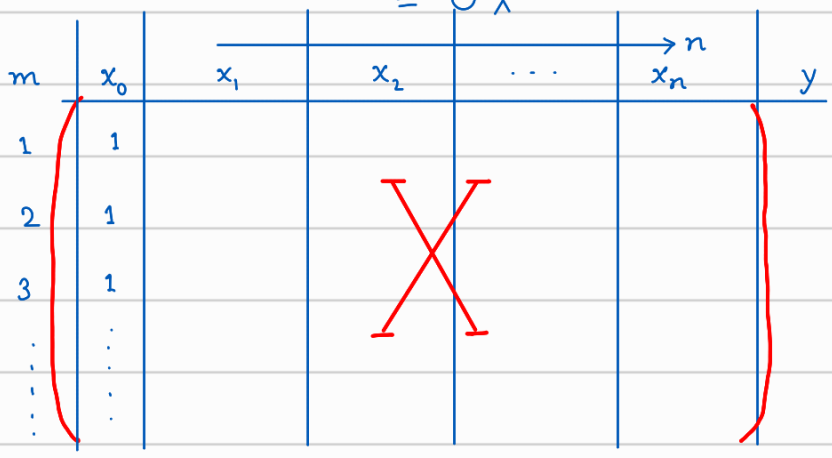
PHY654
Machine Learning in Particle Physics

Monday, 5 August, 2024

Now suppose $h_\theta(x) = \theta_0 x_0 + \theta_1 x_1 + \cdots \theta_n x_n$ where $x_0 = (1, 1, \ldots 1)$

$= \theta_i x_i$

$= \Theta^T X$



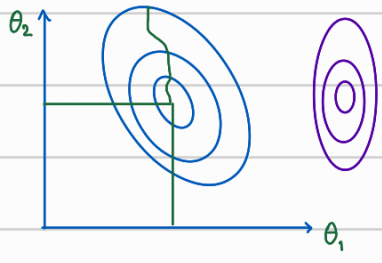
Previously, $J = J(\theta_0, \theta_1)$. Now, $J = J(\vec{\theta})$

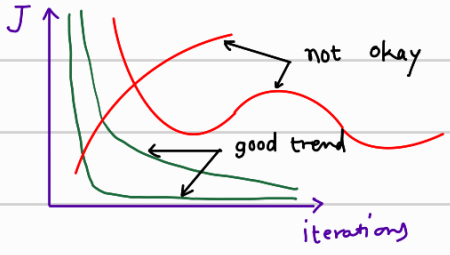
* Feature Scaling :

$x_1 \rightarrow$ size (20 - 400)

$x_2 \rightarrow$ rooms (1 – 10) → grad. desc. has problems here

$x_3 \rightarrow$ floor (1 – 5) → one order of mag. larger.





α might be problematic

↓

learning rate.

☆ Polynomial Regression :

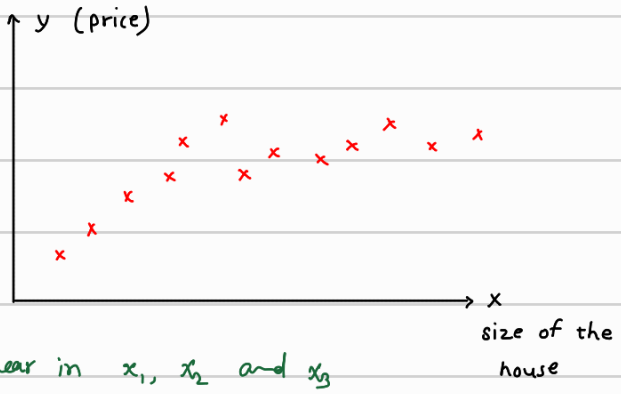
Let's fit a 3rd-order polynomial

$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3$

where $x_1 = x, \quad x_2 = x^2, \quad x_3 = x^3$

(new) features (introduced)

so that this cubic becomes linear in x_1, x_2 and x_3 albeit in three variables now.



Normal Equation method [to find θ_0, θ_1 (or $\vec{\theta}$)]:

set $\frac{dJ}{d\theta_j} = 0$ to find the corresponding parameter (θ_j)

$$\theta = (X^T X)^{-1} X^T Y$$

If n is large (~ 800 and so on) → the normal equation method will run slower.

⚠ Sometimes $X^T X$ may not be invertible

[for example, if $x_2 \rightarrow$ size in sq. ft
and $x_3 \rightarrow$ size in sq. mtr
→ x_2 and x_3 are not linearly independent
hence $X^T X$ is singular
→ solution is to delete one of the features x_2 or x_3, effectively reducing the # of features]

for classification problem, linear regression does not work well.

Logistic Regression (a misnomer, since regression's output is a continuous variable and classification's output is discrete)