

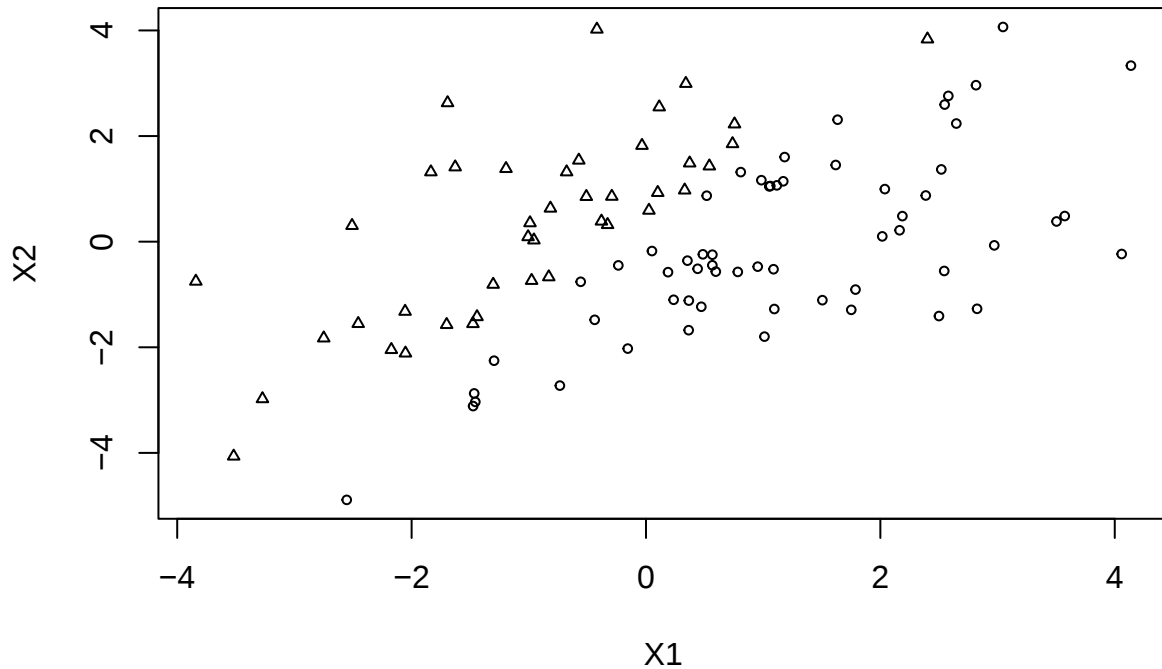
ASML: domande di esercizio sulla classificazione

L.Egidi, N.Torelli

I. Quale delle seguenti osservazioni riguardo la procedura di boosting ADABOOST non è vera:

- (a) Se al primo passo il classificatore debole produce un classificatore perfetto la procedura si può arrestare.
- (b) A ogni classificazione successiva la percentuale di corrette classificazioni è più elevata della precedente.
- (c) A ogni successivo passaggio i dati classificati correttamente avranno meno peso.
- (d) classificatori più performanti in ciascuna iterazione del boosting avranno maggiore peso nella combinazione dei risultati d'insieme.

II. Si consideri il grafico seguente che riporta i dati di un training set per ottenere un algoritmo di classificazione dei punti rappresentati da triangolo o dal cerchio sulla base delle due variabili X_1 e X_2 . Quale delle tecniche di classificazione supervisionata potrebbe condurre a problemi computazionali?



- (a) Gli alberi di classificazione.
- (b) L'analisi discriminante lineare.
- (c) La regressione logistica.
- (d) Le support vector machines.

III. In una classificazione con un glm con legame logistico uso la regola di classificazione per cui la stima di $P(y = 1) > k$. In quale caso porreste $k \neq 0.5$ per classificare nella classe 1?

- (a) Se nel test set ho ottenuto un'accuratezza pari a 1.
- (b) Se c'è perfetta separazione fra le due classi.
- (c) Se si ritiene che sia più grave fare un errore classificando un falso positivo rispetto all'errore in cui si classifica un falso negativo.
- (d) Se classificare correttamente un positivo è altrettanto importante che classificare correttamente un falso negativo.

IV. Considerare due curve ROC di due metodi di classificazione binaria, un GAM e un SVM, che danno, rispettivamente, un' Area Under the Curve (AUC) di 0.65 e 0.73, rispettivamente. Quale delle seguenti affermazioni è vera?

- (a) L'indice di Youden è sicuramente maggiore per la SVM.
- (b) La sensibilità è sempre maggiore per la SVM rispetto al GAM.
- (c) La sensibilità è sempre maggiore della specificità per entrambi i metodi.
- (d) Nessuna delle tre sopra è vera.

V. Disegnare una possibile curva ROC per la domanda IV, usando linea solida e tratteggiata per i due metodi e marcando l'indice di Youden per i due metodi.