Recall Task: Predicting house prices

How much is my house worth?

How much is my house worth?

Look at recent sales in my neighborhood

- How much did they sell for?

Regression fundamentals: data, model, task

Data

input *output*

($\mathbf{x}_1$ = sq.ft., $\mathbf{y}_1$ = \$)

($\mathbf{x}_2$ = sq.ft., $\mathbf{y}_2$ = \$)

($\mathbf{x}_3$ = sq.ft., $\mathbf{y}_3$ = \$)

($\mathbf{x}_4$ = sq.ft., $\mathbf{y}_4$ = \$)

($\mathbf{x}_5$ = sq.ft., $\mathbf{y}_5$ = \$)

⋮

Data

input *output*

$(x_1$ = sq.ft., y_1 = \$)

$(x_2$ = sq.ft., y_2 = \$)

$(x_3$ = sq.ft., y_3 = \$)

$(x_4$ = sq.ft., y_4 = \$)

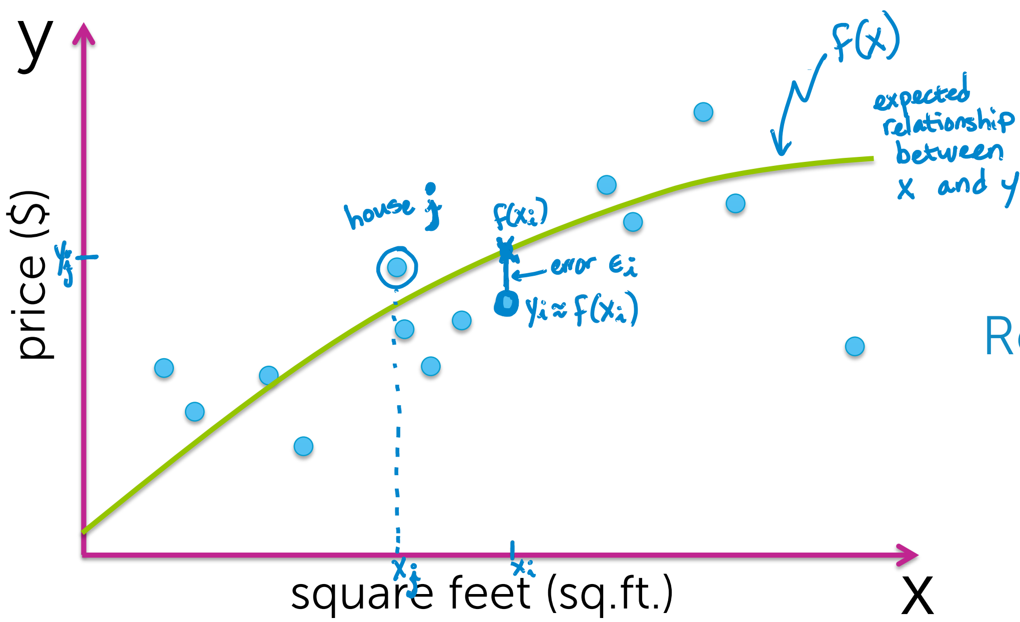
$(x_5$ = sq.ft., y_5 = \$)

⋮

Input vs. Output:

- y is the quantity of interest
- assume y can be predicted from x

Model –
How we *assume* the world works



Regression model:

$y_i = f(x_i) + \epsilon_i$

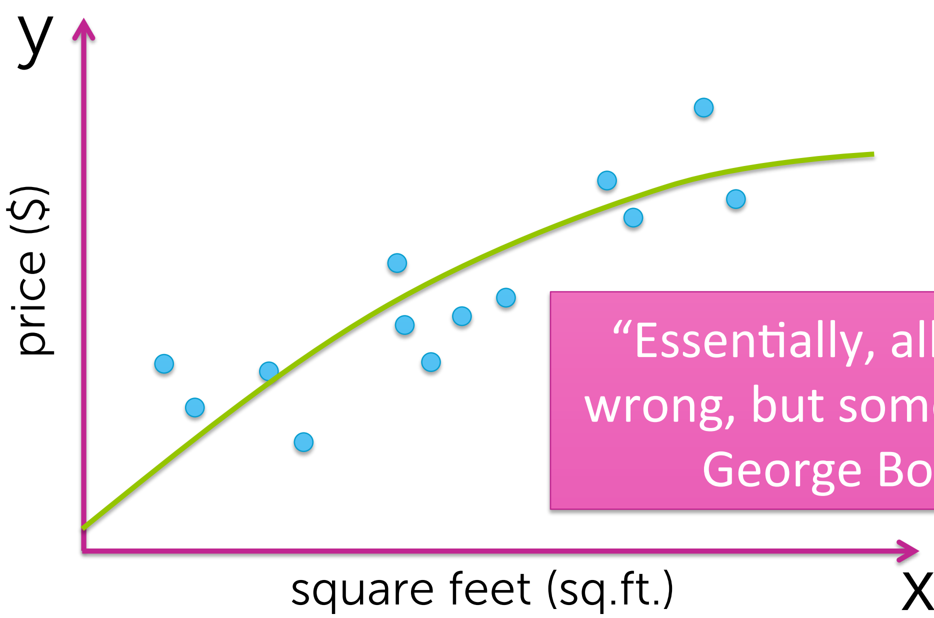
$E[\epsilon_i] = 0$ ← equally likely that error is + or −

↑ expected value

⇓

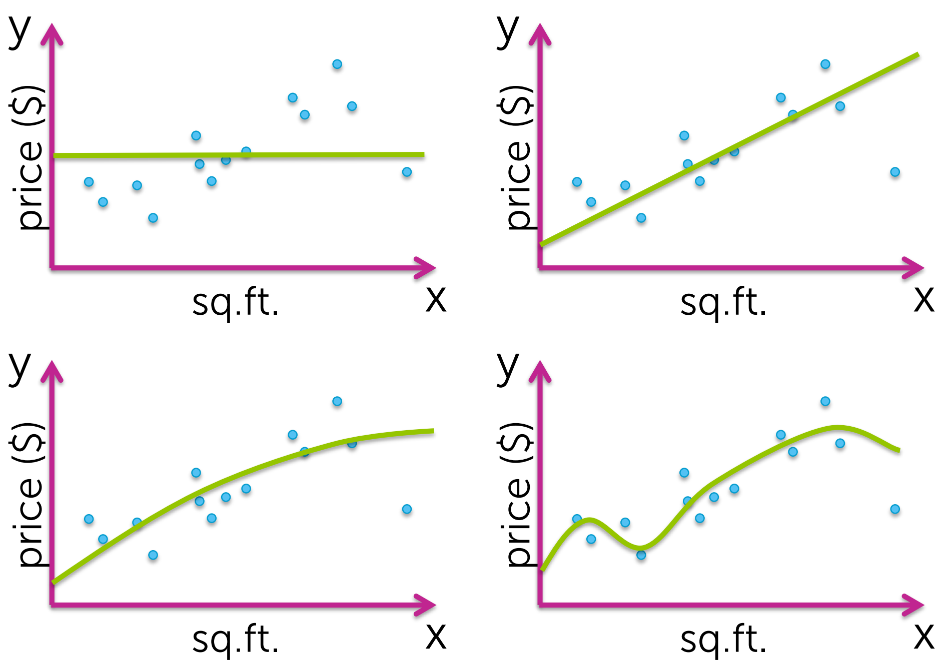
y_i is equally likely to be above or below $f(x_i)$

Model –
How we *assume* the world works

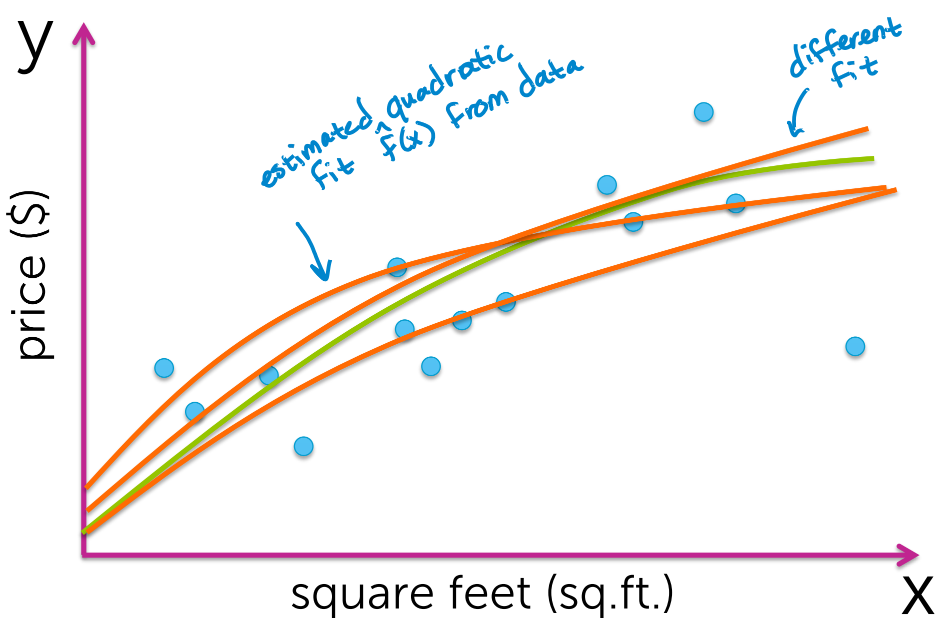


"Essentially, all models are wrong, but some are useful." George Box, 1987.

