

Introduzione all'apprendimento automatico

Matteo Mazzaretto

2024/2025

Indice

1	Prima parte	1
1.1	Si descrivano nel modo più accurato possibile i concetti di bias e variance, il loro rapporto e come nella pratica possano essere affrontati e ridotti i problemi. A tal fine si riportino anche esempi concreti che aiutino a chiarire i diversi aspetti coinvolti	1
1.2	Si descriva in modo accurato il modello di logistic regression, le sue principali caratteristiche ed il contributo dei diversi elementi presenti nella funzione di costo. Si riporti inoltre una comparazione con il modello di classificazione lineare, evidenziando elementi in comune e differenze principali. Infine, si descriva chiaramente la procedura di addestramento mediante l'applicazione di gradient descent	2
1.3	Si descriva in modo dettagliato il modello di logistic regression (con regolarizzazione), le sue principali caratteristiche ed il contributo dei diversi elementi presenti nella funzione di costo. Si riporti infine una descrizione accurata delle differenze di tale modello rispetto ad un semplice classificatore lineare, anche mediante esempi qualitativi	2
1.4	Cosa si intende per inductive bias? Se ne riporti una definizione e si discutano le maggiori implicazioni in termini pratici (si tenga primariamente conto degli algoritmi di machine learning visti a lezione). Si introduca infine il concetto di bias-variance tradeoff	2
1.5	Spiegare in dettaglio gli elementi fondamentali del perceptron e, più in generale, delle reti neurali. Si riporti inoltre una breve descrizione di come tale modello possa essere esteso mediante la realizzazione di un'architettura a più strati, fornendo un esempio che evidenzi le differenze/vantaggi di tale architettura	2
1.6	Si descriva dettagliatamente la procedura di crossvalidation, motivandone scopo ed utilità, e fornendo una chiara descrizione della (corretta) procedura di addestramento di un qualunque sistema di machine learning. Si descrivano inoltre i concetti di true error ed empirical error e se ne evidenzino le relazioni con la procedura di cross-validation	3
1.7	Quali sono i principali paradigmi del machine learning? Se ne riporti una descrizione sintetica – chiarendo quali siano le principali differenze – con particolare enfasi per il caso del supervised learning. Si distinguano in particolare classificazione e regressione	3

1.8	Cosa si intende per “one learning algorithm hypothesis” e come tale ipotesi si relaziona con le reti neurali artificiali? Si fornisca inoltre una descrizione esaustiva degli elementi/ingredienti principali che permettono la definizione di una rete neurale multistrato	3
1.9	Si descriva dettagliatamente la procedura di model selection (aiutandosi con un esempio concreto) e si fornisca una chiara giustificazione teorica/concettuale a tale procedura	3
1.10	Spiegare in dettaglio gli elementi fondamentali del perceptron e delle reti neurali multistrato. Si riporti inoltre un esempio di rete neurale per la realizzazione della porta logica NAND (indicando i valori dei parametri e la funzione di attivazione prescelta)	3
1.11	Spiegare in dettaglio gli elementi fondamentali del perceptron e, più in generale, delle reti neurali. Si riporti inoltre una breve descrizione di come tale modello possa essere esteso mediante la realizzazione di un’architettura multistrato, fornendo un esempio che evidenzi le differenze ed i vantaggi di tale architettura	4
1.12	Spiegare in dettaglio gli elementi fondamentali del perceptron e, più in generale, delle reti neurali multistrato, illustrando chiaramente le due fasi di feedforward e backpropagation. Si riporti inoltre un esempio di rete neurale per la realizzazione di un semplice operatore logico AND, ed uno per la funzione XOR	4
2	Seconda parte	4
2.1	Spiegare in dettaglio gli elementi fondamentali di SVM; in particolare: 1) la sua interpretazione geometrica, 2) la funzione di costo, 3) le differenze/similitudini con altri modelli di ML. Infine, si introduca brevemente l’estensione di SVM basata sul kernel trick	4
2.2	Si descrivano nel modo più accurato possibile gli alberi di decisione, i loro vantaggi e svantaggi rispetto ad altri modelli (ad es. reti neurali) e si evidenzino il principale inductive bias di tale algoritmo. Si fornisca inoltre un semplice esempio di albero di decisione. Infine, si illustri brevemente l’estensione di tale modello attraverso random forest	4
2.3	Si descriva in modo accurato l’algoritmo k-NN, illustrando il ruolo dei principali iperparametri, i vantaggi e le debolezze del modello nei confronti di altri algoritmi affrontati nel corso, e si evidenzino il principale inductive bias di tale algoritmo	5
3	Appelli	5
3.1	Si descriva accuratamente un esempio di rete neurale per la realizzazione di un operatore logico XOR (indicando valori dei parametri e funzione di attivazione prescelta); si fornisca inoltre l’analoga soluzione utilizzando un albero di decisione, e si discutano pro e contro delle due soluzioni	5
3.2	Si descrivano nel modo più accurato possibile i modelli di linear classification e logistic regression; si evidenzino differenze, vantaggi e svantaggi dell’uno rispetto all’altro. Si descriva infine il processo di apprendimento tramite gradient descent e le rispettive funzioni di costo, motivando in modo adeguato la particolare forma utilizzata in entrambi i casi.	5

