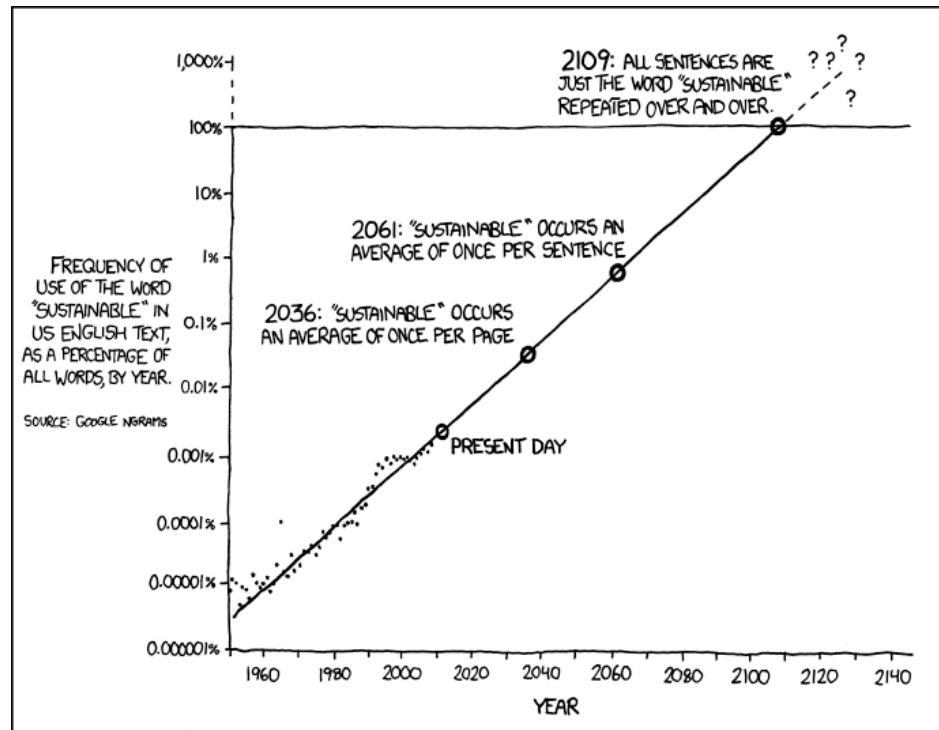


Linear Regression



THE WORD "SUSTAINABLE" IS UNSUSTAINABLE.

<https://xkcd.com/1007/>

Training set of housing prices (Portland, OR)	Size in feet² (x)	Price (\$) in 1000's (y)
	2104	460
	1416	232
	1534	315
	852	178

Notation:

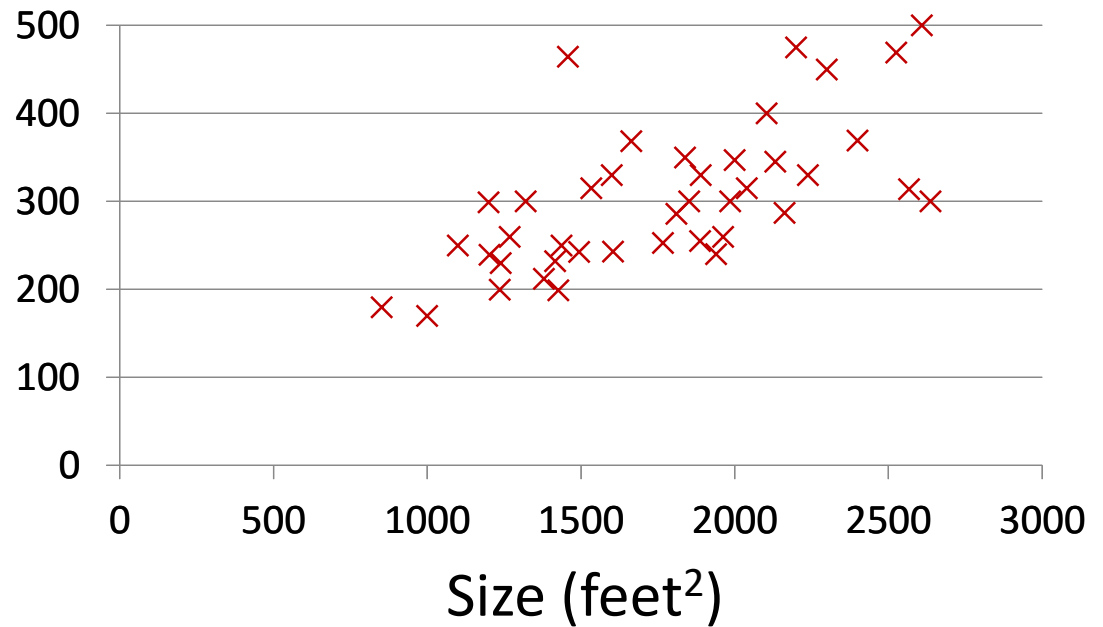
m = Number of training examples

x's = "input" variable / features

y's = "output" variable / "target" variable

Housing Prices (Portland, OR)

Price
(in 1000s
of dollars)



Supervised Learning

Given the “right answer” for each example in the data.

Regression Problem

Predict real-valued output

Training Set



Learning Algorithm



Size of
house



h



Estimated
price

How do we represent h ?

Linear regression with one variable.
Univariate linear regression.

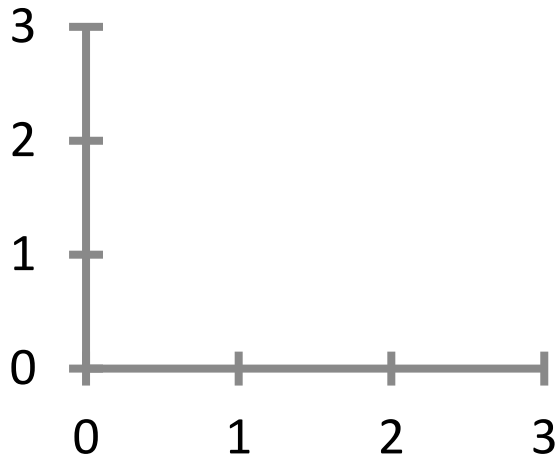
Training Set	Size in feet ² (x)	Price (\$) in 1000's (y)
	2104	460
	1416	232
	1534	315
	852	178

Hypothesis: $h_{\theta}(x) = \theta_0 + \theta_1 x$

θ_i 's: Parameters

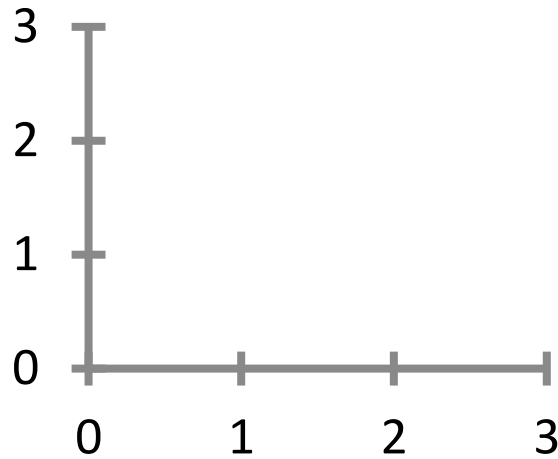
How to choose θ_i 's ?