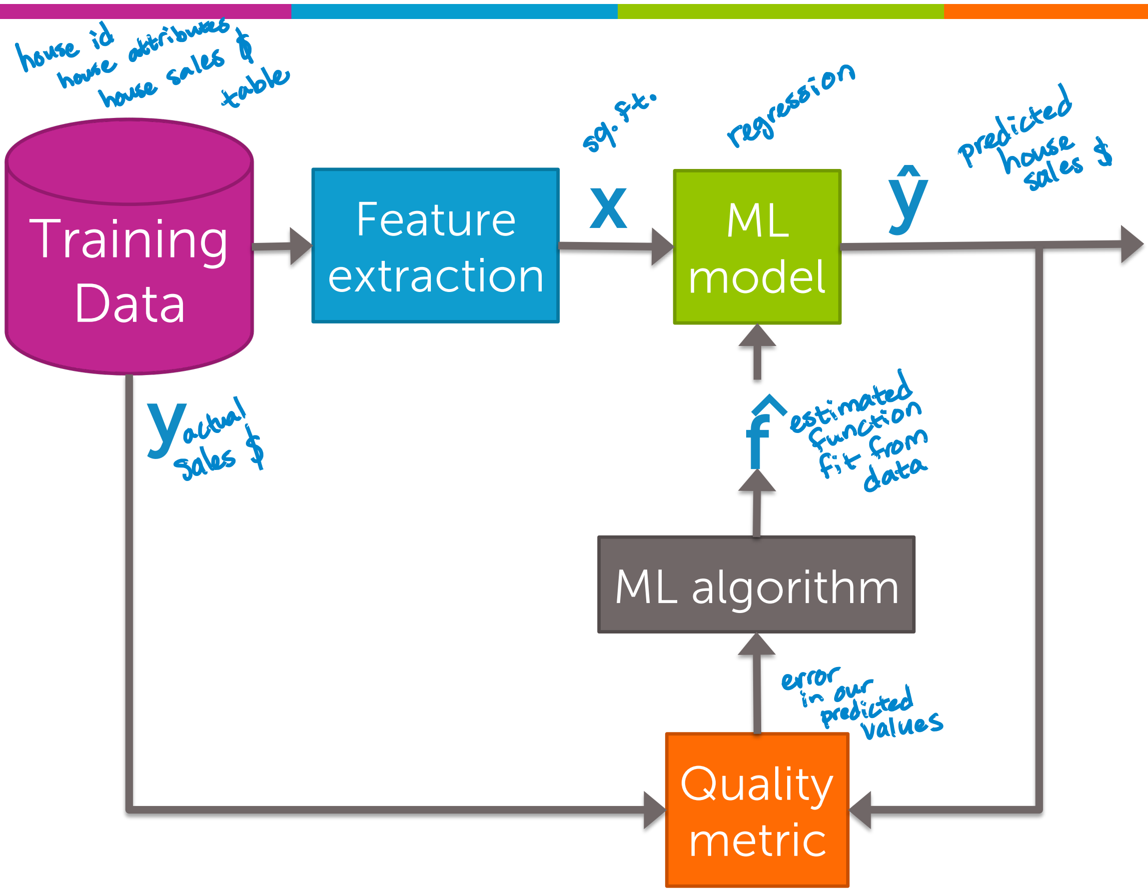
Task 1–
Which model f(x)?



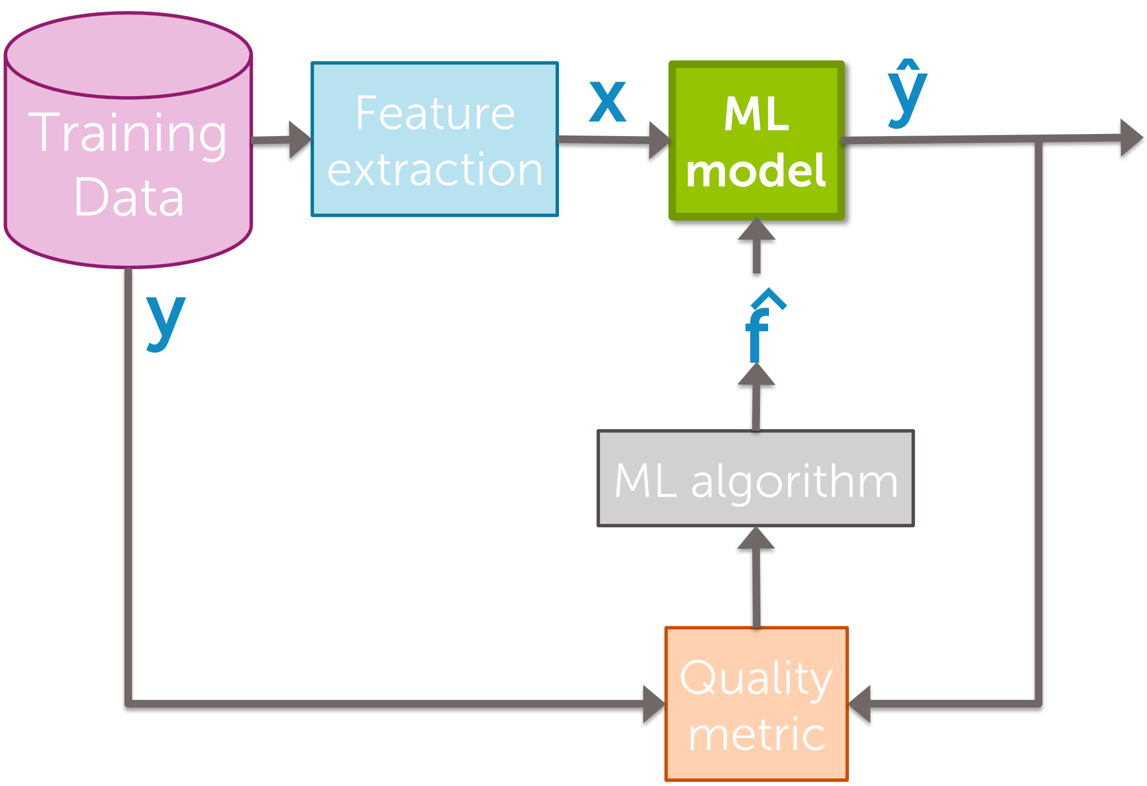
Task 2 – For a given model f(x), estimate function $\hat{f}(x)$ from data



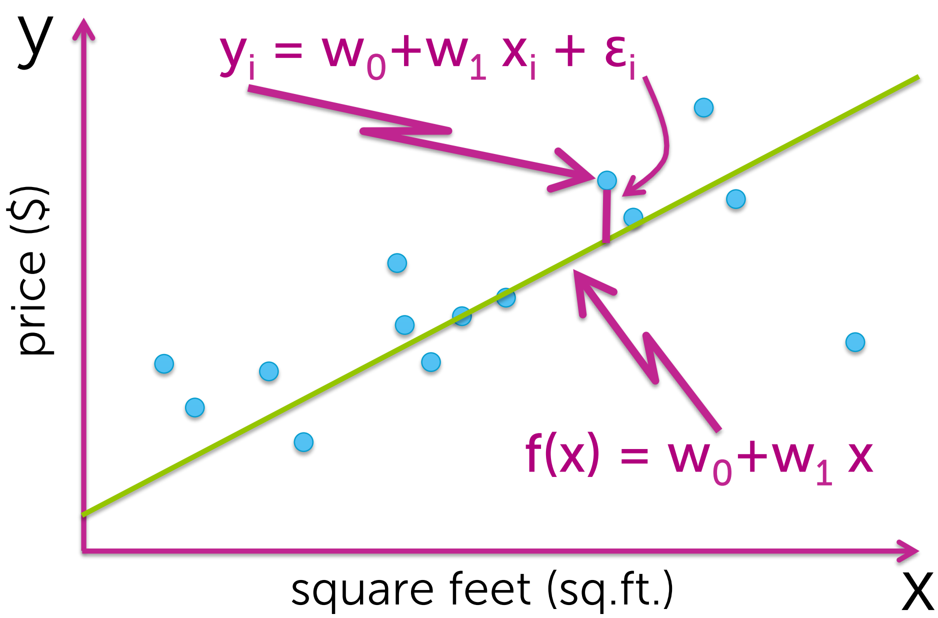
Assume model f(x) is a quadratic function


house id
house attributes
house sales $
table
Training Data
Feature extraction
sq. ft.
x
regression
ML model
ŷ
predicted house sales $
y actual sales $
f̂
estimated function fit from data
ML algorithm
error in our predicted values
Quality metric

