

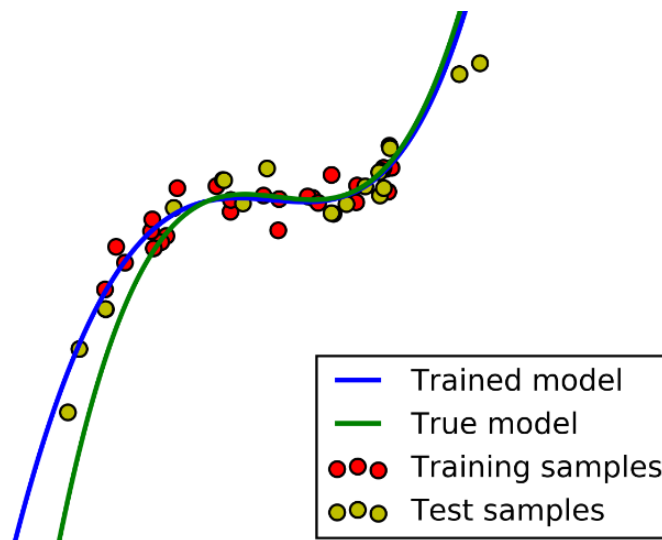
BÀI TẬP: Hồi quy và phân lớp

Bài tập 1: Cho bộ dữ liệu sau. Hãy tìm hàm ước lượng đủ đơn giản và đủ tốt cho bộ dữ liệu này. Vẽ đồ thị minh họa kết quả.

x	y
[7.79667589]	[318.28185696]
[2.79825217]	[20.48143891]
[2.06174503]	[11.97873995]
[4.4713877]	[7.56902114]
[7.20443649]	[224.15497306]
[7.36014312]	[235.04403786]
[4.70688117]	[17.75040067]
[-0.40338389]	[-107.86335911]
[4.72266607]	[1.1140603]
[1.20453709]	[-7.67492972]
[6.07593449]	[87.4263873]
[7.69651292]	[293.22569099]
[3.89733971]	[-11.49557421]
[4.7856351]	[6.4415876]
[-0.59932188]	[-152.88870565]
[4.1507473]	[-4.95755333]
[0.04186784]	[-79.53431819]
[4.89562846]	[34.97246059]
[2.38650347]	[-4.50098315]
[6.42758034]	[95.09276699]

Gợi ý:

- Thực hiện hồi quy đa thức
- Lựa chọn bậc của đa thức để fit với dữ liệu, tránh underfit hoặc overfit



Bài tập 2: Cho bộ dữ liệu gồm giá nhà và quyết định mua nhà tương ứng, trong đó giá nhà có đơn vị là tỷ VNĐ, quyết định mua nhà là có (1) hoặc không (0). Hãy xác định hàm ước lượng cho mô hình dự đoán quyết định mua nhà này.

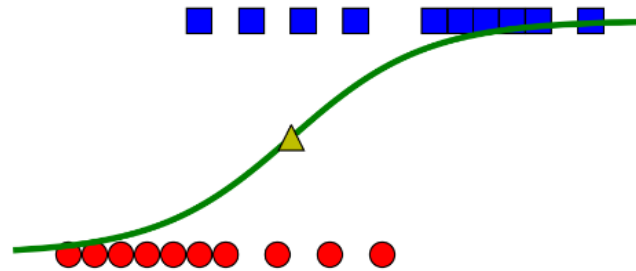
Giá nhà = [0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 1.75, 2.00, 2.25, 2.50, 2.75, 3.00, 3.25, 3.50, 4.00, 4.25, 4.50, 4.75, 5.00, 5.50]

Quyết định mua nhà = [1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0]

Vẽ đồ thị minh họa kết quả.

Gợi ý:

- Sử dụng mô hình hồi quy logistic.



Bài tập 3: Cho 2 tập điểm X_0 và X_1 như sau. Hãy xây dựng mô hình tuyến tính tách 2 tập điểm này. Vẽ đồ thị minh họa kết quả.

X_0	X_1
[1.92306918, 0.775673],	[4.49775285, 1.46545116],
[2.90509186, 1.7997662],	[3.07311718, 3.76212796],
[1.58909188, 1.16143907],	[3.52528933, 1.59844519],
[2.76874122, 1.07663514],	[2.39091046, 2.33431976],
[2.52416203, 1.53020387],	[3.12302646, 1.41945943],
[2.02414192, 1.80692632],	[4.62359547, 2.44921113],
[2.49174878, 2.62943405],	[3.38696098, 2.46494505],
[1.11439322, 2.88348991],	[4.36167918, 1.88637824],
[2.62561276, 2.89077234],	[3.21261415, 2.40558547],
[3.27183166, 0.75454543],	[2.71754956, 2.33882965],
[3.55617919, 0.66250438],	[4.01073111, 0.96947283],
[1.45945603, 2.28222634],	[3.11892586, 4.10522222],
[2.87575608, 2.52637908],	[2.29402636, 2.07905375],
[2.30375703, 2.46497356],	[3.7365999 , 2.41022672],
[1.08925412, 2.01982447],	[4.29699439, 2.79542218],
[4.09096119, -0.08330889],	[4.63617269, 2.00962462],
[1.80350003, 1.91837255],	[2.84870815, 1.77280105],
[1.25827634, 1.8856175],	[4.45368062, 2.44805003],

[2.08532169, 1.79005729], [1.9340609 , 1.09208652]	[4.97052399, 1.86194687], [3.01324102, 1.54377016]
---	---

Bài tập 4: Cho dữ liệu về chiều cao và cân nặng khi thực hiện lấy mẫu của 15 người.

Chiều cao (cm) = [147, 150, 153, 158, 163, 165, 168, 170, 173, 175, 178, 180, 183, 184, 150]

Cân nặng (kg) = [49, 50, 51, 54, 58, 59, 60, 62, 63, 64, 66, 67, 68, 42, 90]

Hãy xây dựng mô hình tuyến tính chỉ ra mối liên hệ giữa chiều cao và cân nặng. Để mô hình được chính xác hơn, có cần phải thực hiện bước tiền xử lý dữ liệu hay không? Nếu có thì thực hiện như thế nào?

Bài tập 5: Cho 2 tập điểm X_0 và X_1 như sau. Hãy xây dựng 3 mô hình tuyến tính tách 2 tập điểm này. Minh họa kết quả trên cùng một đồ thị. Trong 3 mô hình này mô hình nào tốt hơn? Tìm mô hình tách tốt nhất. Ở đây tiêu chí đánh giá là độ rộng của dải phân cách giữa 2 lớp (xem mô hình SVM).

X_0	X_1
[2.37319011, 1.71875981],	[3.42746579, 0.71254431],
[1.51261889, 1.40558943],	[4.24760864, 2.39846497],
[2.4696794 , 2.02144973],	[3.33595491, 1.61731637],
[1.78736889, 1.29380961],	[3.69420104, 1.94273986],
[1.81231157, 1.56119497],	[4.53897645, 2.54957308],
[2.03717355, 1.93397133],	[3.3071994 , 0.19362396],
[1.53790057, 1.87434722],	[4.13924705, 2.09561534],
[2.29312867, 2.76537389],	[4.47383468, 2.41269466],
[1.38805594, 1.86419379],	[4.00512009, 1.89290099],
[1.57279694, 0.90707347]	[4.28205624, 1.79675607]