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$



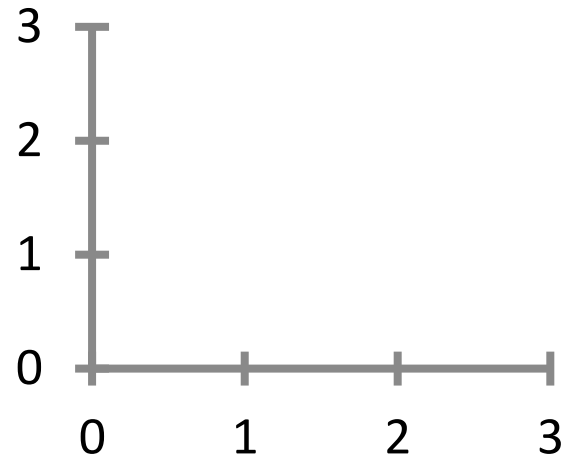
$$\theta_0 = 1.5$$

$$\theta_1 = 0$$



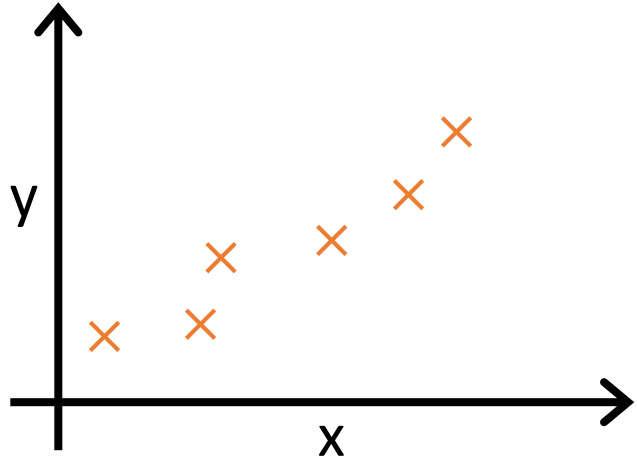
$$\theta_0 = 0$$

$$\theta_1 = 0.5$$



$$\theta_0 = 1$$

$$\theta_1 = 0.5$$

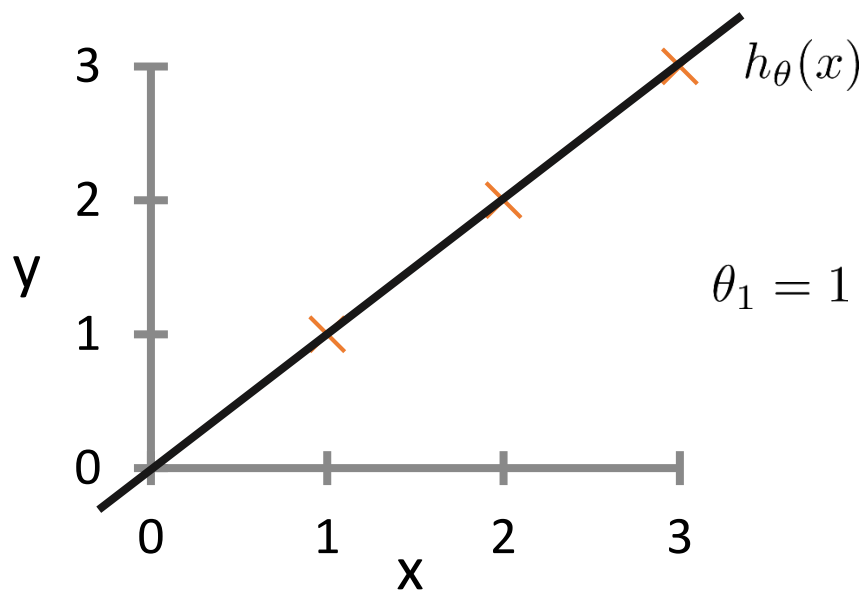


Idea: Choose θ_0, θ_1 so that $h_{\theta}(x)$ is close to y for our training examples (x, y)

Quadratic cost function – on the board

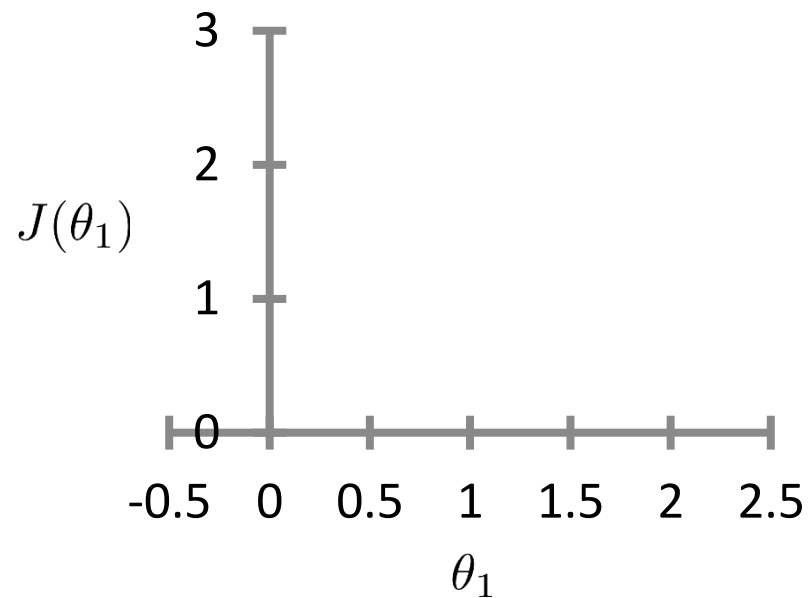
$$h_{\theta}(x)$$

(for fixed θ_1 , this is a function of x)



$$J_{\theta_0}(\theta_1)$$

(function of the parameter θ_1 , for a fixed θ_0)



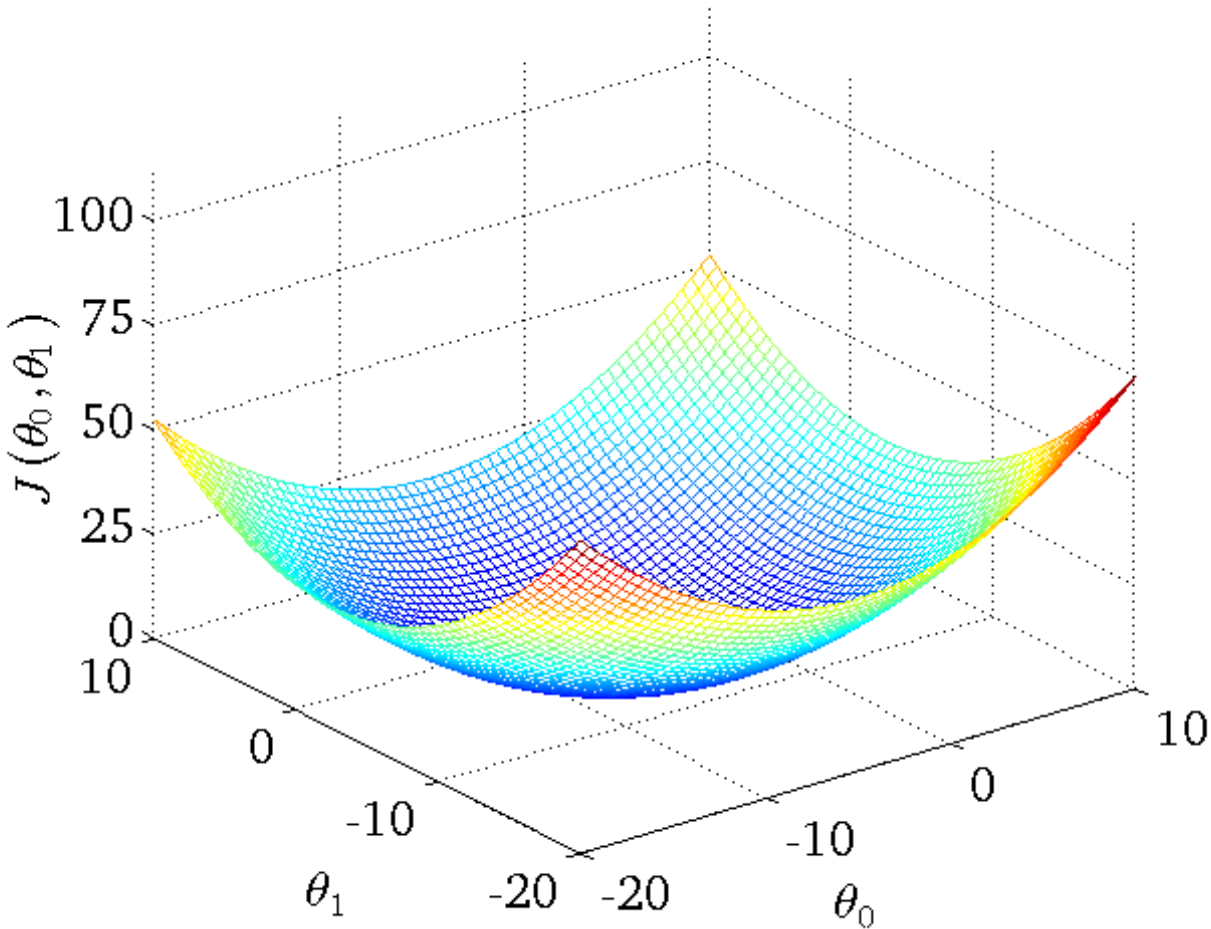
Hypothesis: $h_{\theta}(x) = \theta_0 + \theta_1 x$

