigole di predizione, spesso un'enorme quantità di dati. Quindi supponiamo che abbiamo a disposizione un insieme di misurazioni (x_i, y_i) o

(x_i, g_i) , $i = 1, \dots, N$, conosciuto come i *dati di allenamento*, con i quali costruire la nostra rigola di predizione.

1.3. Due