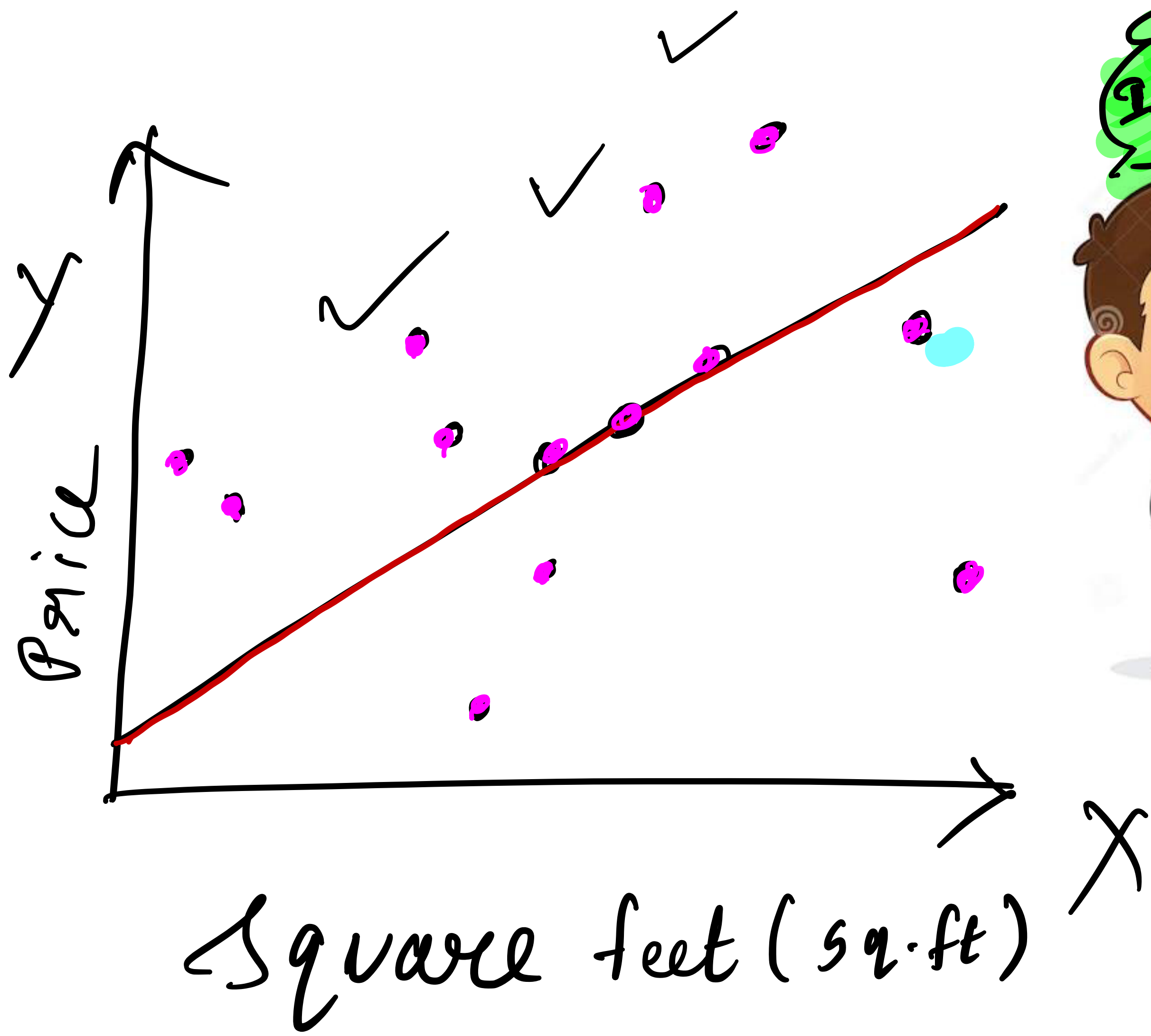
