Với MSSV của mình, sinh viên hãy sử dụng dataset tại trang ứng với chữ số cuối cùng của MSSV.

Ví dụ

Sinh viên có mã số 12345678**9** sẽ dùng dataset ở trang 9

Với dataset và test record của mình, hãy

Câu 1.

1. Tìm một phương pháp để tính khoảng cách d giữa 2 record trong dataset
2. Cho biết kết quả phân loại khi chúng ta áp dụng phương pháp phân loại được quyết định như sau: Gọi x là số hàng chục của MSSV và y = x mod 3. Phương pháp tương ứng với giá trị của y là

y = 0: k-NN

y = 1: Cây quyết định

y = 2: Bayes.

Câu 2. Hãy phân tích ưu khuyết điểm của phương pháp được dùng trong Câu 1a

Câu 3. Cho bảng phân loại như sau

|  |  |  |
| --- | --- | --- |
| Actual\Predicted class | Buy-computer = yes | Buy-computer = no |
| Buy-computer = yes | 5954 | 46 |
| Buy-computer = no | 412 | 3588 |

1. Hãy tính recall, precision, F-measure
2. Nếu chúng ta muốn giảm rủi ro bỏ qua khách hàng tiềm năng, chúng ta nên tập trung cải thiện recall hay precision. Vì sao.