3.3	Si descriva accuratamente un esempio di rete neurale per la realizzazione di un operatore logico AND, ed uno per la funzione OR (indicando valori dei parametri e funzione di attivazione prescelta). Si riporti infine un esempio di rete neurale per lo XNOR	5
3.4	Si descrivano nel modo più accurato possibile il modello di regressione lineare e la sua “estensione” al problema di classificazione. Infine, si compari il modello di classificazione lineare con quello di logistic regression, evidenziando le principali differenze	5

1 Prima parte

1.1 Si descrivano nel modo più accurato possibile i concetti di bias e variance, il loro rapporto e come nella pratica possano essere affrontati e ridotti i problemi. A tal fine si riportino anche esempi concreti che aiutino a chiarire i diversi aspetti coinvolti

Il bias rappresenta l'errore sistematico introdotto da un modello nel semplificare troppo il problema, portando a una rappresentazione inaccurata dei dati
L'inductive bias rappresenta l'insieme di ipotesi che un algoritmo assume sulla funzione target per poter generalizzare dai dati di training ai dati nuovi
Infatti, nel caso della regressione lineare, l'inductive bias principale è che l'insieme di input e output può essere rappresentato come una funzione lineare
Ci sono esattamente due tipi:

1. restriction: limita lo spazio delle ipotesi
2. impone l'ordine delle preferenze

Il bias alto causa all'algoritmo la mancanza di relazioni rilevanti fra feature e target, ad esempio un modello di regressione lineare ha un bias alto se i dati reali seguono una relazione quadratica: la sua rigidità lo porterà a fare errori sistematici. La variance misura quanto il modello è sensibile alle variazioni nei dati di training

Un modello ha varianza alta quando è molto sensibile ai cambiamenti nei dati di training: piccole variazioni nei dati possono portare a grandi differenze nelle predizioni

Questo accade quando il modello è troppo complesso e tende a sovradattarsi (overfitting)

Il loro rapporto principale porta al bias-variance tradeoff, ovvero ciò che rappresenta il rapporto tra bias e varianza: un modello con bias alto semplifica troppo il problema (underfitting), mentre un modello con varianza alta si adatta troppo ai dati di training (overfitting)

L'obiettivo è trovare un equilibrio tra i due per minimizzare l'errore totale

La cosa migliore per un algoritmo sarebbe avere bias basso, varianza bassa

Alcuni metodi pratici per bilanciare bias e varianza:

- 1) Ridurre il bias: usare modelli più complessi (es. passare da regressione lineare a polinomiale)
- 2) Ridurre la varianza: usare più dati di training

Diverse tecniche di apprendimento hanno diversi tipi di inductive bias:

1. Regressione Lineare: assume che la relazione tra input e output sia lineare. Se i dati seguono una relazione non lineare, il modello farà errori sistematici
2. Reti Neurali: hanno un bias più flessibile, ma la loro architettura impone certe strutture (es. una rete convoluzionale è adatta a dati con correlazioni spaziali, come immagini)
3. Nearest Neighbors (NN): assume che dati simili abbiano etichette simili, basandosi sulla vicinanza in uno spazio di feature