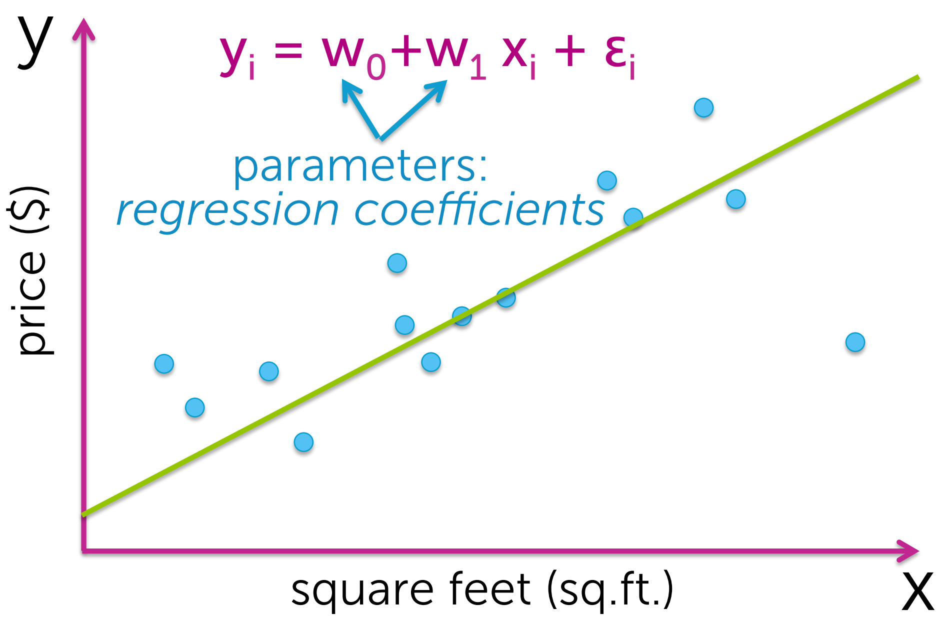
Simple linear regression


Training Data
Feature extraction
x
ML model
ŷ
y
f̂
ML algorithm
Quality metric

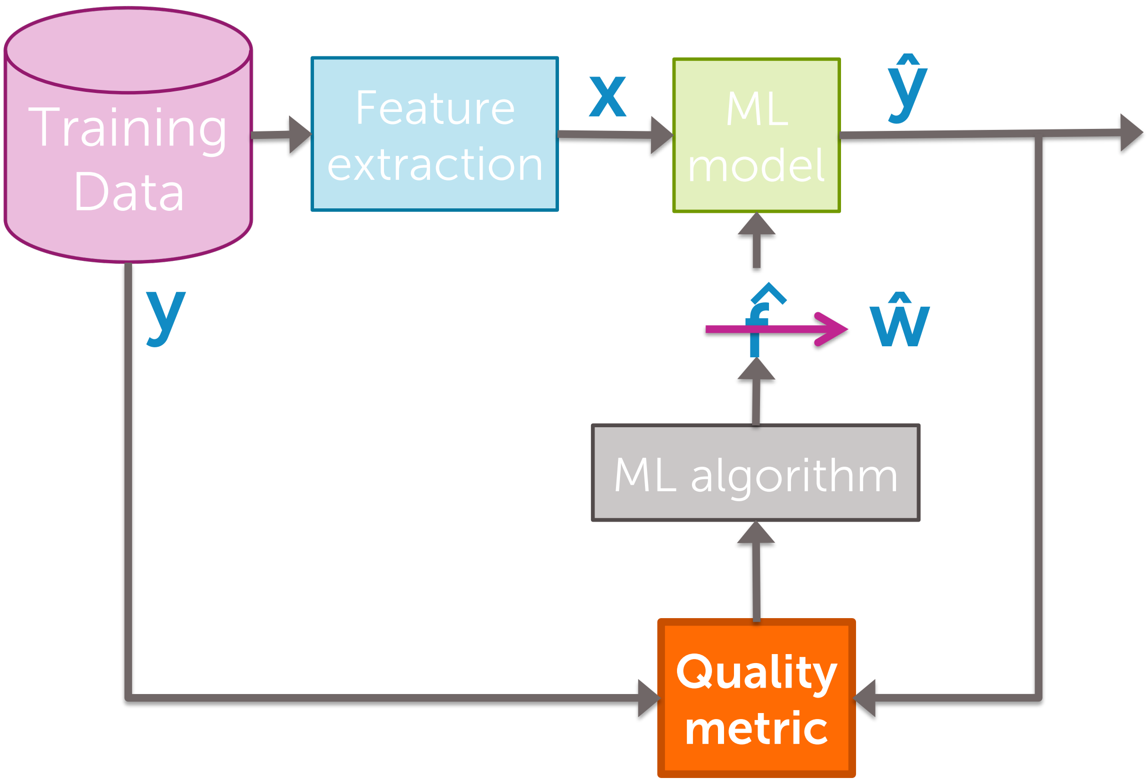
Simple linear regression model



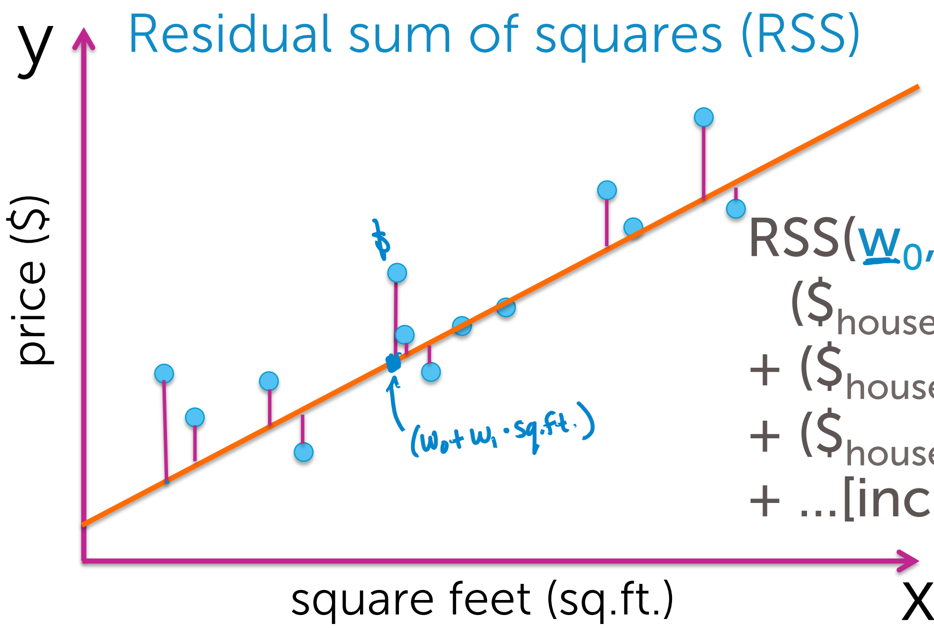
Simple linear regression model



Fitting a line to data


Training Data
Feature extraction
x
ML model
ŷ
y
f̂
ŵ
ML algorithm
Quality metric

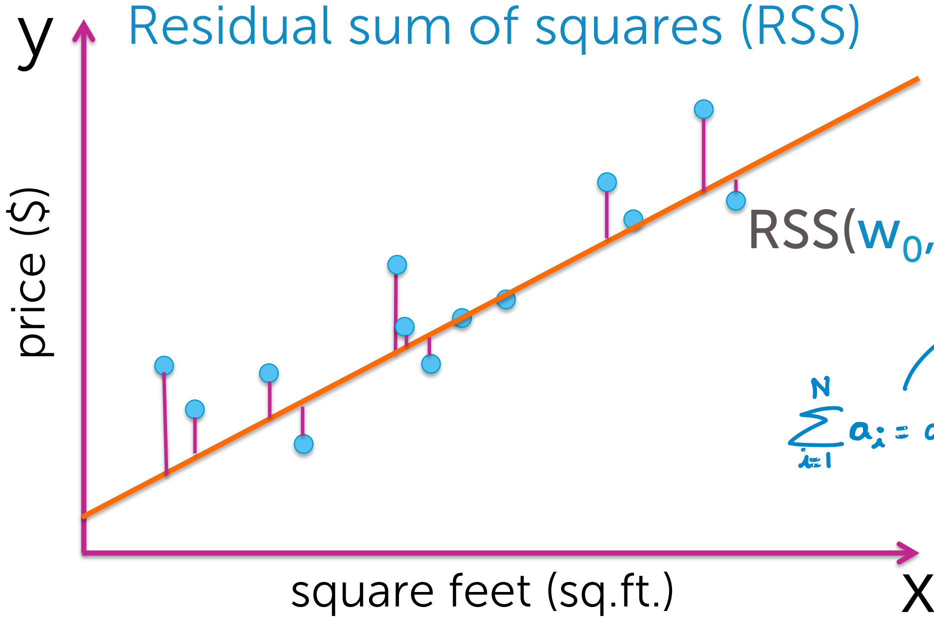
"Cost" of using a given line



RSS($\underline{w_0}$,$\underline{w_1}$) =
($\$_{\text{house 1}}$-[$w_0$+$w_1$sq.ft.$_{\text{house 1}}$])^2
+ ($\$_{\text{house 2}}$-[$w_0$+$w_1$sq.ft.$_{\text{house 2}}$])^2
+ ($\$_{\text{house 3}}$-[$w_0$+$w_1$sq.ft.$_{\text{house 3}}$])^2
+ ...[include all training houses]

"Cost" of using a given line

Residual sum of squares (RSS)

