

C. BÀI TẬP THỰC HÀNH CHIỀU 4

1. Download tập dữ liệu “**winequality-white**” và gán vào biến **wineWhite**
2. Với tập dữ liệu **wineWhite** sử dụng nghi thức K-Fold để phân chia tập dữ liệu huấn luyện với $K=100$, sử dụng tham số “Shuffle” để xáo trộn tập dữ liệu trước khi phân chia.
Xác định số lượng phần tử có trong tập test và tập huấn luyện nếu sử dụng nghi thức đánh giá này.
3. Sử dụng giải thuật **Bayes thơ ngây** (hàm sklearn) để dự đoán nhãn cho tập kiểm tra theo nghi thức đánh giá của câu 2 với phân phối xác suất Gaussian
4. Đánh giá độ chính xác cho từng phân lớp dựa vào giá trị dự đoán của câu 3 cho mỗi lần lặp.
Chép lại kết quả độ chính xác cho từng phân lớp của lần lặp cuối đưa vào trong file code.
5. Tính độ chính xác tổng thể cho mỗi lần lặp và độ chính xác tổng thể của trung bình 100 lần lặp.
6. Sử dụng giải thuật Cây quyết định (Decision Tree) ở buổi thực hành số 2 để so sánh hiệu quả phân lớp của giải thuật Bayes thơ ngây và cây quyết định với nghi thức đánh giá hold-out