

Prima prova parziale e traccia della soluzione

Mercoledì 2 novembre 2016

- Riportare il proprio nome, cognome e numero di matricola in cima a questo foglio e a tutti i fogli, di bella e di brutta copia.
- Al termine dello svolgimento della prova, è necessario riconsegnare *tutti* i fogli, comprese le brutte copie e il presente testo. In caso di riconsegna parziale la prova non verrà valutata.
- Il presente foglio non deve riportare alcuna scritta ad eccezione dei dati di identificazione.
- Durante lo svolgimento della prova non è consentito l'uso di libri, appunti, dispositivi elettronici.
- Non è consentito uscire prima della consegna, che può avvenire in qualunque momento. Una volta usciti, non sarà consentito il rientro.
- Gli esercizi 1 e 2 valgono 10 punti ciascuno. Le 13 domande dell'esercizio 3 valgono 1.2 punti ciascuna (+1.2 se la risposta è corretta, -1.2 se è errata, 0 per le risposte non date).

Esercizio 1

È dato il seguente dataset, di 12 campioni $i = 1, \dots, 12$, composti da una sola feature scalare $x_i \in [0, 10]$ come variabile indipendente e una classe a due valori $y_i \in \{\text{Freddo}, \text{Caldo}\}$ come valore da prevedere:

i	x_i	y_i
1	0.7	Caldo
2	8.7	Freddo
3	3.5	Freddo
4	1.5	Caldo

i	x_i	y_i
5	4.4	Freddo
6	6.1	Caldo
7	6.8	Caldo
8	7.2	Caldo

i	x_i	y_i
9	2.9	Freddo
10	1.9	Freddo
11	3.0	Caldo
12	6.0	Freddo

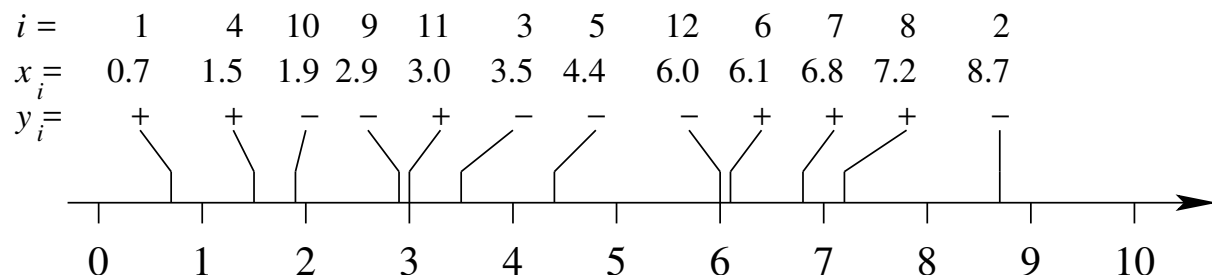
1.1) Tramite la metodologia leave-one-out, valutare la prestazione dell'algoritmo K -nearest-neighbors con $K = 1$ e $K = 3$ sulla base degli indici di accuratezza, precisione, sensibilità e dell' F_1 -score.

1.2) Quale scelta si rivela essere la migliore per il parametro K ? Lo è in maniera univoca, oppure dipende dal criterio considerato?

Suggerimento — Riportare i valori x_i su un asse in modo da semplificare la ricerca dei vicini. In caso di parità di distanze, scegliere l'elemento di indice minore.

Soluzione 1

Consideriamo “Caldo” come positivo, “Freddo” come negativo. In base al suggerimento, ecco i valori riportati sulle ascisse:



i	Vicini			y_i	\tilde{y}_i	
					$K = 1$	$K = 3$
1	+	-	-	+	+	-
2	+	+	+	-	+	+
3	+	-	-	-	+	-
4	-	+	-	+	-	-
5	-	+	-	-	-	-
6	-	+	+	+	-	+
7	+	+	-	+	+	+
8	+	+	-	+	+	+
9	+	-	-	-	+	-
10	+	-	+	-	+	+
11	-	-	-	+	-	-
12	+	+	+	-	+	+