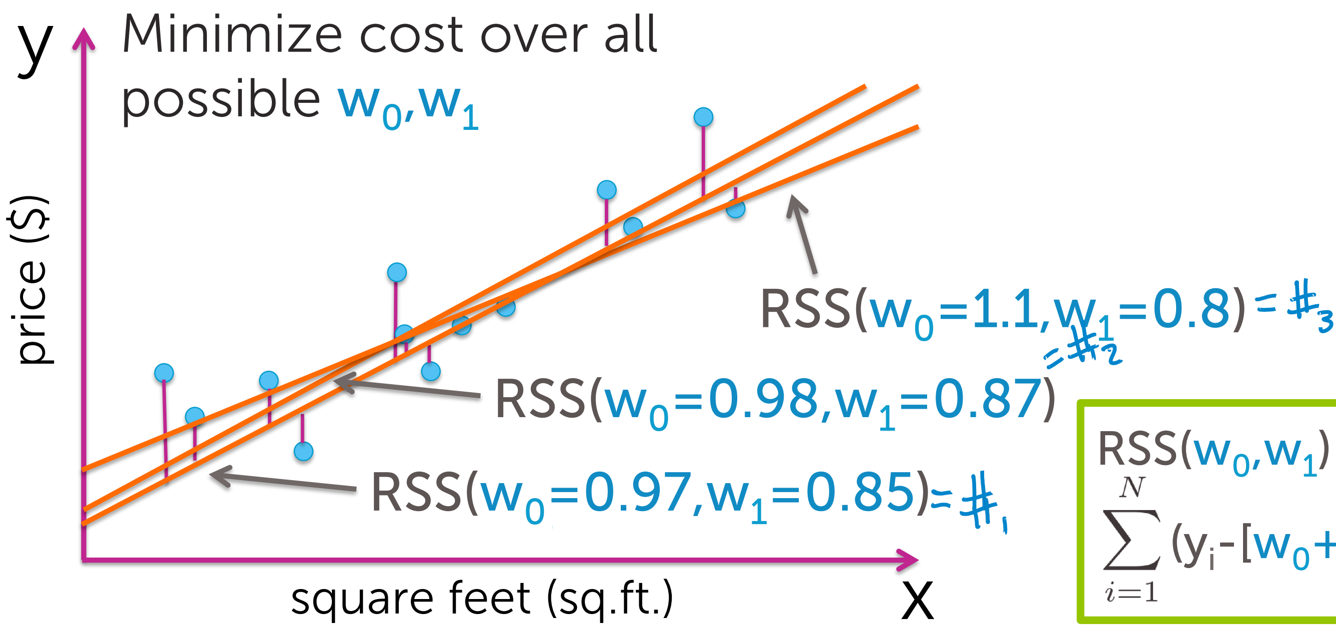


$$RSS(w_0,w_1) = \sum_{i=1}^{N} (y_i-[w_0+w_1x_i])^2$$

$\sum_{i=1}^{N} a_i = a_1 + a_2 + \dots + a_N$

Here,
$a_i = (y_i - [w_0 + w_1 x_i])^2$

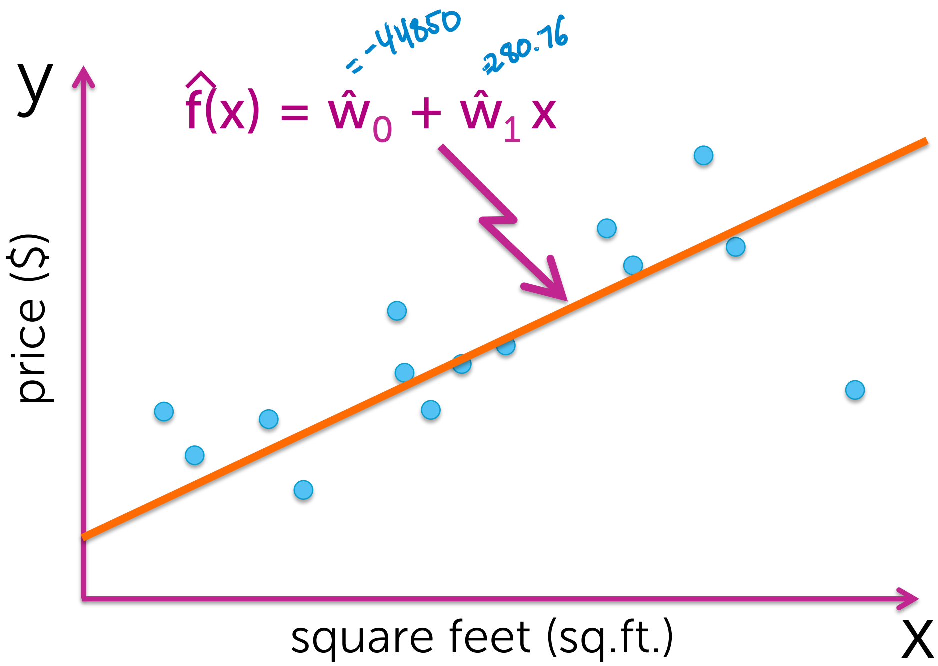
Find "best" line



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

The fitted line: use + interpretation

Model vs. fitted line



Regression model:

$y_i = w_0 + w_1 x_i + \varepsilon_i$

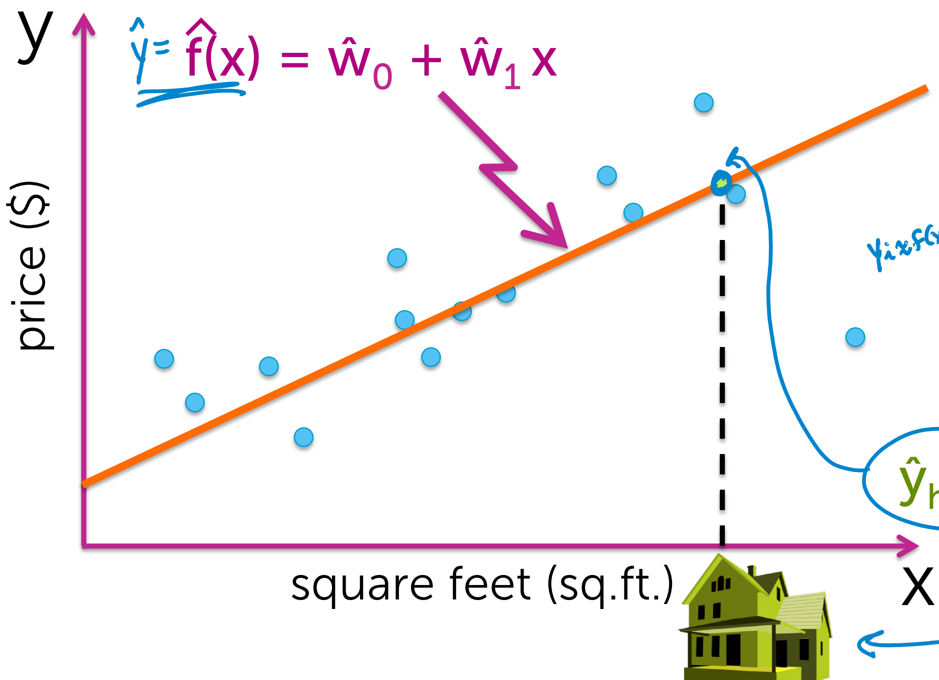
parameters (unknown variables)

Estimated parameters:

$\hat{w}_0 = -44850, \hat{w}_1 = 280.76$

take actual values

Seller:
Predicting your house price



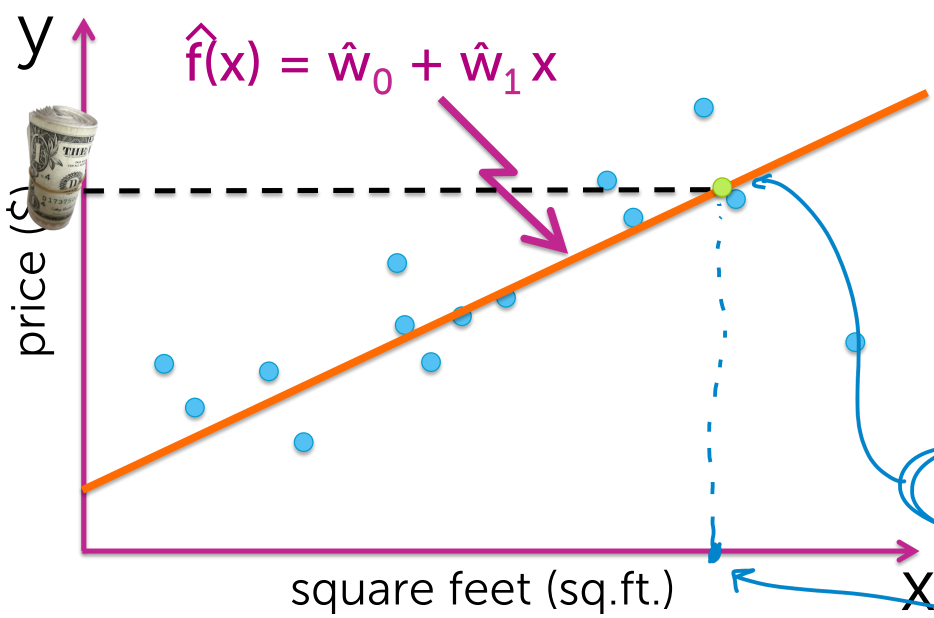
Regression model:

$y_i \approx f(x_i)$: $y_i = w_0 + w_1 x_i + \varepsilon_i$

Best guess of your house price:

$\hat{y}_{house} = \hat{w}_0 + \hat{w}_1\, sq.ft._{house}$

Buyer:
Predicting size of house



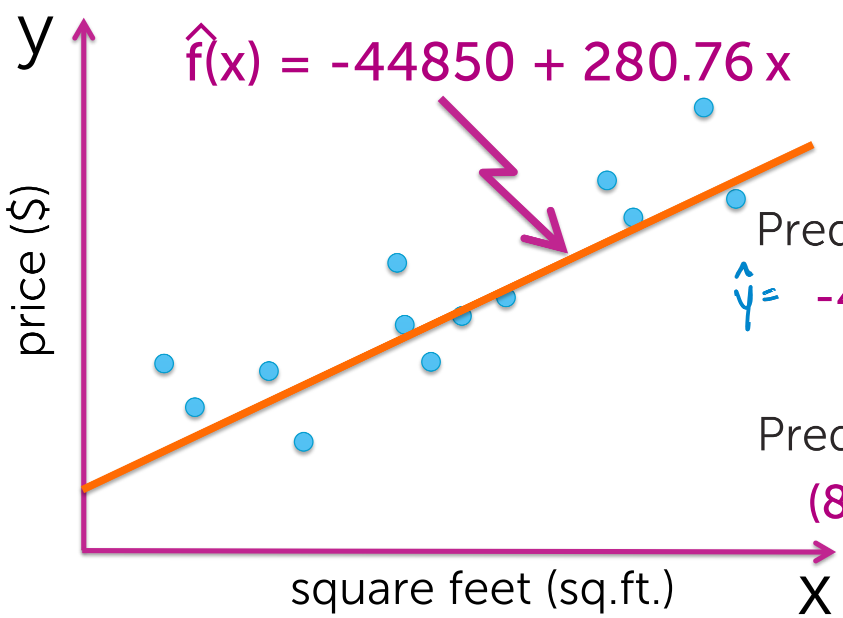
Regression model:

$$y_i = w_0 + w_1 x_i + \varepsilon_i$$

Best guess of size of house you can afford:

$$\$_{\text{in bank}} = \hat{w}_0 + \hat{w}_1 \widehat{\text{sq.ft.}}$$

A concrete example



Predict $ of 2,640 sq.ft. house:

$\hat{y} = -44850 + 280.76 * \underline{2,640}$

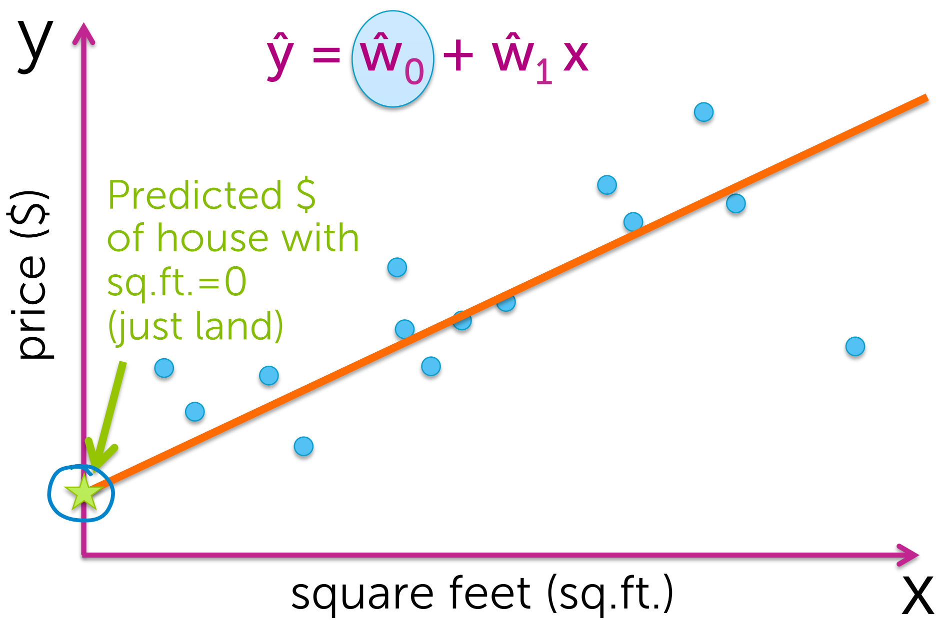
$= \$696,356$

Predict sq.ft. of <u>$859,000</u> sale:

$(859000+44850)/\ 280.76$

$= 3,219$ sq.ft.

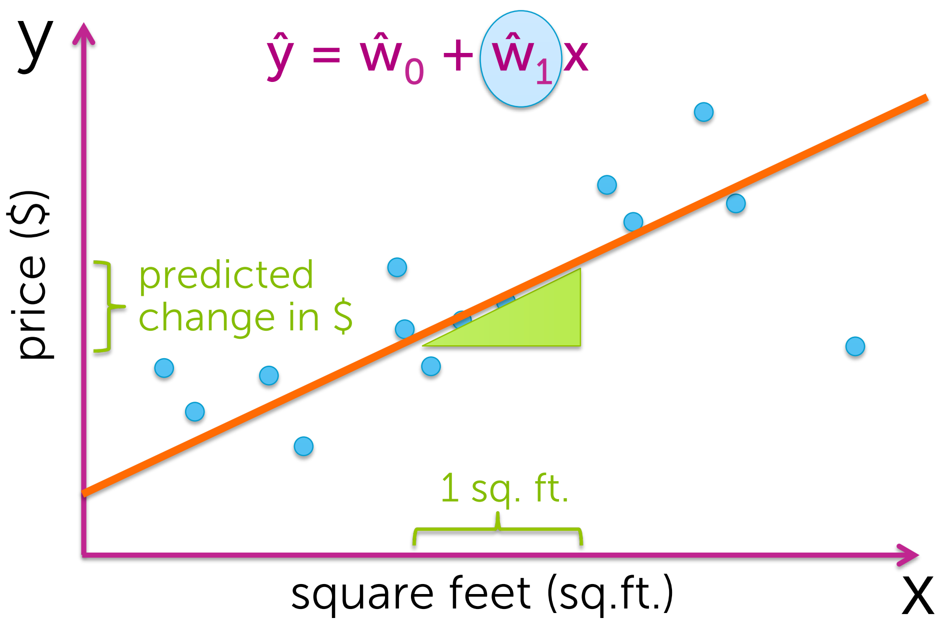
Interpreting the coefficients



$\hat{y} = \hat{w}_0$
when $x = 0$

not very meaningful

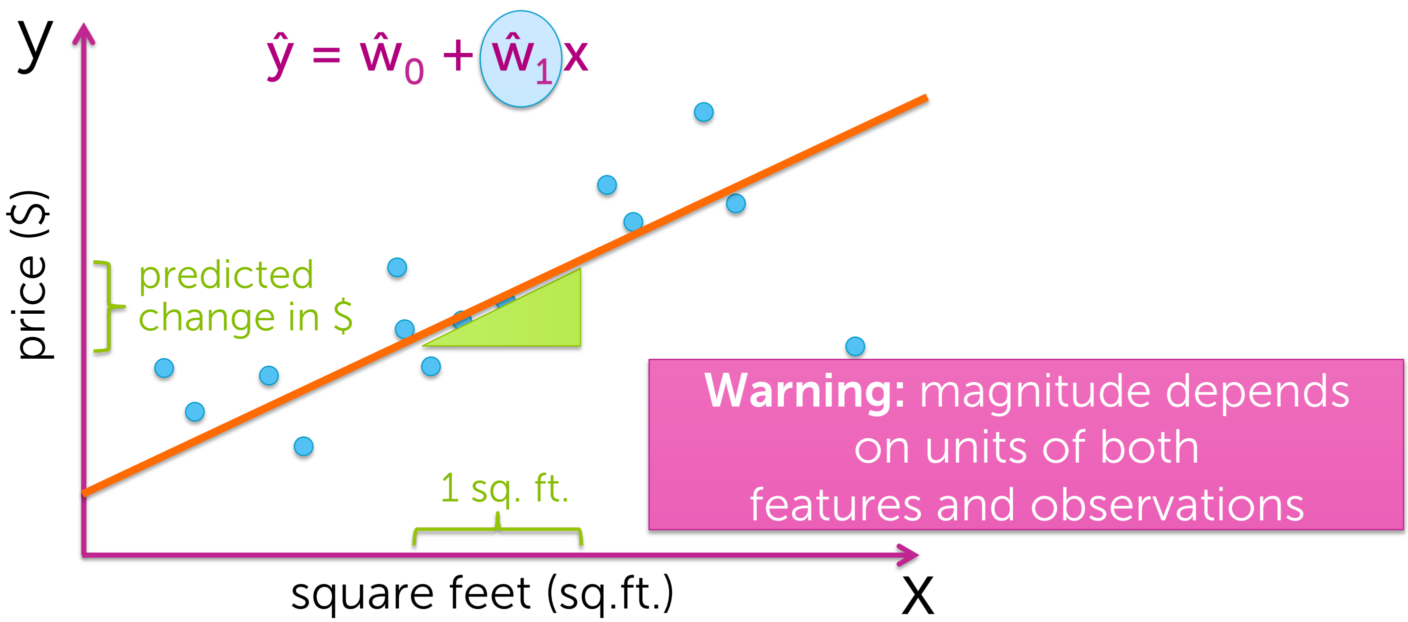
Interpreting the coefficients



$$\hat{\$}_{1001\ sq.ft.} - \hat{\$}_{1000\ sq.ft.}$$
$$= \cancel{\hat{w}_0} + \hat{w}_1 \cdot 1001\ sq.ft.$$
$$- (\cancel{\hat{w}_0} + \hat{w}_1 1000\ sq.ft.)$$
$$= \hat{w}_1$$