Parameters: θ_0, θ_1

Cost Function: $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$

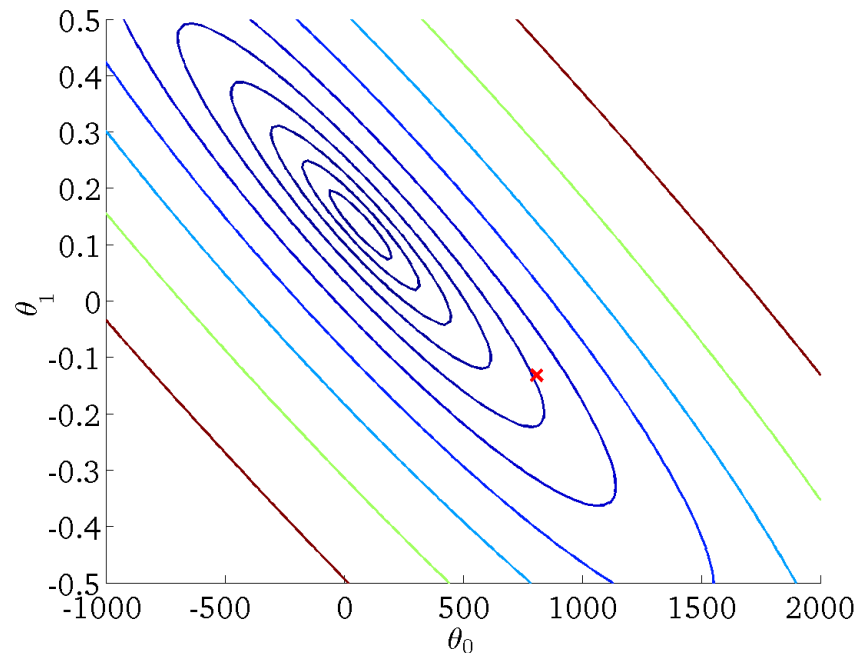
Goal: minimize $J(\theta_0, \theta_1)$
 θ_0, θ_1

Cost Function Surface Plot



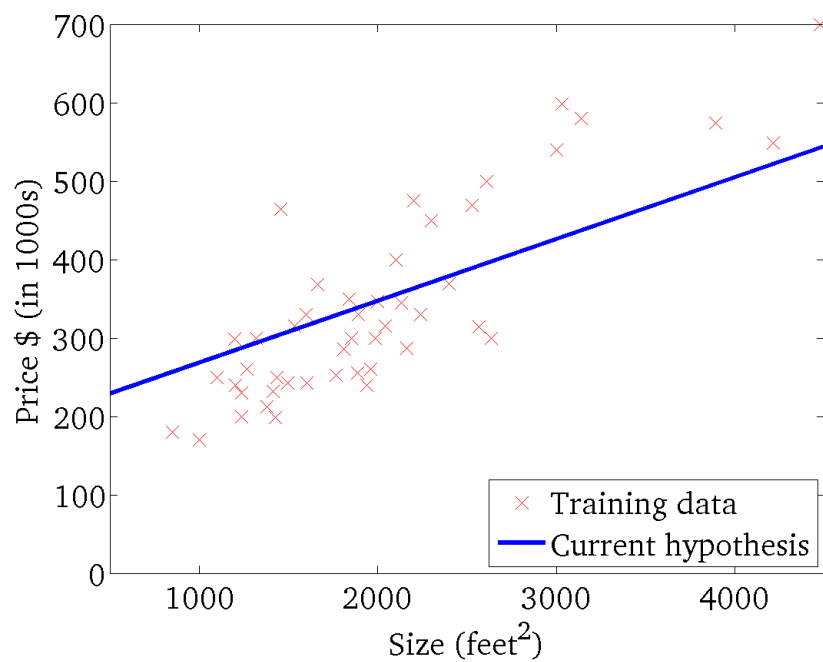
Contour Plots

- For a function $F(x, y)$ of two variables, assigned different colours to different values of F
- Pick some values to plot
- The result will be *contours* – curves in the graph along which the values of $F(x, y)$ are constant



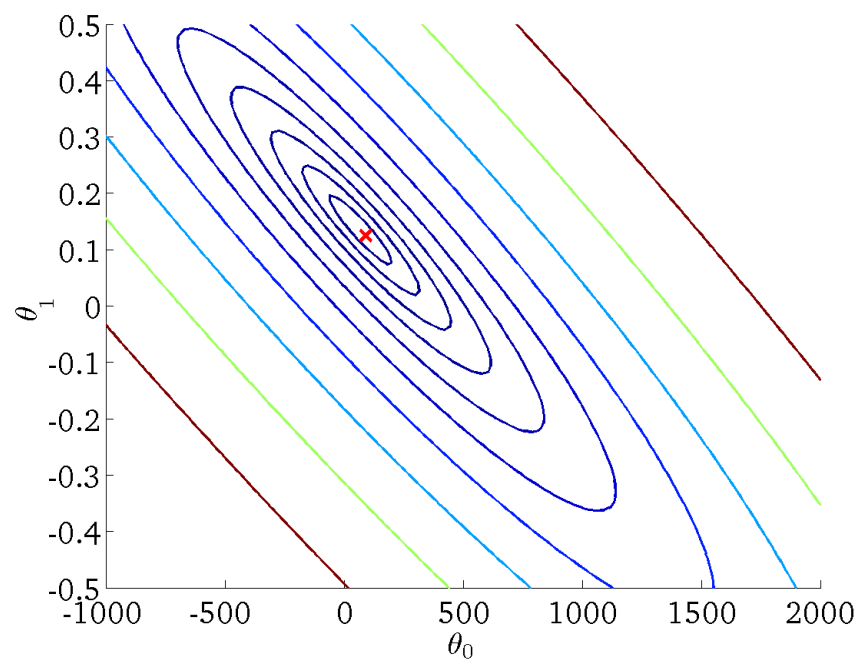
$$h_{\theta}(x)$$

(for fixed θ_0, θ_1 , this is a function of x)



