I/ Mô tả dữ liệu

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Trung bình | Độ lẹch chuẩn | Min | Max | Max - Min |
| X | 0.487188 | 0.296494 | 0.002611 | 0.995166 | 0.992555 |
| Y | 0.508382 | 0.272406 | 0.011963 | 0.990540 | 0.978577 |

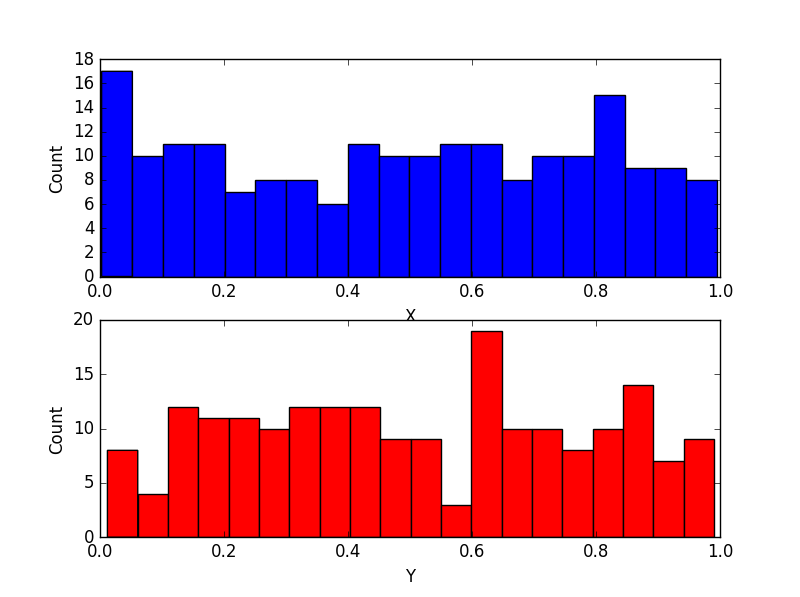


Figure 1 Phân phối dữ liệu của X, Y

Bảng so sánh độ chính xác trung bình (đánh giá theo 10-folds)

|  |  |
| --- | --- |
| **SVM** | **Linear Regression** |
| 76% | 29% |

Dù độ chính xác của phân lớp dùng SVM chưa thực sự cao nhưng vẫn tốt hơn nhiều so với phân lớp sử dụng Linear Regression.

Đánh giá ưu nhược điểm của SVM và Linear Regression

|  |  |  |
| --- | --- | --- |
|  | SVM | Linear Regression |
| Ưu điểm | - Thường thực hiện phân lớp tốt trên các lớp rời rạc.  - Sử dụng kernel để ánh xạ dữ liệu qua không gian mới giúp việc phân lớp dễ hơn.  - Luôn tìm được global minimum. | - Thường giải quyết tốt bài toán hồi quy: Y = b[0] + b[1]\*X[1] + b[2]\*X[2] + ... + b[n] \* X[n]. Dễ nhận thấy Y liên tục và phụ thuộc vào X = (X[0], ..., X[n])  - Việc học cũng được thực hiện tốt bằng công cụ “Gradient descent”, và thường tìm được global minimum. |
| Nhược điểm | - Không giải quyết được bài toán hồi quy như Linear regression.  Mở rộng: biến thể Suppor vector regression, được đề xuất bởi [Vladimir N. Vapnik](https://en.wikipedia.org/wiki/Vladimir_N._Vapnik" \o "Vladimir N. Vapnik) và đồng sự, dùng để giải quyết bài toán regression. | - Linear regression (LR) giả định các biến đầu vào là độc lập với nhau.  - Linear regression giả định có quan hệ tuyến tính giữa các điểm dữ liệu. Hiểu đơn giản là tồn tại “đường thẳng” đi qua các điểm. - *Điều này dẫn đến kết quả khá tệ trong bộ dữ liệu nêu trên khi giá trị hàm f(X,Y) tuần hoàn theo X, Y (khi X, Y tăng tới ngưỡng nhất định thì hàm f thay đổi đột ngột).* - Có thể giải quyết điều trên bằng cách ánh xạ điểm dữ liệu qua một chiều không gian khác nhưng điều này là vô cùng khó khăn và không có bất kì một quy tắc chung nào cho mọi bộ dữ liệu để tăng độ chính xác.  - Ngoài ra, LR dễ bị ảnh hưởng bởi nhiễu. |