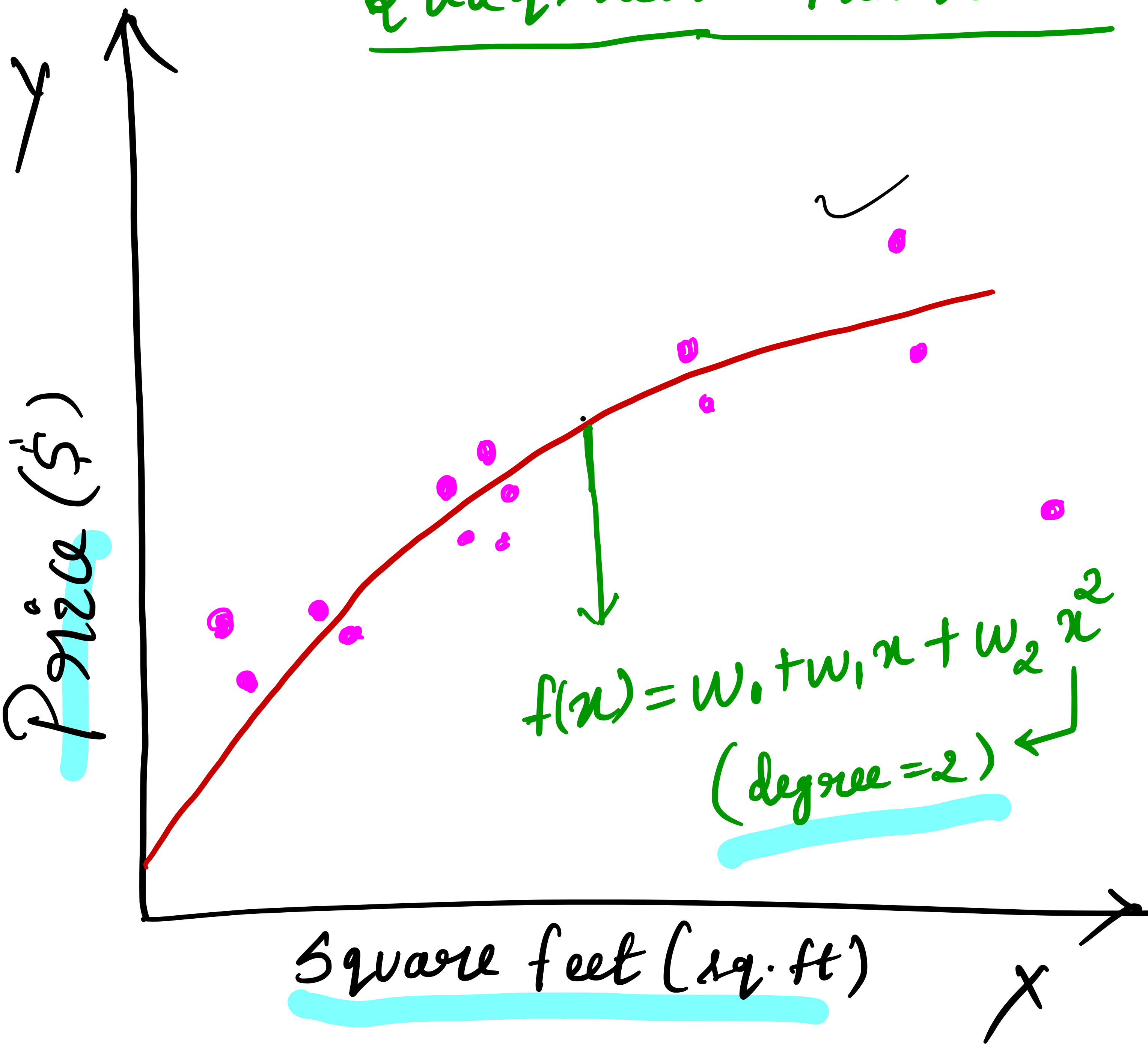
Regression