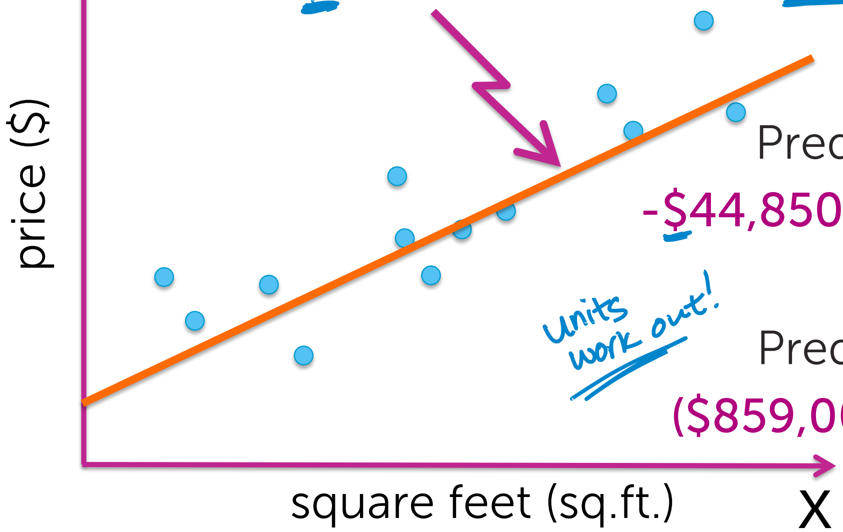
predicted change in the output per unit change in input

Interpreting the coefficients



A concrete example

$\hat{f}(x)$ = -\$44,850 + 280.76 (\$/sq.ft.) x



Predict \$ of 2,640 sq.ft. house:

-\$44,850 + 280.76 (\$/sq.ft.) * 2,640 sq.ft.

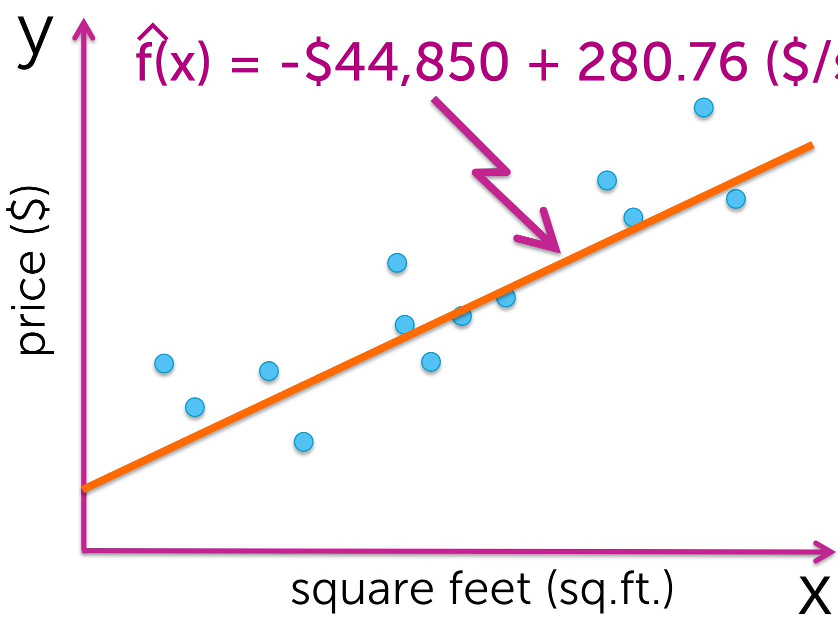
= \$696,356

Predict sq.ft. of \$859,000 sale:

(\$859,000+\$44,850)/ 280.76 (\$/sq.ft.)

= 3,219 sq.ft.

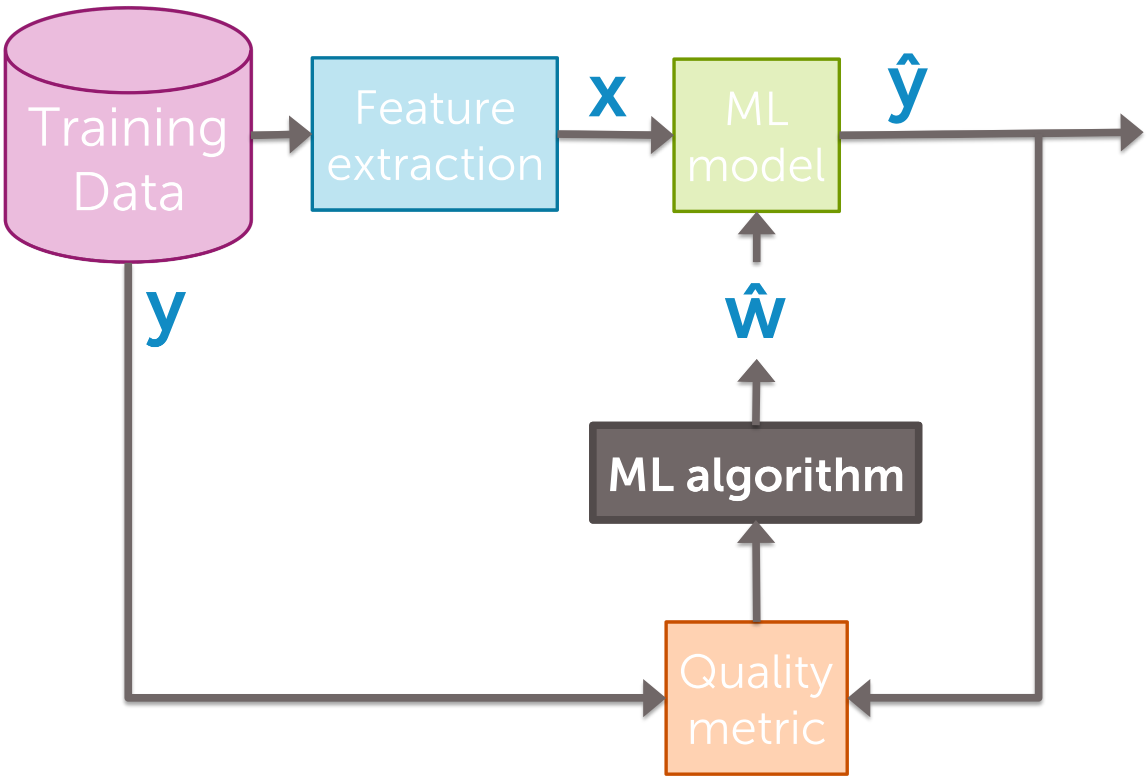
A concrete example



But what if:

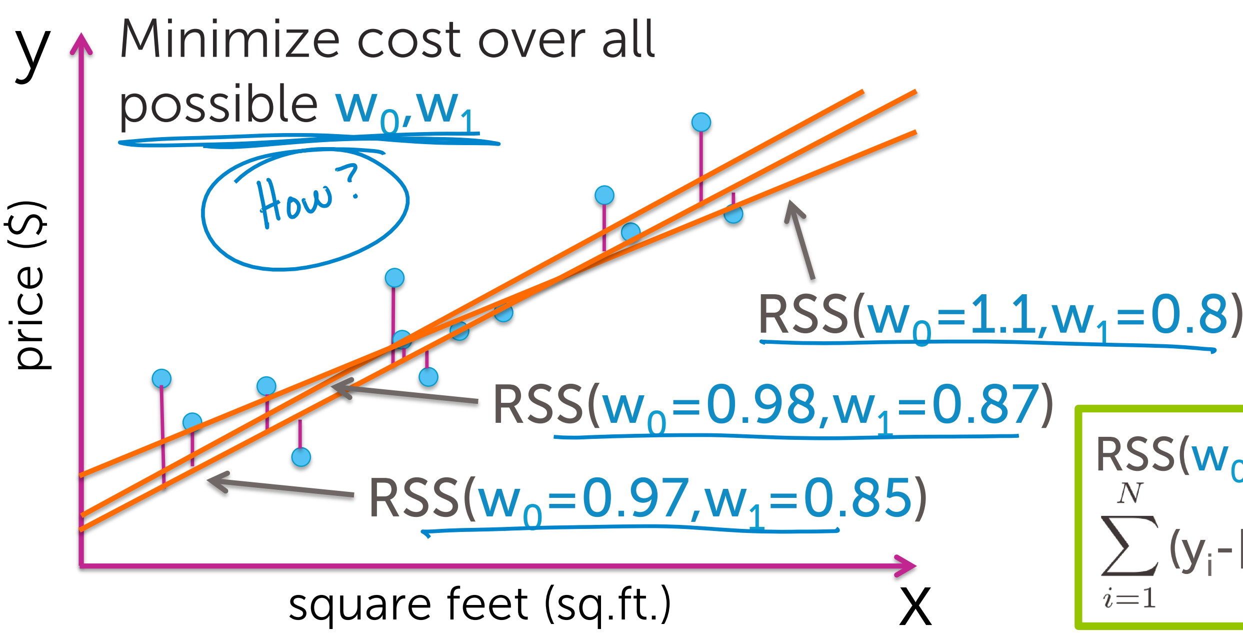
- House was measured in square meters?
- Price was measured in RMB?

