1. Khái niệm học quá khớp?

Máy thể hiện tốt khi huấn luyện nhưng lại cho kết quả kém khi kiểm tra.

1. Học có giám sát?

Chứa cả dữ liệu đầu vào và dữ liệu đầu ra ( gọi là nhãn ), sau đó huấn luyện theo mô hình dữ liệu đã có

1. Dự đoán giá nhà theo diện tích và số phòng đâu là gán nhãn, thuộc phương pháp nào(hồi quy hay phân loại? Vì sao?)

Nhãn là giá nhà, Hồi quy. Vì chúng ta cần thu thập 1 tập dữ liệu …

1. cho ví dụ hồi quy tuyến tính, giải thích vì sao?

Dự đoán tình trạng của một người dựa trên cân nặng chiều cao

1. Đâu là học không có giám sát?

chỉ chứa dữ liệu đầu vào, chưa được gắn thẻ hoặc được phân loại và các yếu tố phổ biến đã được nhân dạng.

1. Ví dụ phân loại ? **ví dụ**, giả sử chúng ta muốn dự đoán khả năng mưa ở một số vùng dựa trên một số thông số. Sau đó, sẽ có hai nhãn mưa và không có mưa theo đó các khu vực khác nhau có thể được phân loại.
2. Hàm mất mát( viết công thức hàm mất mát, tên tiếng anh của nó là gì?)

Hàm mất mát(loss function)

MSE (mean square error) = 1/n\*(tổng từ 1 🡪 n) (y-y mũ )\*\*2

1. Không gian 2 chiều hồi quy tuyến tính có dạng?

Đồ thị hqtt có dạng đường thẳng

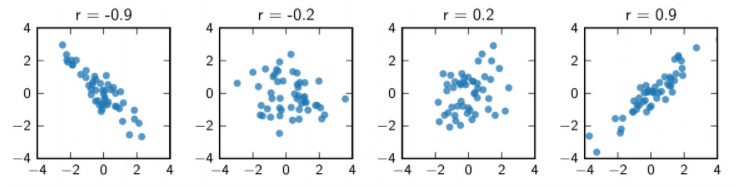
1. Tên tiếng anh của hồi quy tuyến tính? **Linear Regression**
2. Công thức Mối quan hệ giữa 2 biến trong hồi quy tuyến tính?

y = w0 + x1\*w1

1. Cho hệ số tương quan là - 0.8 hình vẽ ở phần correlation coefficient vẽ hình hqtt?

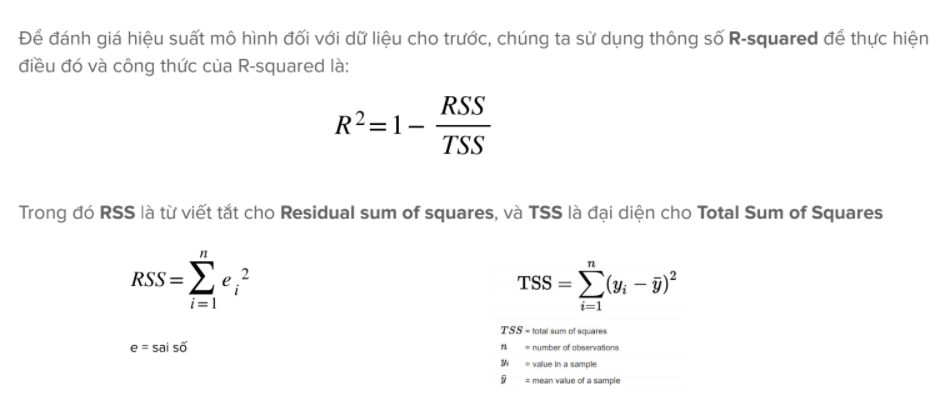
R Từ -1 🡪 0 và 1 🡪 0

Perfect 🡪 weak 🡪 none

Vẽ hình ra 

1. Công thức sai số hqtt, r squared đạt khoảng giá trị nào thì hiệu suất tốt, viết công thức đánh giá mô hình hqtt, yếu điểm hồi quy tuyến tính, hồi quy tuyến tính trong học máy có phải là học có giám sát kh vì sao?

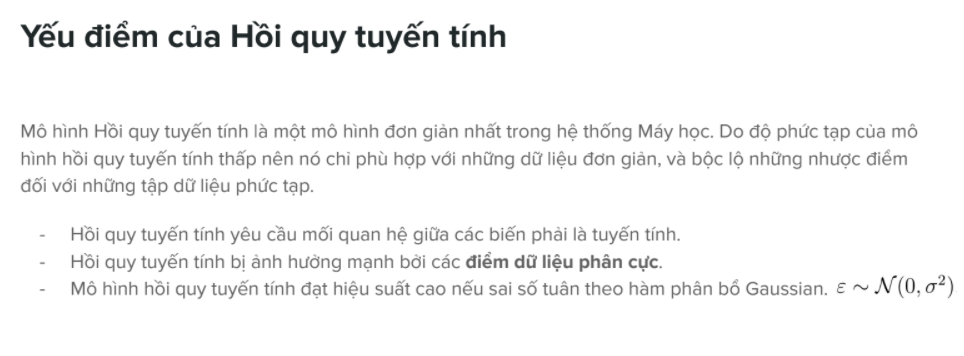
Ct sai số e = y – y mũ



- Giá trị càng lớn tức là gần 1 thì mô hình tuyến tính có hiệu suất tốt và độ biến đổi của dữ liệu là thấp.

- Giá trị càng lớn tức là gần 0 thì mô hình tuyến tính có hiệu suất tồi và dữ liệu của chúng ta gần như xuất hiện ngẫu nhiên.

Hqtt là học có giám sát



1. Yếu điểm của hồi quy tuyến tính phụ thuộc vào yếu tố nào, cách khắc phục?
2. --+kmeans-+++ kmeans thuộc học có giám sát hay kh gs, đặc tính?

K mean là học không có giám sát

Các điểm gần nhau thì có đặc tính, đặc điểm tính chất giống nhau

1. Công thức tính khoảng cách của 2 điểm AB?

Căn (x2 - x1)\*\*2 + (y2 - y1)\*\*2

1. Yếu điểm của kmeans...?

Phải chọn k khởi đầu,

Không đảm bảo đạt được tối ưu toàn cục và kết quả đầu ra phụ thuộc nhiều vào việc chọn k điểm khởi đầu

Khó xác định số cụm thực sự mà không gian dữ liệu có. Do đó có thể phải thử với các giá trị k khác nhau

Khó phát hiện các loại cụm có hình dạng phức tạp và nhất là các dạng cụm không lồi.

1. Tên tiếng anh của ct tính khoảng cách?
2. Rất nhiều câu hỏi logic trong chương này ở giai đoạn huấn luyện số cụm ctroid, các bước huấn luyện kmeans( mục thuật toán chương 3) flowchat, thông số wss, wcss, câu hỏi nào là học có giám sát chọn đáp án đúng, cách tính điểm centroid cho 3 điểm ABC vẽ hình, các bước làm xác định centroid, tính tọa độ centroid( tính xi, yi bằng cách tính trung bình trục tọa độ) neight học có giám sát kh? Cho hình vẽ bài tập x là cc y là cn tìm điểm tại x = 1.8 hỏi cn bao nhiêu? K-NN cho bài toán phân loại(votting điểm K) thuật toán k-NN nhiều bài tập về cái này, vs: về quá trình huấn luyện của kmeans và neighbor... Thuộc bài toán ploai hay hồi quy đầu vào đầu ra, đầu vào là gì( diện tích đất và số phòng) thuật toán k-NN cho hồi quy và phân loại?