Polynomial Regression.




May be a quadratic fit

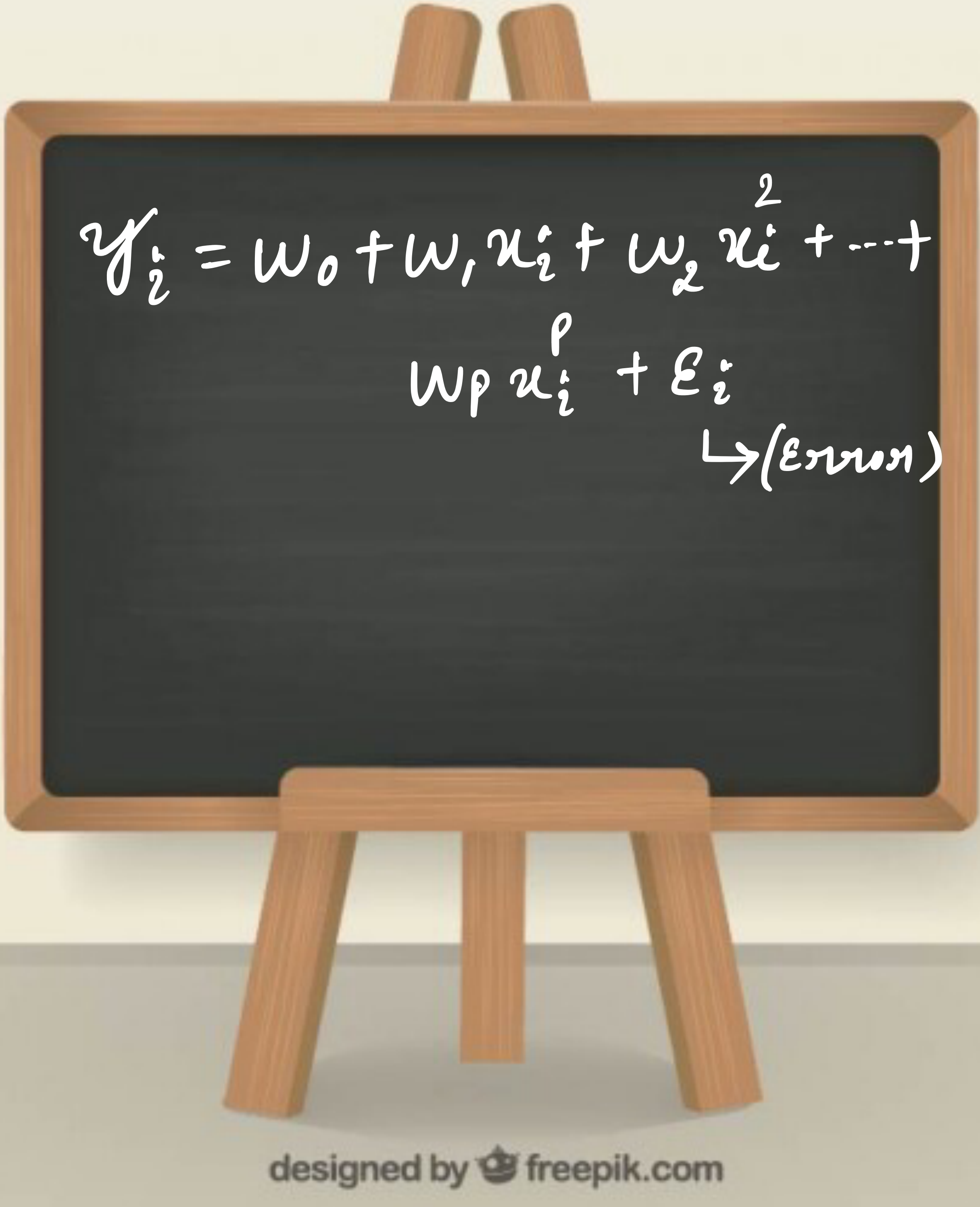
Quadratic function:



$$\boxed{f(x) = w_0 + w_1 x + w_2 x^2 + \cdots + w_p x^p}$$

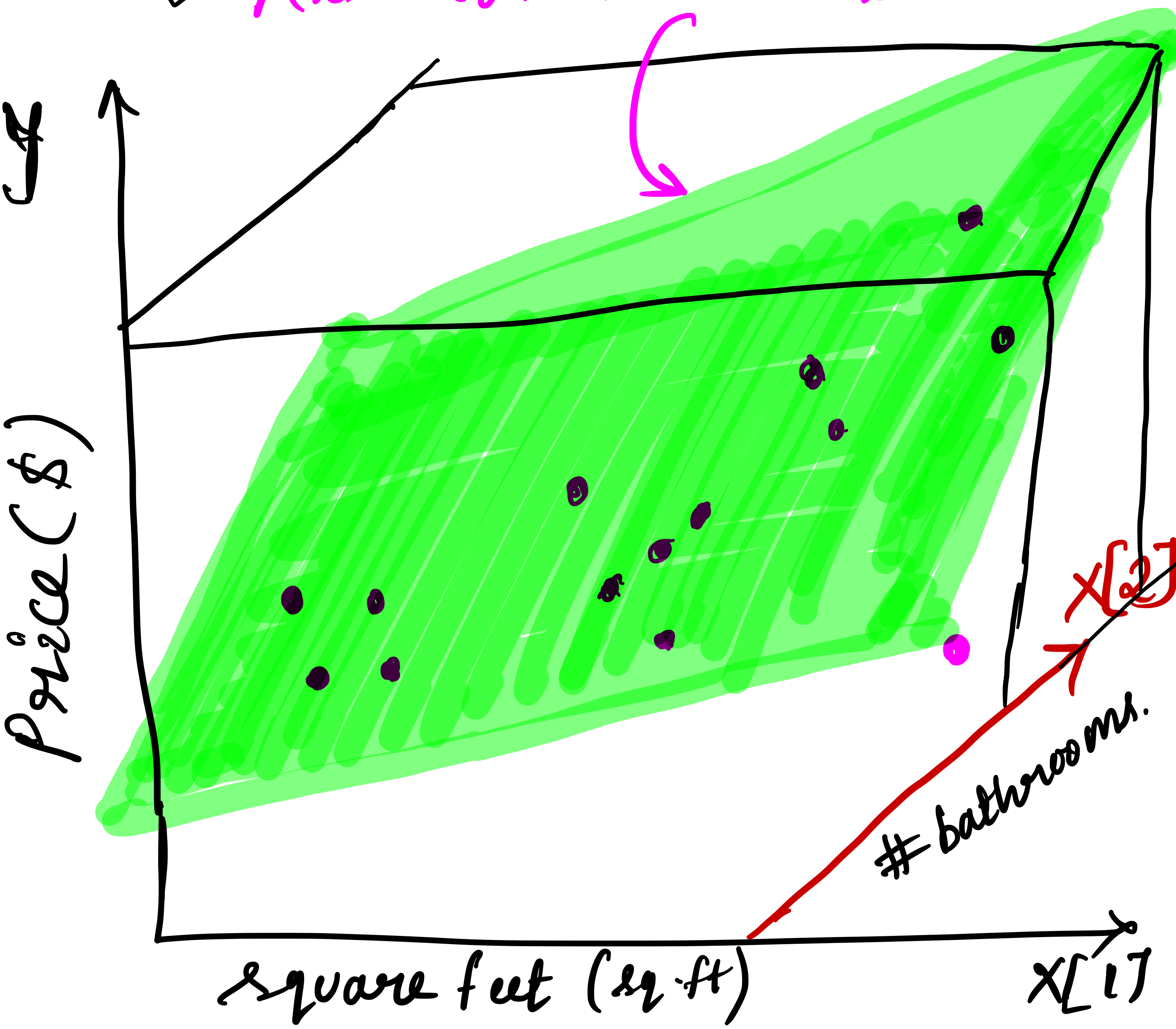
→ Higher order polynomial.

Polynomial regression:



Incorporating multiple inputs:

$f(x) = w_0 + w_1 \, sq.ft + w_2 \# bathroom$



General Notation:

Output = y ✓

Input = x = $(x[1], x[2], \ldots x[d])$

$x[j]$ = j^{th} input

$h_j[x]$ = j^{th} feature

x_i = input of i^{th} datapoint (vector)

$x_i[j]$ = j^{th} input of i .. . . point.

