1. Linear regression

Dự đoán giá trị đầu ra dựa vào vector đặc trưng của giá trị đầu vào. Dựa vào bộ dữ liệu có sẵn để xây dựng hàm f(x) ~~y.

GT đầu vào là tập hợp dữ liệu độc lập, GT đầu ra là tập hợp dữ liệu phụ thuộc.

Giá trị đầu ra phụ thuộc vào giá trị đầu vào -> tuyến tính.

Có thể nhận nhiều giá trị đầu vào đc -> hồi quy

1. Polymonial regression

Đối với dạng dữ liệu k phân bố theo đường thẳng mà theo đường cong (theo như quan sat tu thuc te), k thể xây dựng đường thẳng để dự đoán cho TH này, nên cần nâng bậc của tham số đầu vào (>1) để model tìm đc fit với bộ dữ liệu hơn. Từ đó cho ra kq dự đoán chính xác hơn.

1. Lasso regression

Gần giống với Linear regression nhưng có them hệ số penalty = λ\*||W||1 (hệ số phạt). Với lamda càng nhỏ thì Lasso càng giống với Linear, lamda càng lớn thì W càng nhỏ, do đó các đặc trưng không quan trọng sẽ tiến về 0 (hoặc bằng 0) trong model. Lamda quá lớn có thể gây underfitting

* Cần chọn hệ số lamda phù hợp với mục đích của bài toán.

1. Least mean square (<https://towardsdatascience.com/linear-regression-using-least-squares-a4c3456e8570>)

Giả sử tìm đc đường thẳng y = m\*x + c

Với mỗi điểm (x0, y0) ta có sai số l = (y0 – y)2 (bình phương để lấy số dương)

= (y0 – m\*x0 + c)^2

Tập hợp tất cả các điểm: đc sai số cho bộ data

L(x) = sum((yi – m\*xi +c)^2) i chạy từ 1 -> N

Bài toán trở thành tìm m và c để L(x) min.