1.1) La matrice di confusione per $K = 1$ è dunque

		Previsione	
		+	-
Risposta corretta	+	3	3
	-	5	1

da cui risulta:

$$\text{Accuratezza} = \frac{3+1}{12} = \frac{1}{3}, \quad \text{Precisione} = \frac{3}{3+5} = \frac{3}{8}, \quad \text{Sensibilità} = \frac{3}{3+3} = \frac{1}{2}, \quad F_1 = \frac{2\frac{3}{8}\frac{1}{2}}{\frac{3}{8}+\frac{1}{2}} = \frac{3}{7}.$$

Si osservi che, invertendo l'interpretazione della classe positiva e negativa, si ottiene invece la seguente matrice:

		Previsione	
		+	-
Risposta corretta	+	1	5
	-	3	3

da cui risulta:

$$\text{Accuratezza} = \frac{1+3}{12} = \frac{1}{3}, \quad \text{Precisione} = \frac{1}{1+3} = \frac{1}{4}, \quad \text{Sensibilità} = \frac{1}{1+5} = \frac{1}{6}, \quad F_1 = \frac{2\frac{1}{4}\frac{1}{6}}{\frac{1}{4}+\frac{1}{6}} = \frac{1}{5}.$$

Allo stesso modo, la matrice di confusione per $K = 3$ risulta

		Previsione	
		+	-
Risposta corretta	+	3	3
	-	3	3

da cui risulta:

$$\text{Accuratezza} = \frac{3+3}{12} = \frac{1}{2}, \quad \text{Precisione} = \frac{3}{3+3} = \frac{1}{2}, \quad \text{Sensibilità} = \frac{3}{3+3} = \frac{1}{2}, \quad F_1 = \frac{1}{2}.$$

Visto che la precisione e la sensibilità sono uguali, allora anche la loro media F_1 è lo stesso valore.

1.2) Rispetto a tutti gli indici considerati, $K = 3$ risulta preferibile o indifferente rispetto a $K = 1$.

Esercizio 2

È dato il seguente dataset con variabili indipendenti scalari $x_i \in [0, 100]$ e variabile dipendente continua $y_i \in \mathbb{R}$:

i	1	2	3	4	5	6
x_i	27	12	69	70	55	31
y_i	57	36	95	90	77	51

Si desidera modellare il dataset tramite una regressione affine nella forma

$$y \sim \beta_0 + \beta_1 x.$$

Determinare graficamente (tracciando i punti e le rette su un piano cartesiano) quale tra le seguenti coppie di coefficienti genera l'errore quadratico minore:

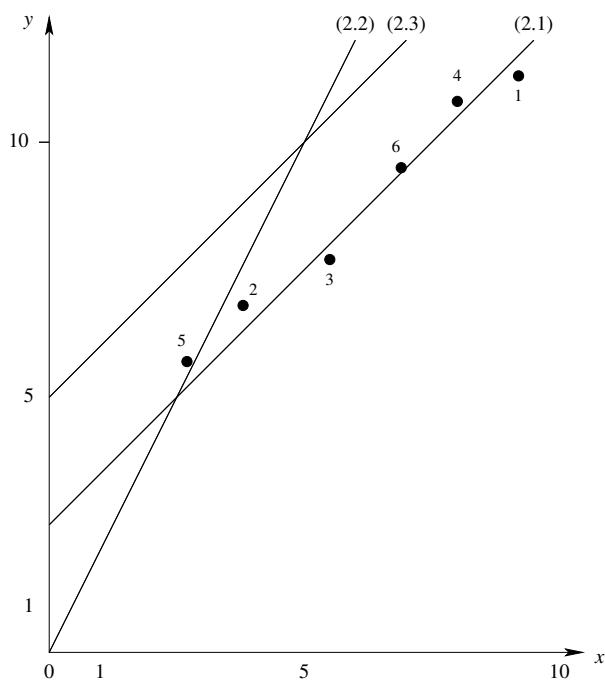
2.1) $\beta_0 = 0, \beta_1 = 2$

2.2) $\beta_0 = 25, \beta_1 = 1$

2.3) $\beta_0 = 50, \beta_1 = 1$

Soluzione 2

Ecco i punti e le rette richieste riportate nel piano cartesiano:



Risulta immediatamente chiaro che la retta che meglio approssima i punti fra le tre fornite è la **(2.1)**, corrispondente all'equazione affine $y \sim 2.5 + x$.