Simple hyperplane:

model:

$$\underline{y_i = w_0 + w_1 x_i[1] + \cdots + w_d x_i[d]} + \varepsilon_i$$

$\varepsilon_i \rightarrow$ noise term

✓ feature 1 = 1

✓ feature 2 = $x[1]$ ---- eg - sq.ft

✓ feature 3 = $x[2]$ ---- eg - #bath

....

✓ feature $d+1$ = $x[d]$ --- eg - lot size.

More generally,

D-dimensional curve:

model:

$$y_i = w_0 h_0(x_i) + w_1 h_1(x_i) + \dots + w_D h_D(x_i) + \varepsilon_i$$

$$= \sum_{j=0}^{D} w_j h_j(x_i) + \varepsilon_i$$

feature 1 = $h_0(x)$ ... e.g. 1

feature 2 = $h_1(x)$ ... e.g. .. $x[1]$ = sq.ft

feature 3 = $h_2(x)$ ... e.g. .. $x[2]$ = #bath

or $\log(x[7])\, x[2]$ =

$\log(\#bed) \times \#bath$

....

feature D+1 = $h_D(x)$ ... some other function of $x[1] \dots x[d]$

Interpreting the co-efficients

simple linear regression:

$$\text{Price} = \hat{w}_0 + \hat{w}_1(\text{square feet})$$

when $\hat{w}_1 = 1$, predicted change in the price for 1 square feet change


How to interpret the co-efficients if we have two variables

Interpreting the Co-efficients.

Two linear features:-

$$\hat{y} = \hat{w}_0 + \hat{w}_1 x[1] + \hat{w}_2 x[2].$$

sq.ft → $x[1]$

bath → $x[2]$

fix. the 1st input i.e number of square feet of the house and think about what's the effect of number of bathrooms on price.

Simply, for the same fixed size house if i increase the number of bathrooms how price is going to be effected.

Interpreting the co-efficients - multiple linear features.

$$\hat{y} = \hat{w}_0 + \hat{w}_1 x[1] + \dots + \hat{w}_j x[j] + \dots + \hat{w}_d x[d]$$

fix

no. of bedrooms

what is the predicted change in price if i add one more bedroom keeping all other constant.

Interpreting the co-efficients polynomial regression -

$$\hat{y} = \hat{w}_0 + \hat{w}_1 x + \dots + \hat{w}_j x^j + \dots + \hat{w}_p x^p$$

Can not hold other features fixed.

Computing the least square fit:

<u>Step-1:</u>

For observation i

$$y_i = \sum_{j=0}^{D} w_j h_j(x_i) + \varepsilon_i$$

$$\boxed{y_i} = \overbrace{\begin{bmatrix} w_0 & w_1 & w_2 & \cdots & w_D \end{bmatrix}}^{W^T} \underbrace{\begin{bmatrix} h_0(x_i) \\ h_1(x_i) \\ h_2(x_i) \\ \vdots \\ h_D(x_i) \end{bmatrix}}_{h(x_i)} + \varepsilon_i$$

$$= w_0 h_0(x_i) + w_1 h_1(x_i) + \cdots + w_D h_D(x_i) + \varepsilon_i$$

$$= W^T h(x_i) + \varepsilon_i$$

$$\boxed{y_i} = \overbrace{\underbrace{\begin{bmatrix} \ & \ & \ & \ & \ \end{bmatrix}}_{h_0(x_i)\ h_1(x_i)\ \cdots}}^{h^T(x_i)} \overbrace{\begin{bmatrix} \ \\ \ \\ \ \\ \ \\ \ \end{bmatrix}}^{w} + \boxed{\varepsilon_i}$$

Exactly same as previous

$$= h_0(x_i)w_0 + h_1(x_i)w_1 + \cdots + h_D(x_i)w_D + \varepsilon_i$$

↳ equation for single observation

?