$$J(\theta_0, \theta_1)$$

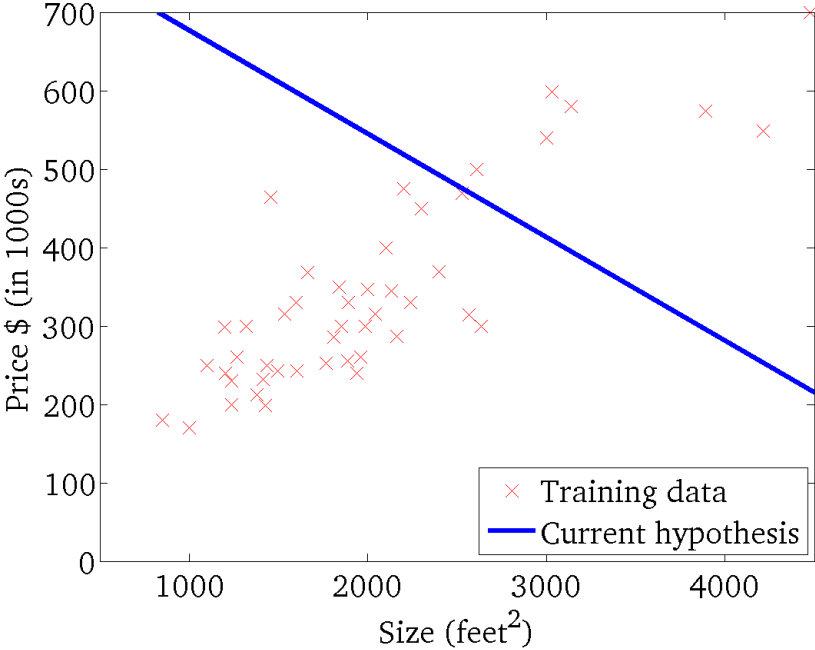
(function of the parameters θ_0, θ_1)



Cost Function Contour Plot

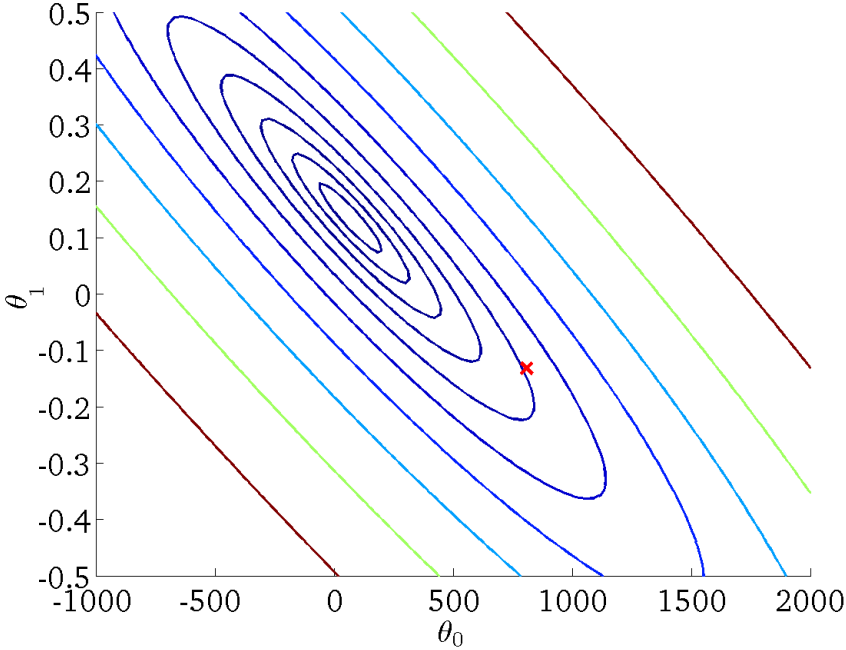
$$h_{\theta}(x)$$

(for fixed θ_0, θ_1 , this is a function of x)



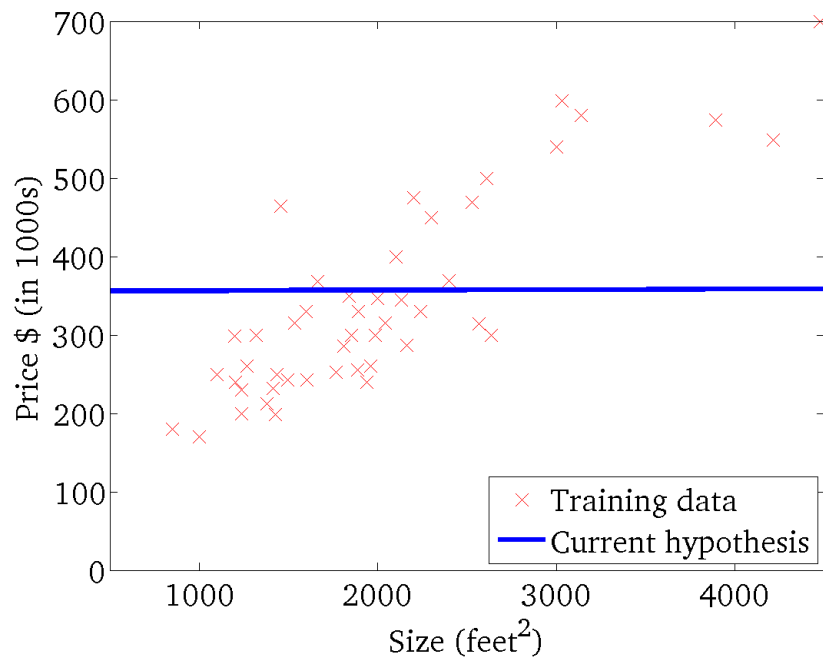
$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)



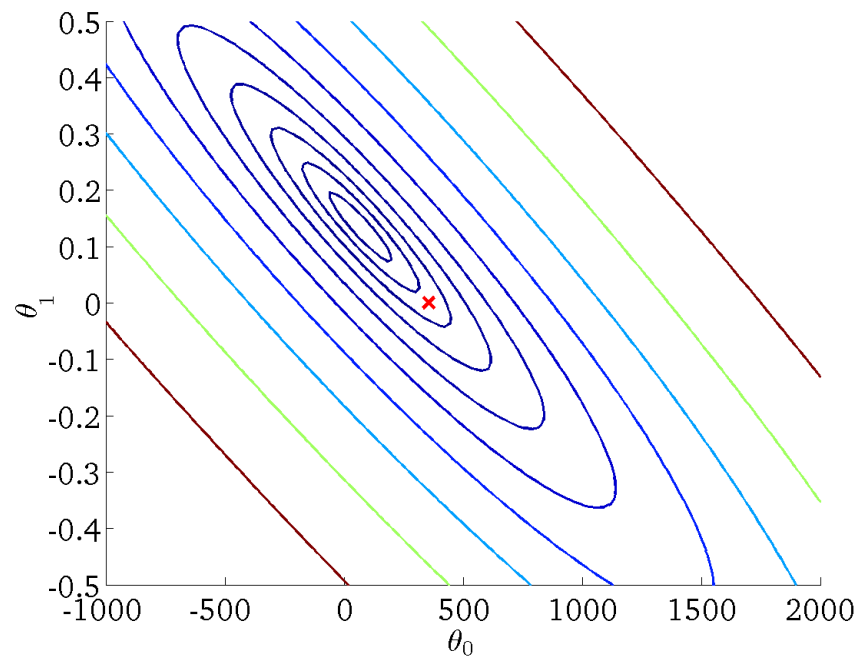
$$h_{\theta}(x)$$

(for fixed θ_0, θ_1 , this is a function of x)



