

Machine Learning & Data Mining

Prova intermedia 4 maggio 2015

1) Concept learning

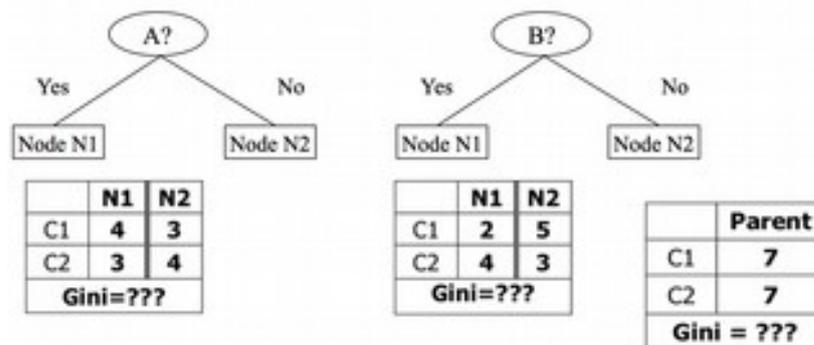
1.a) Descrivere come è possibile rappresentare in modo compatto il version space VS_{HD} rispetto a una ipotesi H e un insieme di training examples D . Dare la definizione formale dei concetti utilizzati.

1.b) Specificare i passi svolti dall'algoritmo "Candidate Elimination" quando esamina un esempio positivo durante un'iterazione del ciclo esterno dell'algoritmo.

2) Alberi di decisione

2.a) Descrive il metodo per decidere il best split che utilizza l'indice di Gini, del quale si dia la definizione.

2.b) Si consideri la seguente distribuzione degli esempi per un nodo "Parent" dell'albero di decisione:



In base al metodo descritto per 2.a) indicare se sia meglio splittare A o B. Motivare dettagliatamente la risposta

2.c) Specificare due condizioni nei dati del training set che possono portare all'overfitting nell'albero di decisione

2.d) Come viene usato il validation set nel metodo "reduced-error pruning" per evitare overfitting?

3) Valutazione di algoritmi

Descrivere brevemente il metodo K-fold Validation

4) Reti Neurali

4.1) Descrivere la versione "Batch" dell'algoritmo Gradient-descent, dove $D = \{<\mathbf{x}, t>\} =$ training examples, \mathbf{x} = vettore di valori di input, t = target output, μ = learning rate, w_i = pesi degli archi.

4.2) Che vantaggi ha la versione incrementale dell'algoritmo rispetto a quella batch?

5) Learning Bayesiano

5.1) Specificare la formula utilizzata dal classificatore Bayesiano naive per classificare una nuova istanza.

5.2) Si consideri il problema "PlayTennis" visto a lezione con il seguente training set:

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Come viene classificata la seguente nuova istanza dal classificatore Bayesiano Naive?

(*Outlook= Rain, Temperature = Cool, Humidity = Normal, Wind = Strong*)

Indicare passaggi che consentono di calcolare la risposta.