

1) Come si calcola l'attivazione (net) di un neurone artificiale? Indicare formula e commentarla.

Dispense "Reti Neurali"

2) Quali sono le condizioni necessarie affinché le tecniche di deep learning siano più efficaci di altri approcci di machine learning?

Dispense "Deep Learning (parte 1)"

3) Cosa si intende per Clustering esclusivo e Clustering soft (o Fuzzy). Quest'ultimo che vantaggi può avere?

Dispense "Clustering"

4) Indicare la formula di Bayes per la probabilità a posteriori, definendo i termini.

Dispense "Classificazione (parte 1)"

5) Dato un training set di 20000 pattern, supponendo di addestrare una rete neurale con SGD (mini-batch size=400) per 60 epoche, riportare, motivandone la risposta:

1. il numero di volte in cui ciascun pattern viene visto dalla rete;
2. il numero di volte in cui ciascun peso della rete viene aggiornato.

Svolgimento

Ad ogni epoca tutti i pattern del training set vengono presentati alla rete esattamente una volta. Pertanto se la rete viene addestrata per 60 epoche, ciascun pattern le verrà mostrato 60 volte.

Nell'approccio SGD, l'aggiornamento dei pesi avviene alla fine di ogni iterazione che corrisponde alla presentazione di un intero mini-batch alla rete. Avendo 20000 pattern raggruppati in mini-batch di dimensione 400, avremo un totale di $\frac{20000}{400} = 50$ iterazioni per epoca. Ne consegue che il numero di volte in cui ogni peso della rete viene aggiornato corrisponde a $60 \cdot 50 = 3000$.

6) Un problema di multiple linear regression viene risolto ai minimi quadrati ottenendo su un training set i seguenti coefficienti $\beta = [-7.4 \quad 2.7 \quad 2.5 \quad -6.1]$ (dove l'ultimo elemento è il termine noto). Dato un test set costituito dai tre pattern \mathbf{x}_1 , \mathbf{x}_2 e \mathbf{x}_3 (di cui y_1 , y_2 e y_3 sono i valori veri della variabile dipendente):

$$\mathbf{x}_1 = [-3.2 \quad 8.9 \quad -8.1], y_1 = -7.4$$

$$\mathbf{x}_2 = [8.5 \quad 4.3 \quad -0.4], y_2 = 8.6$$

$$\mathbf{x}_3 = [-6.7 \quad 1.3 \quad 7.7], y_3 = -2.4$$

determinare il MAE (*Mean Absolute Error*) sul test set, riportando i principali passaggi intermedi del calcolo.

Svolgimento

Un problema di multiple linear regression consiste nella generalizzazione a iperpiani di un problema di linear regression. Dato un pattern di test \mathbf{x} , è possibile predire il valore della variabile dipendente y a partire dai valori dei coefficienti β utilizzando la seguente formula:

$$y = \sum_{i=1 \dots d+1} x_i \cdot \beta_i$$

Con $x_{d+1} = 1$ (per considerare il termine noto β_{d+1}).

Eseguendo i calcoli per i pattern di test si ottengono le seguenti predizioni:

$$\hat{y}_1 = (-3.2) \cdot (-7.4) + 8.9 \cdot 2.7 + (-8.1) \cdot 2.5 + 1.0 \cdot (-6.1) = 21.3$$

$$\hat{y}_2 = 8.5 \cdot (-7.4) + 4.3 \cdot 2.7 + (-0.4) \cdot 2.5 + 1.0 \cdot (-6.1) = -58.4$$

$$\hat{y}_3 = (-6.7) \cdot (-7.4) + 1.3 \cdot 2.7 + 7.7 \cdot 2.5 + 1.0 \cdot (-6.1) = 66.3$$

Il MAE del test set è dato della media (su tutto il test set) del valore assoluto degli errori. In questo caso:

$$MAE = \frac{|y_1 - \hat{y}_1| + |y_2 - \hat{y}_2| + |y_3 - \hat{y}_3|}{3} = \frac{28.7 + 67 + 68.7}{3} = 54.8$$

7) Un multiclassificatore, composto da 4 classificatori combinati a livello di confidenza, viene utilizzato per riconoscere pattern appartenenti a 3 classi (A, B, C). Nella tabella seguente sono riportate le confidenze restituite dai singoli classificatori (C_i) dati in input 3 diversi pattern (\mathbf{p}_j). Completare la tabella riportando, per ogni metodo di fusione (Somma, Prodotto, Massimo e Minimo), le confidenze ottenute e la classe di output restituita dal multiclassificatore.

	C_1			C_2			C_3			C_4		
	A	B	C	A	B	C	A	B	C	A	B	C
\mathbf{p}_1	0.24	0.44	0.32	0.39	0.28	0.33	0.46	0.33	0.21	0.17	0.68	0.15
\mathbf{p}_2	0.15	0.22	0.63	0.64	0.15	0.21	0.37	0.41	0.22	0.16	0.32	0.52
\mathbf{p}_3	0.61	0.24	0.15	0.60	0.25	0.15	0.30	0.41	0.29	0.15	0.17	0.68

Svolgimento

	Somma				Prodotto				Massimo				Minimo			
	A	B	C	Out	A	B	C	Out	A	B	C	Out	A	B	C	Out
\mathbf{p}_1	1.26	1.73	1.01	B	0.01	0.03	0	B	0.46	0.68	0.33	B	0.17	0.28	0.15	B
\mathbf{p}_2	1.32	1.10	1.58	C	0.01	0	0.02	C	0.64	0.41	0.63	A	0.15	0.15	0.21	C
\mathbf{p}_3	1.66	1.07	1.27	A	0.02	0	0	A	0.61	0.41	0.68	C	0.15	0.17	0.15	B