What about all observation

Rewrite matrix notation for all observations together!

$$\begin{bmatrix} y_1\\ y_2\\ y_3\\ \vdots \\ y_N\end{bmatrix} = \begin{array}{c} \mathbf{H} \\ \begin{array}{r} h^T(x_1) \to \\ \\ \\ \\ \end{array}\begin{bmatrix} h_0(x_1) & h_1(x_1) & \cdots & h_D(x_1)\\ h_0(x_2) & h_1(x_2) & \cdots & h_D(x_2)\\ \vdots & \vdots & & \vdots \\ h_0(x_n) & h_1(x_n) & \cdots & h_D(x_n)\end{bmatrix}\end{array} \begin{array}{c} \mathbf{w} \\ \begin{bmatrix} w_0\\ w_1\\ w_2\\ w_3\\ \vdots\\ w_D\end{bmatrix}\end{array} + \begin{array}{c} \boldsymbol{\varepsilon} \\ \begin{bmatrix} \varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\\ \vdots \end{bmatrix}\end{array}$$

$$\to \quad \boxed{y = Hw + \varepsilon}$$

Residual sum of square (RSS) for linear regression →

$$RSS(w_0, w_1) = \sum_{i=1}^{N} \left(y_i - [w_0 + w_1 x_i]\right)^2$$

RSS for multiple regression

$$RSS(w) = \sum_{i=1}^{N} \left(y_i - \hat{y}_i \right)^2$$

$$= \sum_{i=1}^{N} \left(y_i - h^T(x_i) w \right)^2$$

$$\boxed{\hat{y}_i} = \underbrace{\begin{bmatrix} h_0(x_i) & h_1(x_i) & \cdots & h_D(x_i) \end{bmatrix}}_{h^T(x_i)} \; \underbrace{\begin{bmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_D \end{bmatrix}}_{w}$$

RSS in matrix notation!

$$RSS(w) = \sum_{i=1}^{N} \left(y_i - h(x_i)^T w\right)^2$$

$$= (Y - Hw)^T (Y - \underbrace{Hw}_{\hat{y}})$$

$$\approx (y - \hat{y})$$

$$\boxed{\hat{y} = Hw}$$

$$= (Y - \overset{\hat{y}}{HW})^T \cdot (Y - \overset{\hat{y}}{HW}) \checkmark$$

$$\begin{bmatrix} \text{residual}_1 & \cdots & \text{residual}_N \end{bmatrix} \quad \begin{bmatrix} \text{residual}_1 \\ \text{residual}_2 \\ \vdots \\ \text{residual}_N \end{bmatrix}$$

$$= (\text{residual}_1)^2 + (\text{residual}_2)^2 + \cdots + (\text{residual}_N)^2$$

$$= \sum_{i=1}^{N} \text{residual}_i^2$$

$$\triangleq RSS(W)$$

Gradient of RSS:

$$\nabla RSS(w) = \nabla\left[(y-Hw)^T(y-Hw)\right]$$

$$= -2H^T(y-Hw)$$

$$\nabla RSS(w) = -2H^T(y-Hw) = 0$$

$$\frac{d}{dw}(y-hw)(y-hw)$$

$$\frac{d}{dw}(y-hw)^2 = 2(y-hw)'(-h)$$

$$= -2h(y-hw)$$

$$\nabla RSS(w) = -2H^T(y - Hw) = \underline{0}$$

Solve for w:

$$-2H^Ty + 2H^THw = 0$$

$$\Rightarrow H^THw = H^Ty$$

$$\Rightarrow \underbrace{(H^TH)^{-1} \cdot H^TH}_{I}\,\hat{w} = (H^TH)^{-1}H^Ty$$