$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)



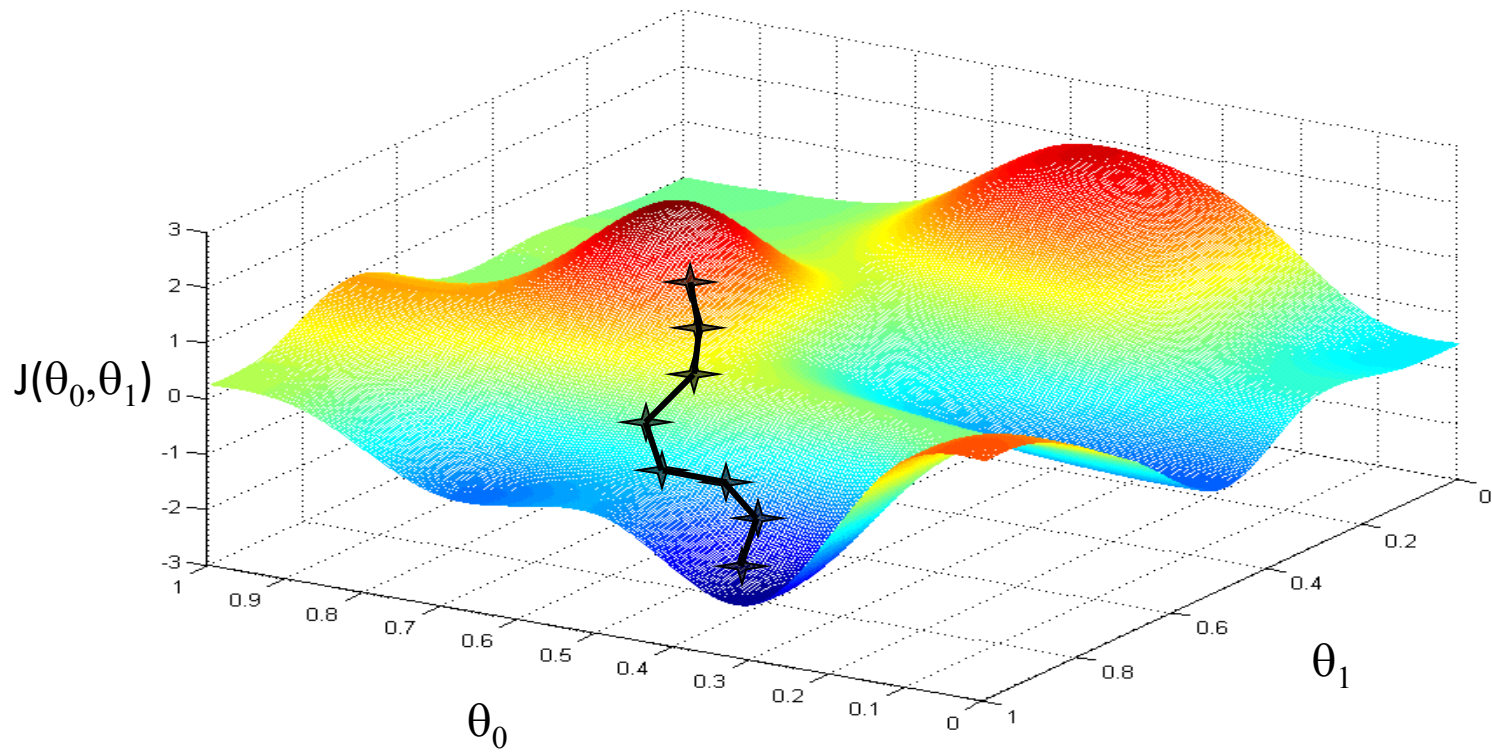
Have some function $J(\theta_0, \theta_1)$

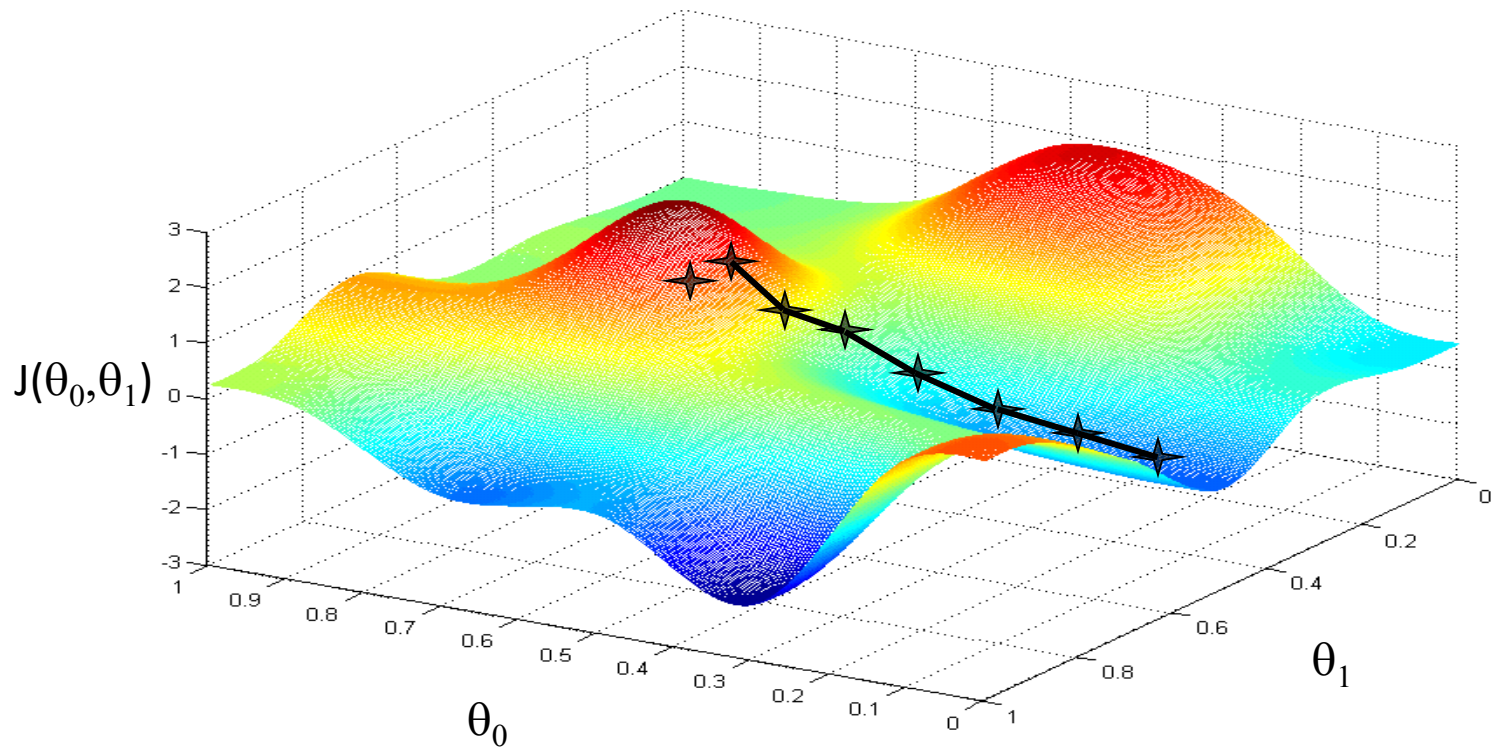
Want $\min_{\theta_0, \theta_1} J(\theta_0, \theta_1)$