- 1.2 Si descriva in modo accurato il modello di logistic regression, le sue principali caratteristiche ed il contributo dei diversi elementi presenti nella funzione di costo. Si riporti inoltre una comparazione con il modello di classificazione lineare, evidenziando elementi in comune e differenze principali. Infine, si descriva chiaramente la procedura di addestramento mediante l'applicazione di gradient descent

g

- 1.3 Si descriva in modo dettagliato il modello di logistic regression (con regolarizzazione), le sue principali caratteristiche ed il contributo dei diversi elementi presenti nella funzione di costo. Si riporti infine una descrizione accurata delle differenze di tale modello rispetto ad un semplice classificatore lineare, anche mediante esempi qualitativi

g

- 1.4 Spiegare in dettaglio gli elementi fondamentali del perceptron e, più in generale, delle reti neurali. Si riporti inoltre una breve descrizione di come tale modello possa essere esteso mediante la realizzazione di un'architettura a più strati, fornendo un esempio che evidenzi le differenze/vantaggi di tale architettura

g

- 1.5 Si descriva dettagliatamente la procedura di cross-validation, motivandone scopo ed utilità, e fornendo una chiara descrizione della (corretta) procedura di addestramento di un qualunque sistema di machine learning. Si descrivano inoltre i concetti di true error ed empirical error e se ne evidenzino le relazioni con la procedura di cross-validation

g

- 1.6 Quali sono i principali paradigmi del machine learning? Se ne riporti una descrizione sintetica – chiarendo quali siano le principali differenze – con particolare enfasi per il caso del supervised learning. Si distinguano in particolare classificazione e regressione

g

- 1.7 Cosa si intende per “one learning algorithm hypothesis” e come tale ipotesi si relaziona con le reti neurali artificiali? Si fornisca inoltre una descrizione esaustiva degli elementi/ingredienti principali che permettono la definizione di una rete neurale multistrato

g

- 1.8 Si descriva dettagliatamente la procedura di model selection (aiutandosi con un esempio concreto) e si fornisca una chiara giustificazione teorica/concettuale a tale procedura

g

- 1.9 Spiegare in dettaglio gli elementi fondamentali del perceptron e delle reti neurali multistrato. Si riporti inoltre un esempio di rete neurale per la realizzazione della porta logica NAND (indicando i valori dei parametri e la funzione di attivazione prescelta)

g

- 1.10 Spiegare in dettaglio gli elementi fondamentali del perceptron e, più in generale, delle reti neurali. Si riporti inoltre una breve descrizione di come tale modello possa essere esteso mediante la realizzazione di un’architettura multistrato, fornendo un esempio che evidenzia le differenze ed i vantaggi di tale architettura

g

- 1.11 Spiegare in dettaglio gli elementi fondamentali del perceptron e, più in generale, delle reti neurali multistrato, illustrando chiaramente le due fasi di feed-forward e backpropagation. Si riporti inoltre un esempio di rete neurale per la realizzazione di un semplice operatore logico AND, ed uno per la funzione XOR

g

2 Seconda parte

- 2.1 Spiegare in dettaglio gli elementi fondamentali di SVM; in particolare: 1) la sua interpretazione geometrica, 2) la funzione di costo, 3) le differenze/similitudini con altri modelli di ML. Infine, si introduca brevemente l'estensione di SVM basata sul kernel trick

g

- 2.2 Si descrivano nel modo più accurato possibile gli alberi di decisione, i loro vantaggi e svantaggi rispetto ad altri modelli (ad es. reti neurali) e si evidenzi il principale inductive bias di tale algoritmo. Si fornisca inoltre un semplice esempio di albero di decisione. Infine, si illustri brevemente l'estensione di tale modello attraverso random forest

g

- 2.3 Si descriva in modo accurato l'algoritmo k-NN, illustrando il ruolo dei principali iperparametri, i vantaggi e le debolezze del modello nei confronti di altri algoritmi affrontati nel corso, e si evidenzi il principale inductive bias di tale algoritmo

g

3 Appelli

- 3.1 Si descriva accuratamente un esempio di rete neurale per la realizzazione di un operatore logico XOR (indicando valori dei parametri e funzione di attivazione prescelta); si fornisca inoltre l'analoga soluzione utilizzando un albero di decisione, e si discutano pro e contro delle due soluzioni

g

- 3.2 Si descrivano nel modo più accurato possibile i modelli di linear classification e logistic regression; si evidenzino differenze, vantaggi e svantaggi dell'uno rispetto all'altro. Si descriva infine il processo di apprendimento tramite gradient descent e le rispettive funzioni di costo, motivando in modo adeguato la particolare forma utilizzata in entrambi i casi.

g

- 3.3 Si descriva accuratamente un esempio di rete neurale per la realizzazione di un operatore logico AND, ed uno per la funzione OR (indicando valori dei parametri e funzione di attivazione prescelta). Si riporti infine un esempio di rete neurale per lo XNOR

g

- 3.4 Si descrivano nel modo più accurato possibile il modello di regressione lineare e la sua "estensione" al problema di classificazione. Infine, si confronti il modello di classificazione lineare con quello di logistic regression, evidenziando le principali differenze

g