$A^{-1} \cdot A = I$

$IW = W$

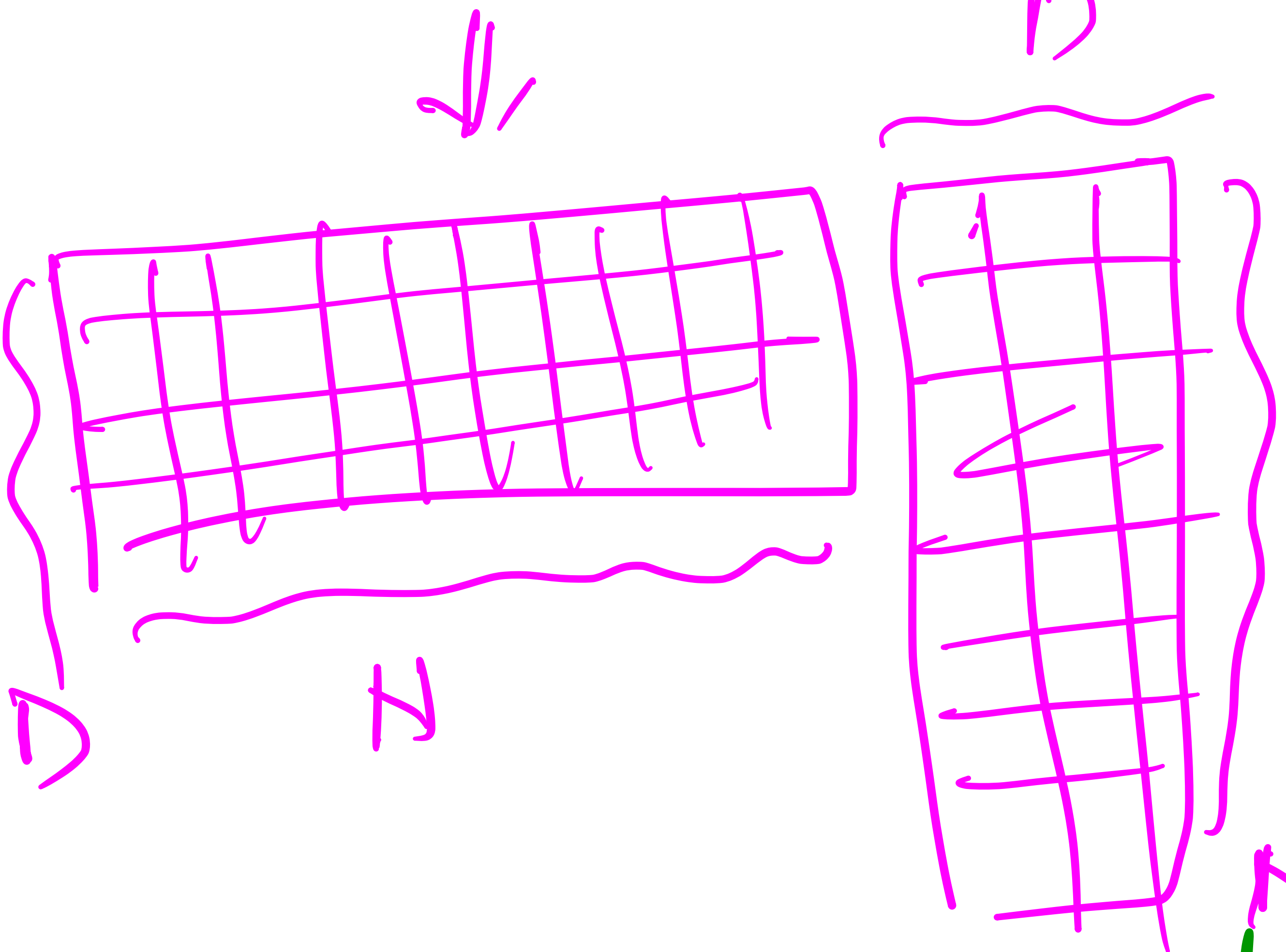
$I \cdot W = W$

$$\hat{w} = (H^TH)^{-1}H^Ty$$

Closed form solution.

Closed form solution:

$$\hat{w} = \underbrace{(H^T H)}^{-1} H^T y$$



No. of independent obss.

$$= \boxed{D \times D} \ \#\text{features}$$

Invertible if:

$$N > D$$

Con, vexity of inverse:

$$O(D^3)$$

$$\hat{w} = \underbrace{(H^T H)}^{-1} H^T y.$$