Outline:

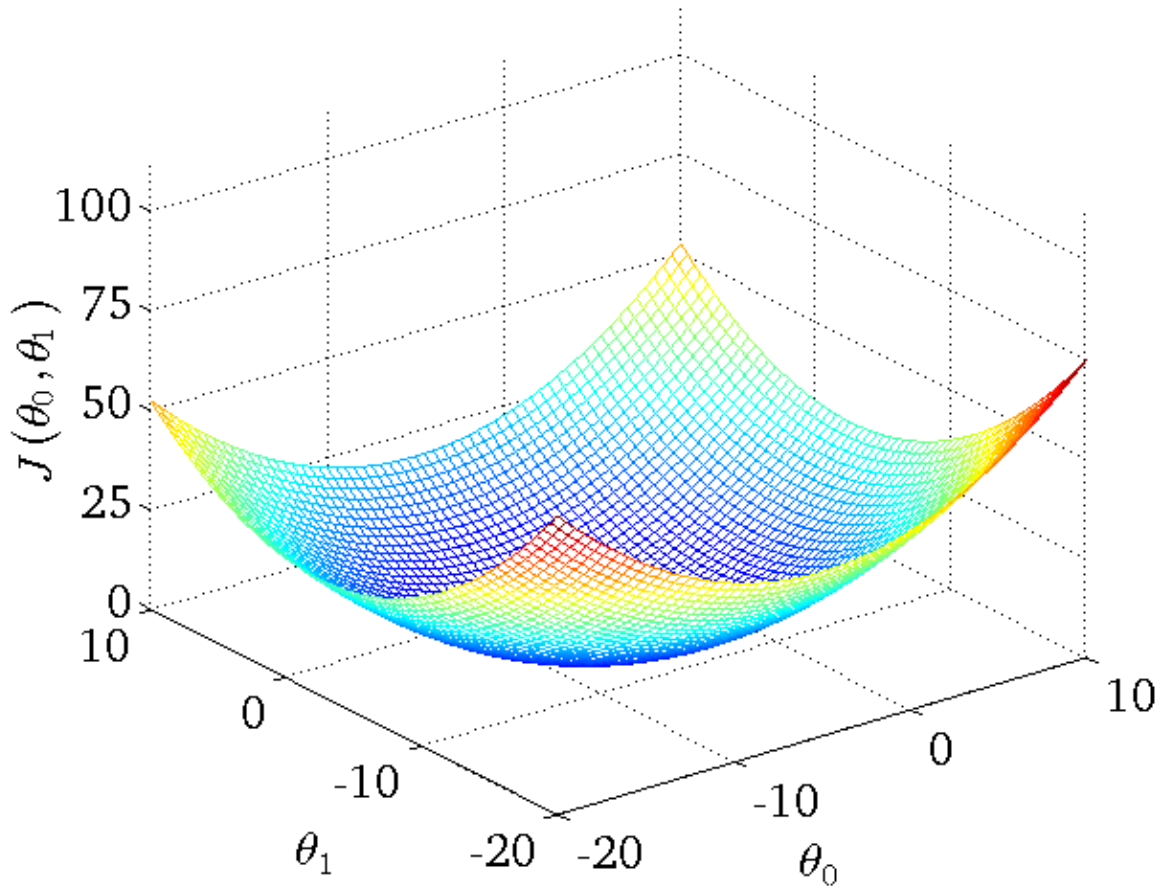
- Start with some θ_0, θ_1
- Keep changing θ_0, θ_1 to reduce $J(\theta_0, \theta_1)$
until we hopefully end up at a minimum

Gradient Descent on the board



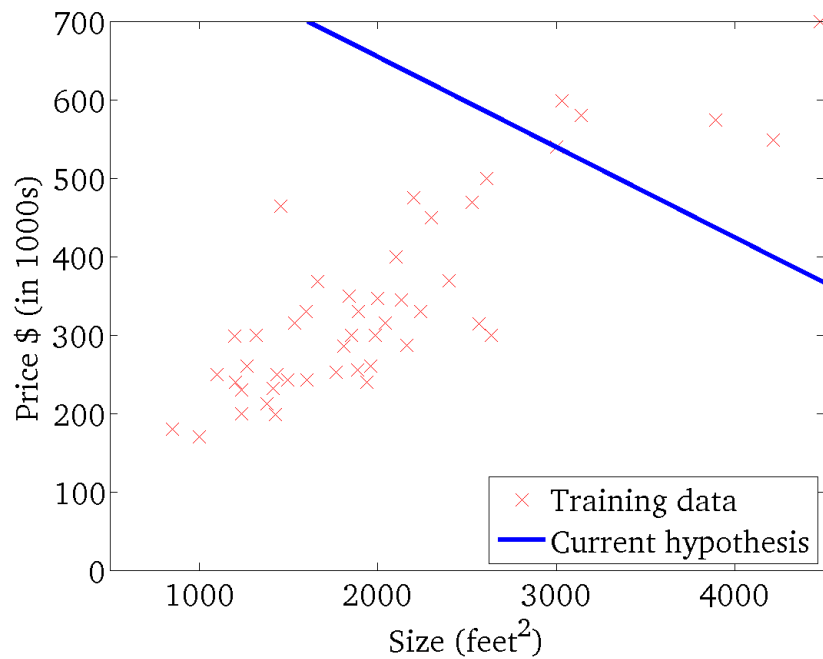


For Linear Regression, J is bowl-shaped (“convex”)



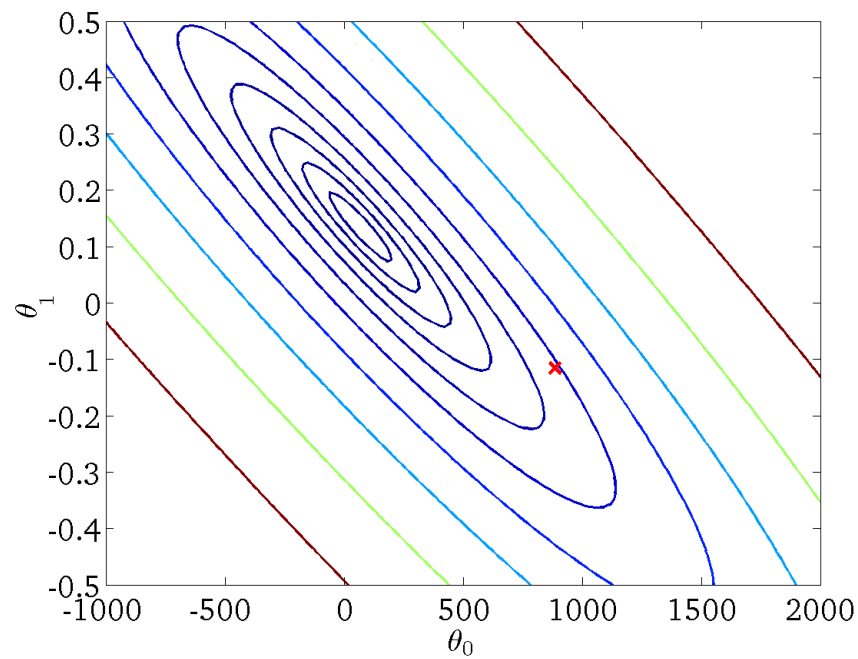
$$h_{\theta}(x)$$

(for fixed θ_0, θ_1 , this is a function of x)



