**Questions and Exercises**

**Course: Machine Learning and Applications**

**Chapter 9: Combination of Classifiers**

**9.1** Describe the differences and the similarities between bagging và boosting.

*(Mô tả sự khác biệt và tương đồng giữa bagging và boost.)*

**9.2**. What is Boosting? How can boosting improve the classification accuracy.

*(Tăng tốc là gì? Làm thế nào có thể thúc đẩy cải thiện độ chính xác phân loại.)*

*[Trâm]*

*Tăng tốc (Boosting) là một kỹ thuật trong học máy được sử dụng để cải thiện độ chính xác của các mô hình phân loại. Kỹ thuật này hoạt động bằng cách kết hợp nhiều mô hình yếu (weak learner) thành một mô hình mạnh (strong learner), trong đó mỗi mô hình yếu được tạo ra dựa trên việc tập trung vào việc xử lý các điểm dữ liệu khó khăn và bị phân loại sai trong quá trình huấn luyện.*

*Có một số cách thức để thúc đẩy cải thiện độ chính xác phân loại khi sử dụng kỹ thuật tăng tốc:*

*Gán trọng số cho từng điểm dữ liệu*

*Kết hợp các mô hình yếu: Tăng tốc kết hợp nhiều mô hình yếu lại với nhau để tạo thành một mô hình mạnh.*

*Sử dụng câu hỏi đặt tiên nguyên (weak learner): Trong tăng tốc, các mô hình yếu thường là các mô hình đơn giản như câu hỏi đặt tiên nguyên.*

*Điều chỉnh trọng số dựa trên độ chính xác.*

**9.3.** For what purpose AdaBoos algorithmt assigns weights to each tuples in the training set?

*(Vì mục đích gì mà thuật toán AdaBoos gán trọng số cho từng bộ dữ liệu trong tập huấn luyện?)*

*Tạo ra một tập dữ liệu huấn luyện "tập con" mới: Bằng cách gán trọng số cho từng bộ dữ liệu, AdaBoost tạo ra một tập dữ liệu huấn luyện "tập con" mới trong mỗi vòng lặp. Việc gán trọng số cho các bộ dữ liệu này cho phép tập trung vào các điểm dữ liệu khó khăn và bị phân loại sai trong quá trình huấn luyện.*

*Tạo ra các weak classifiers tập trung vào các điểm dữ liệu khó khăn: Các weak classifiers (các mô hình yếu) trong AdaBoost được xây dựng dựa trên trọng số của từng bộ dữ liệu. Bằng cách gán trọng số lớn cho các điểm dữ liệu khó khăn, AdaBoost đảm bảo rằng các weak classifier tiếp theo sẽ tập trung vào việc phân loại đúng các điểm này.*

*Xây dựng một mô hình mạnh từ các weak classifier: AdaBoost kết hợp các weak classifier lại với nhau để tạo ra một mô hình mạnh. Việc gán trọng số cho từng bộ dữ liệu trong quá trình huấn luyện giúp xác định mức độ quan trọng của từng weak classifier trong việc xây dựng mô hình mạnh cuối cùng.*

**9.4** (True/false) In AdaBoost algorithm, tuples that are classified incorrectly will have their weights increased while tuples that are classified correctly will have their weights decreased.

*(Đúng/sai) Trong thuật toán AdaBoost, các bộ được phân loại sai sẽ có trọng số tăng lên trong khi các bộ được phân loại đúng sẽ có trọng số giảm.*

*[Tram]ĐÚng*

**9.5** Describe how to draw the ROC curve to compare the performances of the classifiers.

*Mô tả cách vẽ đường cong ROC để so sánh hiệu suất của các bộ phân loại.*

**9.6** Consider the collection of two dimensional patterns:

X1 = (1, 1, 1), X2 = (1, 2, 1), X3 = (2, 1, 1), X4 = (2, 1.5, 1), X5 = (3, 2, 1),

X6 = (4, 1.5, 2), X7 = (4, 2, 2), X8 = (5, 1.5, 2), X9 = (4.5, 2, 2), X10 = (4, 4, 3),

X11 = (4.5, 4, 3), X12 = (4.5, 5, 3), X13 = (4, 5, 3), X14 = (5, 5, 3),

where each pattern is represented by feature 1, feature 2 and the class label. In *bagging*, the following subsets have been taken from the data set:

S1 = {X1, X3}, S2 = {X2, X4, X5}, S3 = {X6, X7},

S4 = {X8, X9}, S5 = {X10, X11}, S6 = {X12, X13 , X14}

Consider the test pattern (3.5, 2.8). When doing bagging, leave out one subset from each class and classify the test pattern. Consider all possible datasets by leaving out one subset from each classifier and classify the test pattern according to majority vote.

Assume that : 1-NN is used as base classifier.

**9.7** Consider the collection of two dimensional patterns:

X1 = (1, 1, 1), X2 = (1, 2, 1), X3 = (2, 1, 1), X4 = (2, 1.5, 1), X5 = (3, 2, 1),

X6 = (3.5, 2, 2), X7 = (4, 1.5, 2), X8 = (4, 2, 2), X9 = (5, 1.5, 2), X10 = (4.5, 2, 2),

where each pattern is represented by feature 1, feature 2 and the class label. Classify the test pattern (3.5, 1) by using ADABOOST algorithm with the following weak classifiers :

(a) If (f1 < 3) then the pattern belongs to Class 1

else the pattern belongs to Class 2

(b) If (f1 < 3.5) then the pattern belongs to Class 1

else the pattern belongs to Class 2

(c) If (4f1 + 5f2) ≤ 20) then the pattern belongs to Class 1

else the pattern belongs to Class 2.

**9.8** Describe two factors of randomness utilized in the RandomForest algorithm.

*Mô tả hai yếu tố ngẫu nhiên được sử dụng trong thuật toán RandomForest.*