

IIA ML Esercitazioni IV (testo degli esercizi)

WARNING: E' estremamente utile per voi stessi provare autonomamente a risolvere gli esercizi (nei file testi-domande) PRIMA di vedere queste soluzioni

Suggerimenti generali:

- Si vedano le note iniziali negli altri file
- Nota: Questi esercizi a seguire sono in forma più di domande, con risposte multiple o libere, adeguate anche alle prove dei test su elearning, soprattutto per gli esercizi **1, 3, 4, 5, 6 e 7** (anche se alcuni erano già presenti nei file di esercitazioni precedenti, qui si vogliono fornire altri esempi)

1 – Confrontare la complessità (flessibilità in fitting) delle seguenti classi di modelli , ove $x > y$ esprime che x ha complessità (potenzialmente) maggiore di y

- a. modello lineare (negli input) con 3 variabili di input
- b. modello lineare (negli input) con 5 variabili di input
- c. K-nn con 3 variabili di input e K=1
- d. K-nn con 3 variabili di input e K=5
- e. modello con 3 variabili di input da algoritmo Find-S
- f. decision tree con 3 variabili di input

In particolare, queste affermazioni sono corrette? (esprimere tutte quelle corrette ma affermazioni indicate come corrette che non lo sono incidono negativamente sul punteggio dell'esercizio).

1. a>b
2. f>e
3. d>c
4. a>c
5. c> e
6. f> a

2 - Rispondere ai seguenti quesiti sul Machine Learning.

Nota: risposte valide se contengono la risposta all'affermazione/domanda (Si/No o dire in che casi sia Si o No) e la motivazione [con testo libero, sintetico].

- a. Si deve scegliere tra un modello Decision Tree e uno lineare per risolvere un problema in cui sapete che la funzione target prevede un “and” tra le variabili (booleane) di ingresso. Un collega vi suggerisce il Decision Tree raffigurando il potenziale limite del modello lineare. Siete d'accordo?
- b. La complessità della SVM è legata al margine?
- c. Vi chiedono di giudicare una procedura per stabilire quando terminare il training (assunto iterativo) di un modello. Si propone in particolare di utilizzare un data set separato dal training set per decidere la terminazione in base al minimo errore raggiunto su quel data set e si suggerisce che per ottenere un modello con la miglior capacità di generalizzazione sarebbe utile valutare in base alla partizione dei dati del test set. Che cosa rispondete in merito al criterio di terminazione sulla base dell'errore e in merito all'intero processo così come proposto? Siete cioè d'accordo?
- d. Dopo che avete calcolato la soluzione (valori alfa) di una SVM hard-margin, vengono aggiunti dei punti nel training set. Cambia la soluzione? Rispondere Si o No o dire in che casi sia vero o falso (e come cambia o perché non cambia).

3 - La ridge regression (Tikhonov) e la LBE polinomiale (di grado M).

Quali tra queste affermazioni sono vere? (vanno indicate tutte quelle vere ma affermazioni indicate come vere che non lo sono incidono negativamente sul punteggio dell'esercizio).

1. In una LBE polinomiale, il grado M regola la complessità del modello
2. Lambda nella ridge regression non regola la complessità del modello
3. Posso mettere M alto e poi usare lambda per controllare l'over-fitting
4. Posso mettere M alto e poi usare lambda per controllare l'under-fitting
5. Se ho M alto il termine con lambda non serve
6. Se ho M basso il termine con lambda può non servire
7. Il parametro lambda influenza il livello di regolarizzazione
8. Più alto è il grado M del polinomio più è bene utilizzare lambda bassi
9. Scegliere sia M che lambda non permette di trovare un bilanciamento tra under- e over-fitting
10. Posso usare la ridge regression anche per LBE di tipo diverso da quelle polinomiali
11. La tecnica di Tikhonov con il lambda si usa solo per problemi di regressione
12. C'è un limite al grado M del polinomio da usare
13. Con pochi dati usare un polinomio di grado M elevato può richiedere un valore di lambda maggiore

4 - Discutere le seguenti affermazioni riguardanti possibili scelte di design di modelli di ML, ossia dire se considerarle vere o false e motivare la risposta* (molto brevemente in una o due righe), alla luce dei principi alla base del ML [con testo libero, sintetico]. (*nei compiti risposte non motivate non sono considerate)

1. Una SVM non può risolvere problemi non linearmente separabili.
2. In modello lineare soggetto a una *linear basis expansion* si aumenta la flessibilità del modello.
3. In un DT limitare il numero dei nodi a un massimo fissato garantisce un buon apprendimento.
4. In una discesa di gradiente non fermarsi presto assicura un buon learning.
5. In un DT il numero di nodi non deve essere inferiore alla metà dei dati di training.
6. Per stimare la (futura) capacità predittiva di un modello è bene considerare accuratamente il risultato della model selection senza guardare al risultato in training.

5 - Giudicate quali tra le seguenti affermazioni sono Vere o False (SVM)

1. I vettori di supporto sono dati del problema
2. I vettori di supporto sono rette nell'iperpiano usate per separare i punti
3. I vettori di supporto sono indicati dal designer dell'applicazione
4. Il numero di vettori di supporto può eccedere il numero di dati di test

6 - Scegliere tra le seguenti le opzioni appropriate in merito a underfitting e overfitting.

Nota: Vanno indicate tra le seguenti **tutte** e solo le opzioni appropriate; le opzioni indicate appropriate che non lo sono incidono con un punteggio negativo o più sul totale dell'esercizio.

1. Il linear model tende all'overfitting se il numero di variabili in input è molto alto
2. Il linear model tende all'underfitting se la LBE prevede molti termini
1. Il linear model può tendere all'overfitting se si usa regolarizzazione di Tikhonov con un lambda molto alto
2. La SVM può tendere all'overfitting con C alto
3. La SVM tende all'overfitting con kernel polinomiale a grado basso e C basso
4. La SVM tende all'overfitting con un margine ampio
5. Un Decision Tree tende all'underfitting con troppi nodi
6. Un Decision Tree tende all'overfitting con pochi dati in test
7. Il K-NN tende all'overfitting con K troppo bassi

7. SLT e Polinomio

In una LBE polinomiale di grado "g", cosa vi aspettate con maggiore probabilità in base alla disequazione della Statistical Learning Theory?

Nota: Vanno indicate **tutte** e solo le opzioni appropriate, ma opzioni indicate come appropriate che non lo sono incidono con un punteggio negativo sul totale dell'esercizio.

1. L'aumento di g tende a far aumentare la VC-confidence (a pari valore degli altri termini e valori presenti)
2. La diminuzione di g tende a ridurre la VC-dim
3. La diminuzione di g fa diminuire il valore di λ (elle)
4. Aumentare g tende a ridurre il R_{emp}
5. La somma dei due termini R_{emp} e VC-confidence aumenta sempre con l'aumentare di g
6. Si può generalizzare bene quando il valore di g permette un bilanciamento tra R_{emp} e VC-confidence