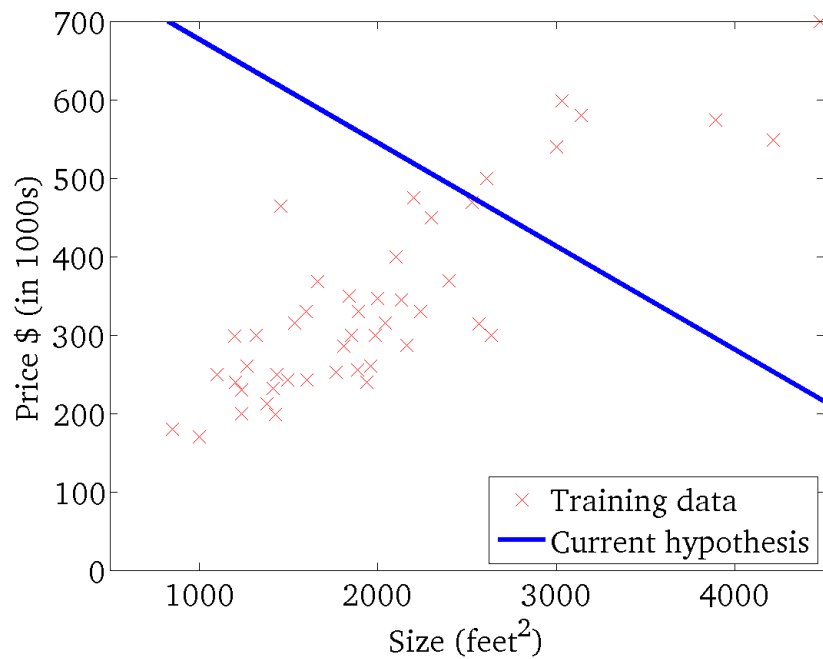
$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)



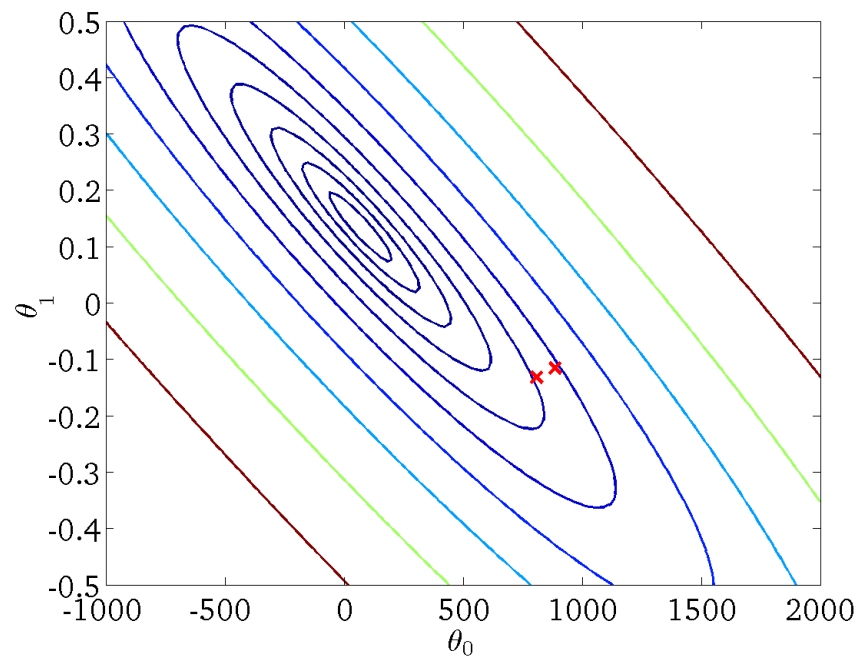
$$h_{\theta}(x)$$

(for fixed θ_0, θ_1 , this is a function of x)



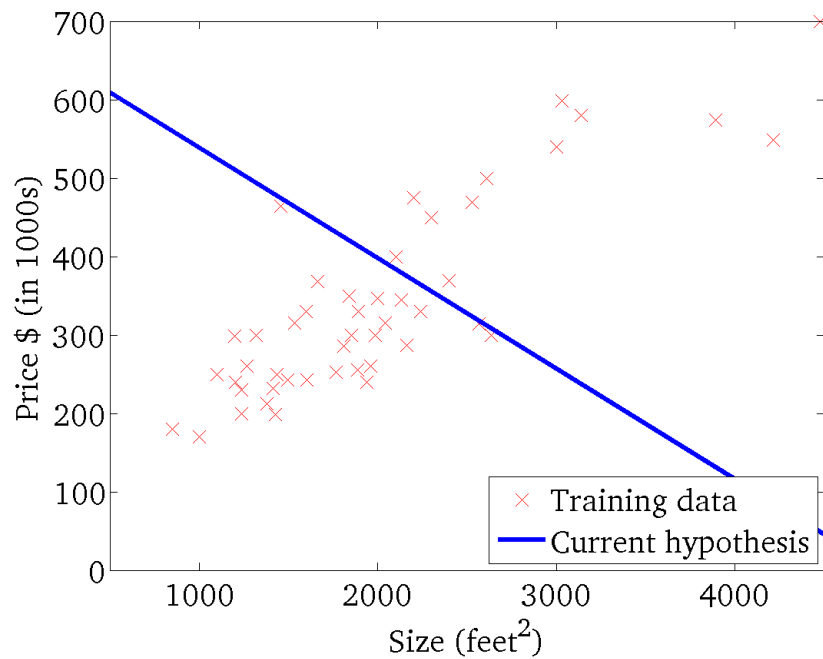
$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)



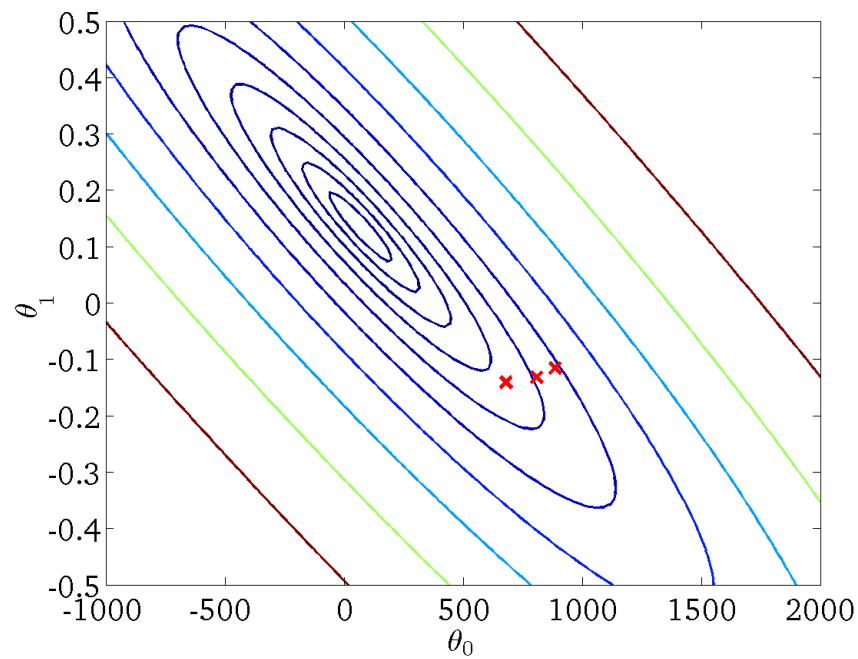
$$h_{\theta}(x)$$

(for fixed θ_0, θ_1 , this is a function of x)



