**Fin\_12\_2021.pdf\_HKII\_(2021-2022)**

**Đề bài tập cuối kỳ học kỳ II (2021-2022)  
Môn : Học Máy và Ứng dụng (**Lớp Cao học)

Thời gian: 90 phút   
(Sinh viên được phép tham khảo slides bài giảng)

Đề gồm 3 trang

**1.(0.75 điểm)** Hãy trả lời hai câu hỏi ngắn sau đây.  
**1.1** Giải thuật Lan Truyền Ngược (Backpropagation) làm việc theo cách tiếp cận tối ưu hóa nào sau đây:  
A. Giải thuật Newton B. Dùng nhân tử Lagrange  
C. Suy giảm độ dốc (gradient descent) D. Gia tăng độ dốc (gradient ascent)  
E. Tất cả các câu trên đều sai. (0.25 điểm)  
**→ Chương 7**

**1.2** Phát biểu nào sau đây là không đúng  
A. Bài toán *quy hoạch toàn phương* (Quadratic Programming) liên quan đến máy vector hỗ trợ thuộc loại tối ưu hóa có ràng buộc (constrained optimization). *→ Chương 8*  
B. Trong giải thuật ADABOOST, việc xây dựng các bộ phận lớp thành phần là độc lập với nhau.*→ Chương 9*  
C. Bagging và Rừng Ngẫu nhiên sử dụng cùng chiến lược phiếu bầu: đa số phiếu (majority voting).*→ Chương 9*  
D. Quá trình huấn luyện mạng nơ ron truyền thẳng thường nhanh hơn quá trình huấn luyện mạng nơ ron RBF.*→ Chương 7*  
Đúng. Do RBF được xử lí bằng hồi quy tuyến tính (linear regression), còn Feedforward NN yêu cầu kiểu hồi quy phi tuyến (non - linear regression). Hồi quy phi tuyến sử dụng giải thuật tối ưu để cực tiểu hóa sai số do khiến cho quá trình hồi quy bị chậm đi.

E. Gibb Sampling là một kỹ thuật toán học được dùng để huấn luyện mạng Deep Belief network mà dựa vào máy Boltzmann giới hạn. *→ Chương 12*  
F. Học không có giám sát không cần đến thông tin nhân lớp của các mẫu trong tập dữ liệu huấn luyện. *→ Chương 6*  (0.5 điểm)

**→ Chương 6 + Chương 7 + Chương 8 + Chương 9 + Chương 12**

**2. (0.75 điểm)** Khi xác định cấu hình của mạng nơ ron cho một bài toán phân lớp, căn cứ vào đâu để xác định số nút của tầng nhập và số nút của tầng xuất? Và bằng cách nào xác định số nút thích hợp cho tầng ẩn ?  
**→ Chương 7**

Tầng nhập:

Tầng ẩn:

Tầng xuất: số lượng nhãn dùng để gán cho mẫu dữ liệu.

**3.** **(1 điểm)**Điều chỉnh lại giải thuật *lan truyền ngược* để giải thuật này có thể làm việc được với các nút sử dụng hàm kích hoạt *tanh* thay vì hàm sigmoid. Hàm *tanh* được định nghĩa như sau:

*tanh*(*x*) = (*ex e-x*)/(*ex* + *e-x*) *a^^Hỉỵ^e e* v*e e*

Tức là điều chỉnh lại qui tắc cập nhật trọng số tại tầng xuất và tầng ẩn. Lưu ý đạo hàm hàm *tanh* có tính chất: *tanh’*(*x*) = 1 - [*tanh*(*x*)]2.

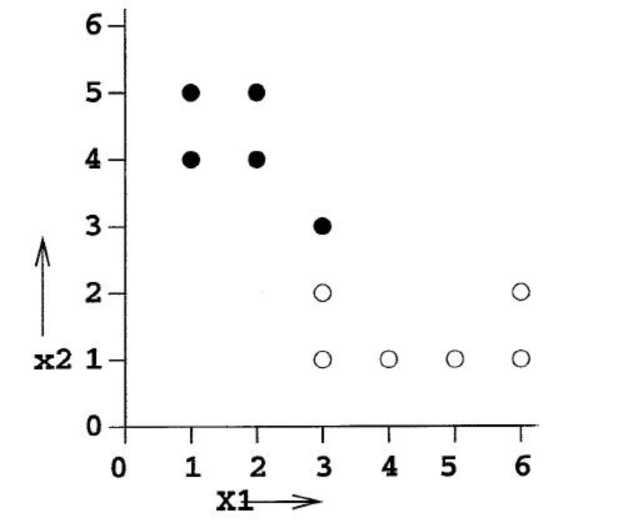
**→ Chương 7 *(cùng đề với câu 1.*** [***Fin\_202.pdf\_HKII\_(2020-2021)***](https://docs.google.com/document/d/1DkbvgkUJOOQcx9eNk-DsMT5MwyhNMvLE/edit#)***)***

**4.** (0.5 điểm) Liệt kê những tham số trong mạng nơ ron RBF mà chúng ta phải xác định khi huấn luyện mạng nơ ron này.

**→ Chương 7**

**5. (2 điểm)  
a.** Cho một tập dữ liệu mà mỗi điểm dữ liệu có hai thuộc tính (x1 and x2) và một nhãn lớp như trong hình vẽ sau. Các điểm thuộc lớp dương ký hiệu bằng dấu chấm đậm và các điểm thuộc lớp âm ký hiệu bằng những hình tròn nhỏ.