Esercizio 3

Per ciascuna delle seguenti domande, riportare nel foglio protocollo il numero della risposta ritenuta corretta. Si prega di non segnare in alcun modo le domande e le risposte sul foglio.

In caso di incertezza è consentito motivare una risposta con una riga di testo.

1. Quando si dice che un sistema di machine learning è “overfitting”?
 - (a) Quando genera previsioni sistematicamente troppo alte.
 - (b) Quando è troppo specializzato sui dati di training e non è in grado di generare previsioni adeguate su dati nuovi.
 - (c) Quando è stato addestrato su un insieme di training troppo ampio e restano pochi dati per validarlo.
2. Quanti coefficienti vanno determinati in una regressione polinomiale di secondo grado in una variabile?
 - (a) Quattro.
 - (b) Tre.
 - (c) Cinque.
3. Quale misura indica il numero di previsioni corrette per un modello di classificazione?
 - (a) Accuratezza.
 - (b) Precisione.
 - (c) Sensibilità.
4. Come possiamo classificare il formato CSV?
 - (a) Non strutturato.
 - (b) Semi-strutturato.
 - (c) Strutturato.
5. Perché è preferibile che gli insiemi di training e di validazione siano disgiunti?
 - (a) Perché siamo interessati a valutare le prestazioni del sistema su esempi non visti durante l’addestramento.
 - (b) Si tratta di una precauzione per evitare che la presenza di due elementi identici nei dataset causi divisioni per zero.
 - (c) Perché dobbiamo minimizzare le dimensioni dei due insiemi.
6. Che effetto ha la funzione sigmoide sull’uscita di un regressore?
 - (a) Mappa il valore reale in uscita sull’intervallo $[0, 1]$.
 - (b) Determina un valore di soglia per la decisione della classe di uscita.
 - (c) Decide la classe di uscita.
7. Come può essere definita una funzione sigmoide?
 - (a) In entrambi i modi.
 - (b) $\frac{1}{1+e^{-t}}$.
 - (c) $\frac{e^t}{e^t+1}$.
8. A cosa serve normalizzare le colonne di un dataset?
 - (a) A eguagliare gli intervalli di variabilità delle colonne.
 - (b) A eliminare eventuali elementi uguali a zero.

- (c) A rimuovere eventuali valori negativi.
9. Quale misura occorre massimizzare per ridurre il numero di falsi negativi per un modello di classificazione?
- (a) Precisione.
 - (b) Sensibilità.
 - (c) Accuratezza.
10. Quando una partizione di un dataset si dice *stratificata*?
- (a) Quando i campioni di ciascun sottoinsieme della partizione si trovano nello stesso ordine in cui compaiono nel dataset originale.
 - (b) Quando le numerosità dei diversi valori della classe di uscita nei sottoinsiemi della partizione sono approssimativamente uguali a quelle dell'intero dataset.
 - (c) Quando campioni simili vengono messi di preferenza nella stessa partizione.
11. È possibile utilizzare KNN per la classificazione se la classe in uscita ha più di due valori?
- (a) Sì, ma è necessario addestrare un classificatore KNN per ciascun valore della classe, e poi confrontarne l'output.
 - (b) Sì, in quanto la funzione di decisione si basa sul valore di maggioranza, indipendentemente dal loro numero.
 - (c) No, è adatto solo a problemi di regressione.
12. In cosa consiste la K -fold cross-validation?
- (a) Si separano i campioni in K gruppi distinti che si usano a rotazione per la validazione.
 - (b) Si usano K campioni alla volta per il training e i restanti per la validazione.
 - (c) Si separano le feature in K gruppi distinti che si usano a rotazione per la validazione.
13. Che cos'è il metodo della discesa lungo il gradiente?
- (a) Un metodo per trovare il valore ottimale di un classificatore KNN.
 - (b) Un metodo per trovare un minimo locale di una funzione differenziabile in più variabili reali.
 - (c) Un metodo per trovare le derivate parziali dell'output di un regressore logistico.

Soluzione 3

N.B.: il compito conteneva 8 domande fra le 13 sopra elencate.

Ecco la griglia delle risposte corrette:

Domanda	1	2	3	4	5	6	7	8	9	10	11	12	13
Risposta	b	b	a	c	a	a	a	a	b	b	b	a	b