Algorithms for fitting the model



Training Data
Feature extraction
x
ML model
ŷ
y
ŵ
ML algorithm
Quality metric

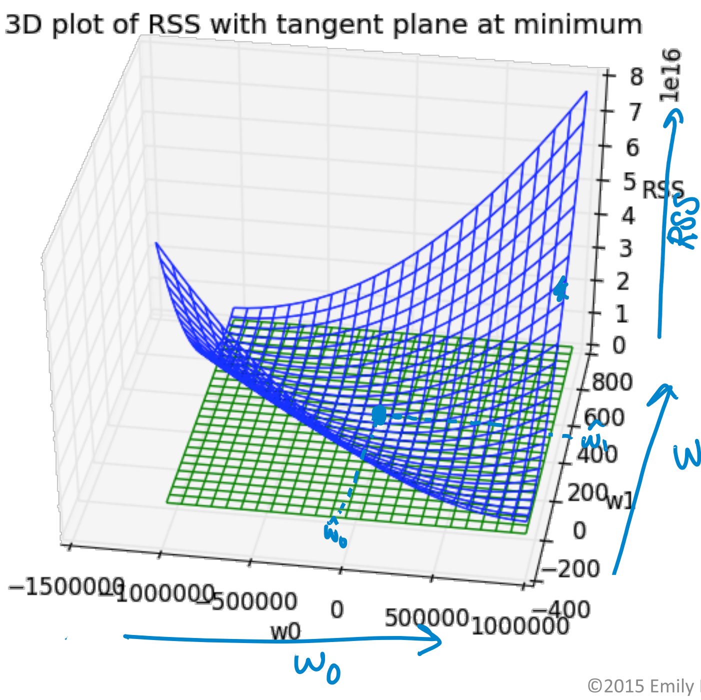
Find "best" line



Recall:

$$RSS(w_0,w_1) = \sum_{i=1}^{N}(y_i-[w_0+w_1x_i])^2$$

Minimizing the cost



Minimize function
over all possible w_0,w_1

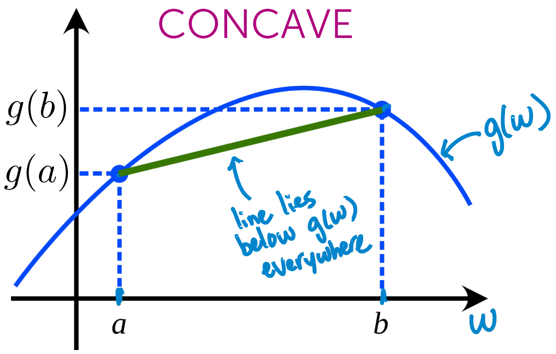
$$\boxed{\min_{w_0,w_1}} \underbrace{\sum_{i=1}^{N}(y_i-[w_0+w_1x_i])^2}$$

RSS(w_0,w_1) is a function
of 2 variables $= g(w_0,w_1)$

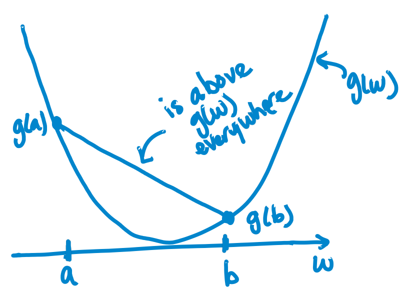
An aside on optimization

Convex/concave functions

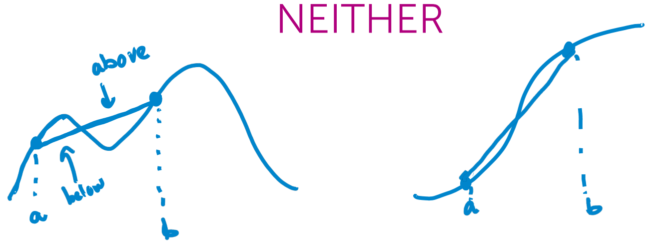
CONCAVE



CONVEX

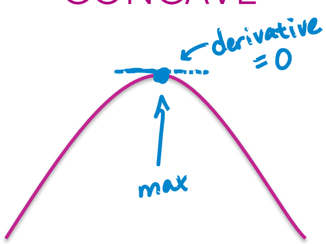


NEITHER



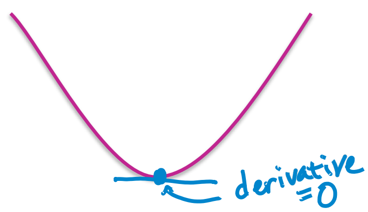
Finding the max or min analytically

CONCAVE



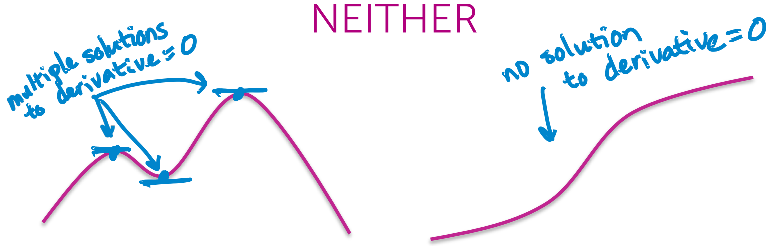
$\max_w g(w)$

CONVEX



$\min_w g(w)$

NEITHER



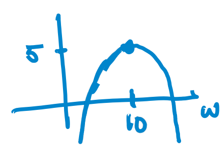
Example:

$g(w) = 5-\underline{(w-10)^2}$

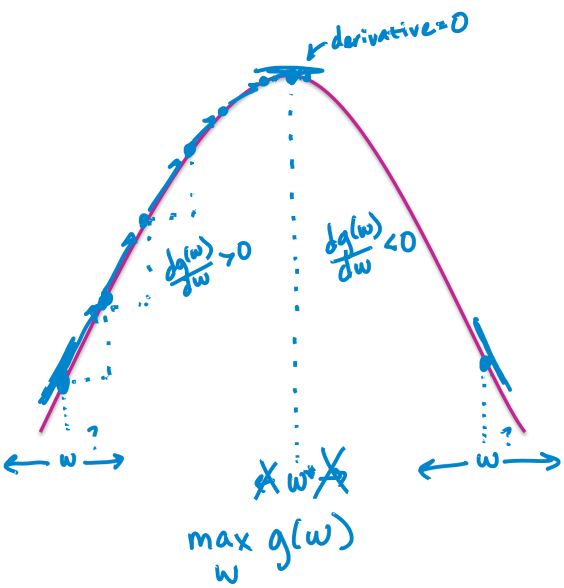
$$\frac{dg(w)}{dw} = 0 - 2(w-10)' \cdot 1$$
$$= -2w + 20$$

set derivate = 0:

$$-2w + 20 = 0$$
$$w = 10$$



Finding the max via hill climbing



How do we know whether to move w to right or left? (inc. or dec. the value of w?)

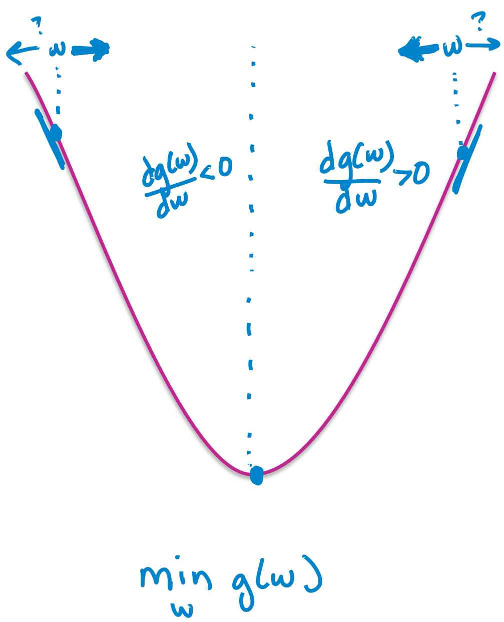
while not converged

$$w^{(t+1)} \leftarrow w^{(t)} + \eta \frac{dg(w)}{dw}$$

iteration t

stepsize

Finding the min via hill descent



when derivative is positive, we want to decrease w
and when derivative is negative, we want to increase w

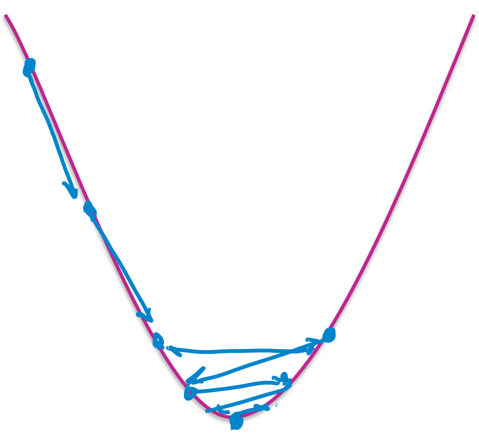
Algorithm:

while not converged

$$w^{(t+1)} \leftarrow w^{(t)} - \eta \left.\frac{dg}{dw}\right|_{w^{(t)}}$$

Choosing the stepsize—
Fixed stepsize

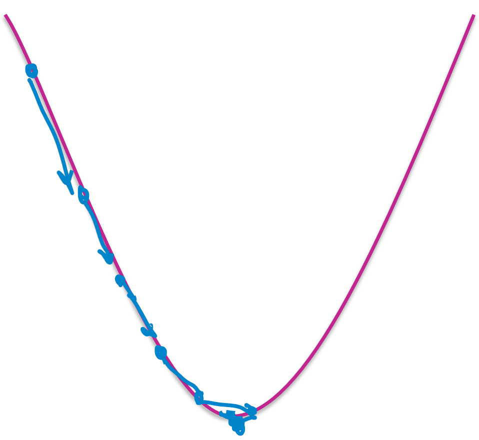
η



$\eta = 0.1$

Choosing the stepsize—
Decreasing stepsize

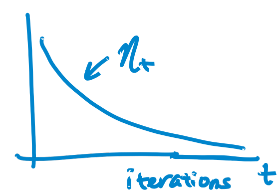
or stepsize schedule



Common choices:

$$\eta_t = \frac{\alpha}{t}$$

$$\eta_t = \frac{\alpha}{\sqrt{t}}$$



Convergence criteria

For convex functions,
optimum occurs when

$$\frac{dg(w)}{dw} = 0$$

In practice, stop when

$$\left|\frac{dg(w)}{dw}\right| < \epsilon$$

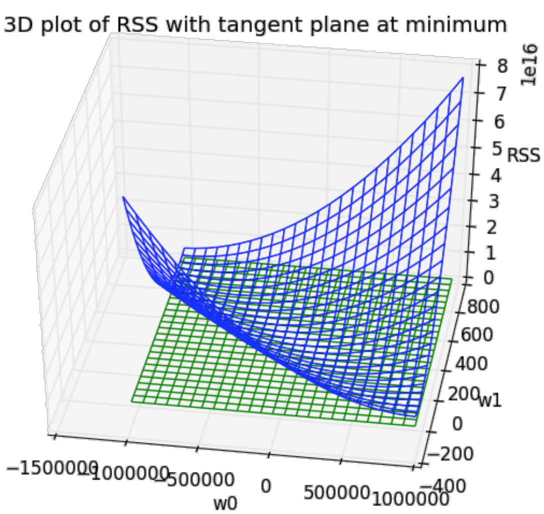
threshold
to be set

Algorithm:

while not converged

$$w^{(t+1)} \leftarrow w^{(t)} - \eta \left.\frac{dg}{dw}\right|_{w^{(t)}}$$

Moving to multiple dimensions: Gradients



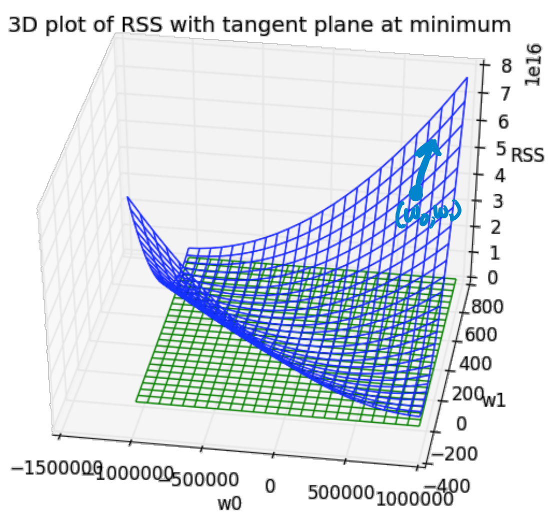
$$\nabla g(\mathbf{w}) = \begin{bmatrix} \frac{\partial g}{\partial w_0} \\ \frac{\partial g}{\partial w_1} \\ \vdots \\ \frac{\partial g}{\partial w_p} \end{bmatrix} \Bigg\} \text{(p+1)-dimensional vector}$$

gradient

$[w_0, w_1, \ldots, w_p]$

partial derivative is like a derivate with respect to w_1 treating all other variables as constants

Gradient example



$$g(\mathbf{w}) = 5w_0 + 10w_0w_1 + 2w_1^2$$

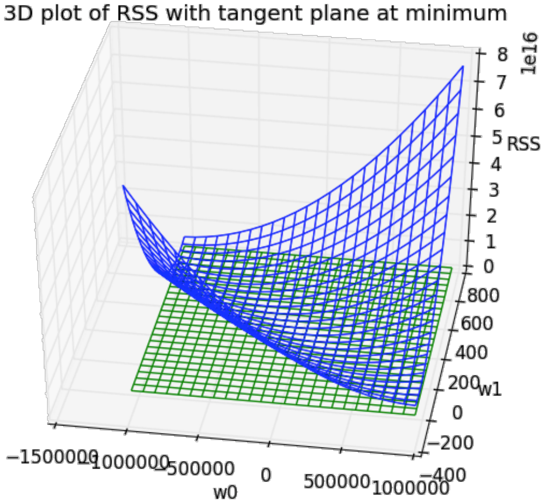
$$\frac{\partial g}{\partial w_0} = 5 + 10w_1$$

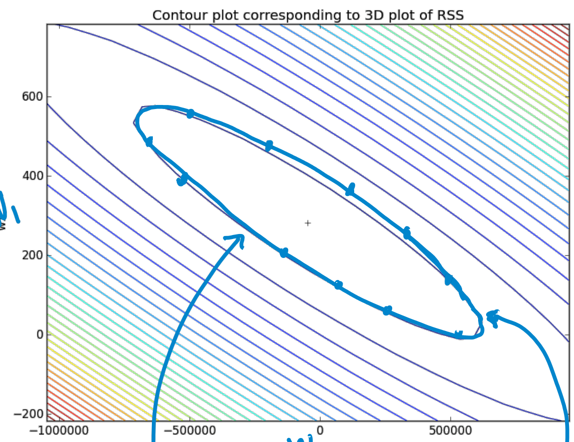
$$\frac{\partial g}{\partial w_1} = 10w_0 + 4w_1$$

$$\nabla g(\mathbf{w}) = \begin{bmatrix} 5 + 10w_1 \\ 10w_0 + 4w_1 \end{bmatrix}$$

Contour plots

bird's eye view





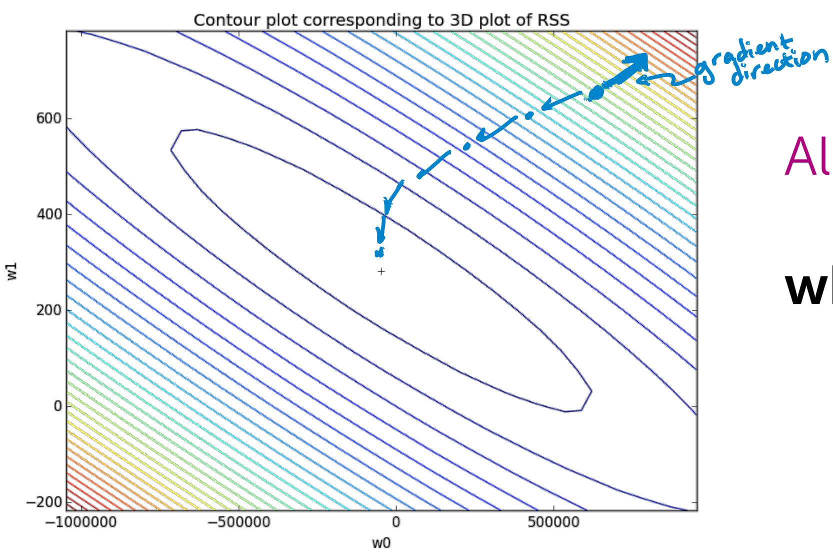
w_1

w_0

a slice of the 3D surface

$g(w_0,w_1)$

Gradient descent



Algorithm:

while not converged

$$w^{(t+1)} \leftarrow w^{(t)} - \eta \nabla \underline{g(w^{(t)})}$$

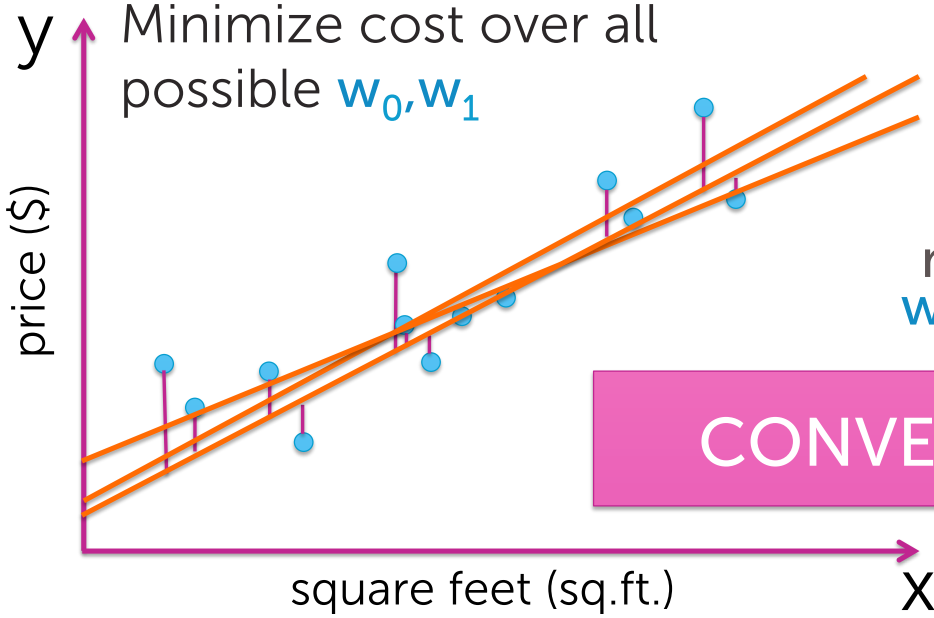
$$\begin{bmatrix} \vdots \end{bmatrix} \longleftarrow \begin{bmatrix} \vdots \end{bmatrix} - \eta \begin{bmatrix} \vdots \end{bmatrix}$$

$$\boxed{\begin{array}{l}\text{Convergence:} \\ \|\nabla g(w)\| < \epsilon\end{array}}$$

Finding the least squares line

Find "best" line

Minimize cost over all possible w_0, w_1



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

CONVEX

$\Rightarrow$ solution is unique

+ gradient descent alg. will converge to minimum