Giả sử chúng ta đang dùng SVM tuyến tính để phân lớp. Hãy vẽ 3 đường thẳng biểu diễn siêu phẳng tách có khoảng biên cực đại và hai siêu phẳng hỗ trợ. Tính khoảng biên cực đại. (*1 điểm*)



**b.** Hãy nêu công dụng của hàm kernel đối với máy véc tơ hỗ trợ trong trường hợp dữ liệu không khả tách một cách tuyến tính. (*0.25 điểm*)

**c.** Giải thích ý nghĩa của các biến bù *Ặ* và thông số *C* trong máy vec tơ hỗ trợ với khoảng biên mềm (soft margin) (*0.5 điểm*)

**→ Chương 8 *(cùng đề với câu 5.*** [***Fin\_202.pdf\_HKII\_(2020-2021)***](https://docs.google.com/document/d/1DkbvgkUJOOQcx9eNk-DsMT5MwyhNMvLE/edit#)***)***

**6. (1.25 điểm)** Cho tập mẫu gồm 14 mẫu hai chiều như sau:

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),

Mỗi mẫu được biểu diễn bởi đặc trưng 1, đặc trưng 2 và nhãn lớp. Ta xây dựng một tổ hợp bộ phân lớp bằng kỹ thuật bagging như sau:

Khi tạo bộ phân lớp thứ nhất, bằng cách lấy mẫu, ta tạo ra tập huấn luyện cho bộ phân lớp này: D1 = {X2, X4, X5, X8, X9, X12, X13, X14}

Khi tạo bộ phân lớp thứ hai, bằng cách lấy mẫu, ta tạo ra tập huấn luyện cho bộ phân lớp này: D2 = { X2, X4, X5, X8, X9, X10, X11}

Khi tạo bộ phân lớp thứ ba, bằng cách lấy mẫu, ta tạo ra tập huấn luyện cho bộ phân lớp này: D3 = { X1, X3, X8, X9, X12, X13, X14}

Khi tạo bộ phân lớp thứ tư, bằng cách lấy mẫu, ta tạo ra tập huấn luyện cho bộ phân lớp này: D4 = { X1, X3, X8, X9, X10, X11}

Khi tạo bộ phân lớp thứ năm, bằng cách lấy mẫu, ta tạo ra tập huấn luyện cho bộ phân lớp này: D5 = { X1, X3, X6, X7, X12, X13 , X14}

Khi tạo bộ phân lớp thứ sáu, bằng cách lấy mẫu, ta tạo ra tập huấn luyện cho bộ phân lớp này: D6 = { X1, X3, X6, X7, X10, X11}

Các bộ phân lớp thành phần đều là bộ phân lớp 1-lân cận gần nhất.

Xét mẫu thử (3.5, 2.8). Bộ phân lớp tổ hợp sẽ quyết định mẫu thử thuộc lớp nào ?  
**→ Chương 9**  **(cùng đề với câu 6** [**Fin\_201.pdf\_HKI\_(2020-2021)**](https://docs.google.com/document/d/1O30YVMcYqIh3FvXyK30TFXqQB4cgPsMO/edit#heading=h.k859l5ku3md)**)**

**7.** **(1 điểm)**

a. Nêu những điểm khác biệt giữa hai phương pháp tổ hợp bộ phân lớp: bagging và boosting (*0.5 điểm*).

b.Nếu hai yếu tố ngẫu nhiên được đưa vào giải thuật tổ hợp phân lớp RandomForest (0.5 điểm).   
**→ Chương 9**

**8. (0.75 pt)**

a. Đường cong ROC là gì? (0.5 điểm)

b. Làm cách nào sử dụng đường cong ROC để đánh giá chất lượng của một bộ phân lớp? (0.25 điểm)  
**→ Chương 9**

**9. (1.75 điểm)**

a, Nêu công dụng chính của kiến trúc nhiều tầng trong mạng nơ ron học sâu. (0.25 điểm)  
b. Trong lãnh vực mạng nơ ron học sâu, mô hình sinh (generative model) và mô hình phân biệt (discriminative model) khác biệt nhau như thế nào? (0.5 điểm)  
c. Nêu tên hai giải thuật dùng trong hai giai đoạn của quá trình huấn luyện mạng Deep Belief Network; một giai đoạn là học không giám sát và một giai đoạn là học có giám sát. (0.5 điểm)  
d. Giải thích công dụng của tầng tích chập (convolution) và tầng gộp (pooling) trong mạng nơ ron tích chập. (0.5 điểm)

**→ Chương 12**

***(cùng đề với câu 9.*** [***Fin\_202.pdf\_HKII\_(2020-2021)***](https://docs.google.com/document/d/1DkbvgkUJOOQcx9eNk-DsMT5MwyhNMvLE/edit#)***,   
cùng đề với câu 8.*** [**Fin\_12\_2021.pdf\_HKI\_(2021-2022)**](https://docs.google.com/document/d/1FBqWzRhwBudjJDBuK2_z72pM98EndxA2/edit#heading=h.t37itj2wog4b)***)***