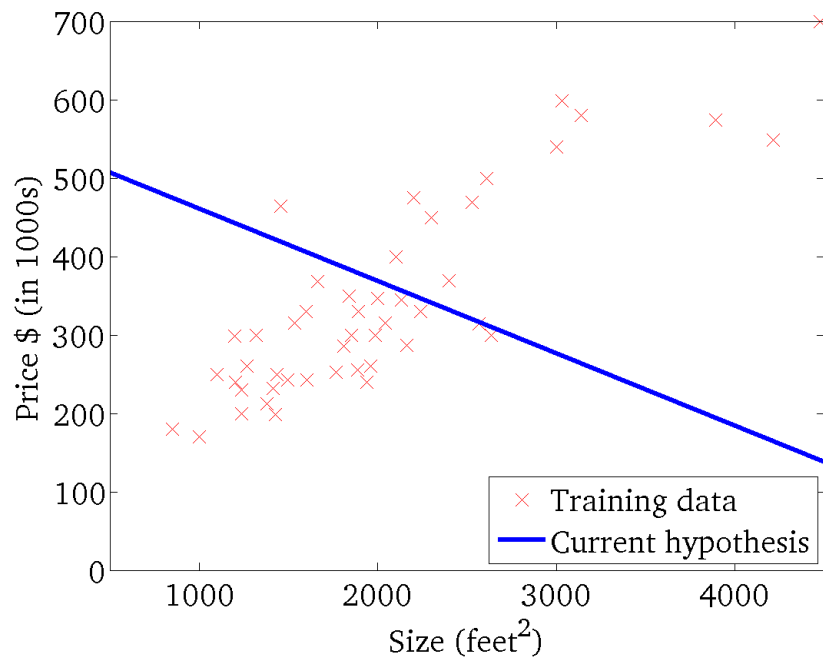
$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)



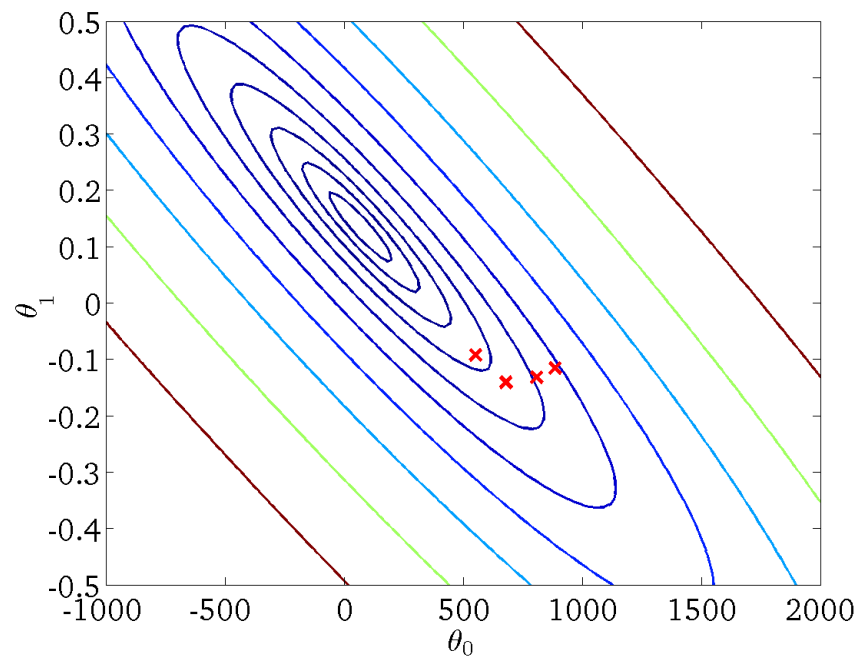
$$h_{\theta}(x)$$

(for fixed θ_0, θ_1 , this is a function of x)



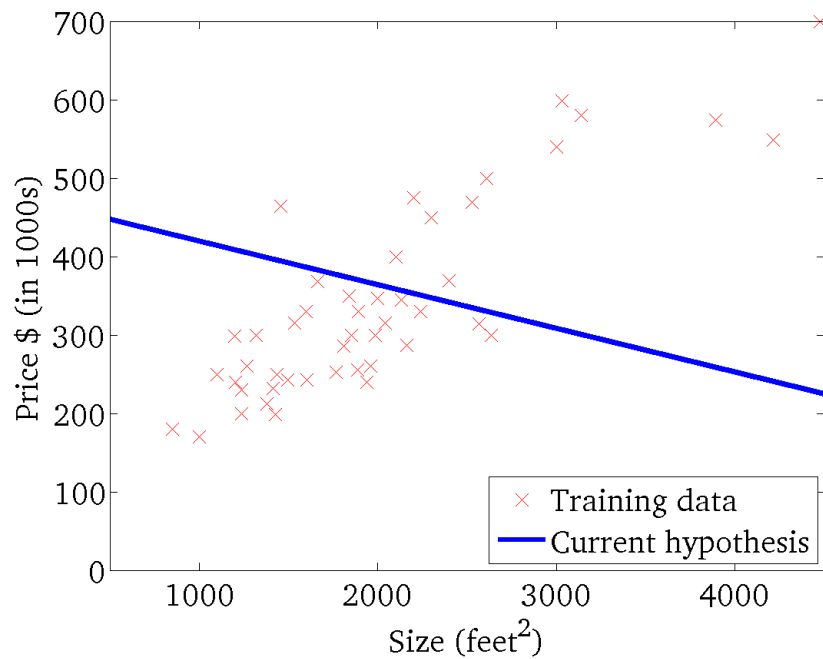
$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)



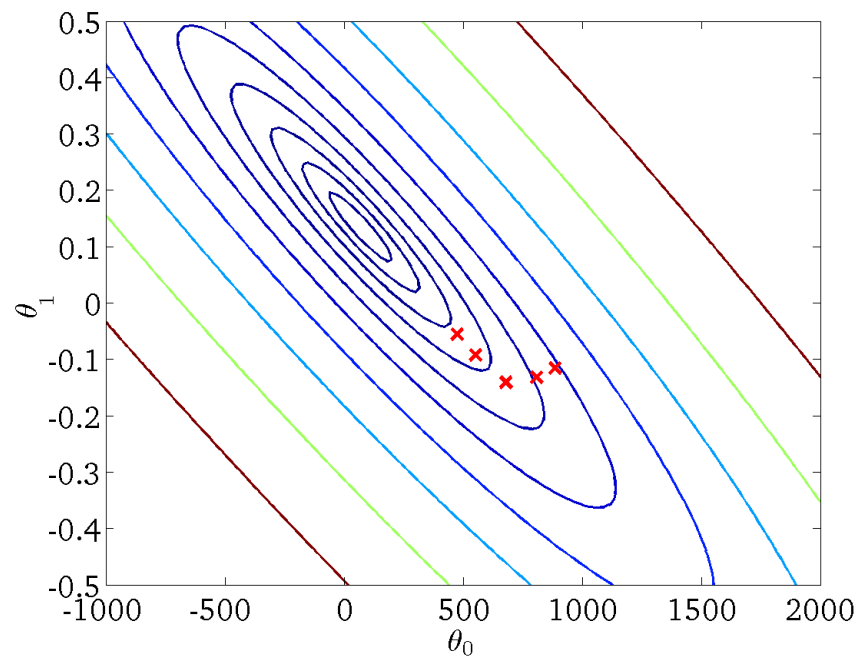
$$h_{\theta}(x)$$

(for fixed θ_0, θ_1 , this is a function of x)



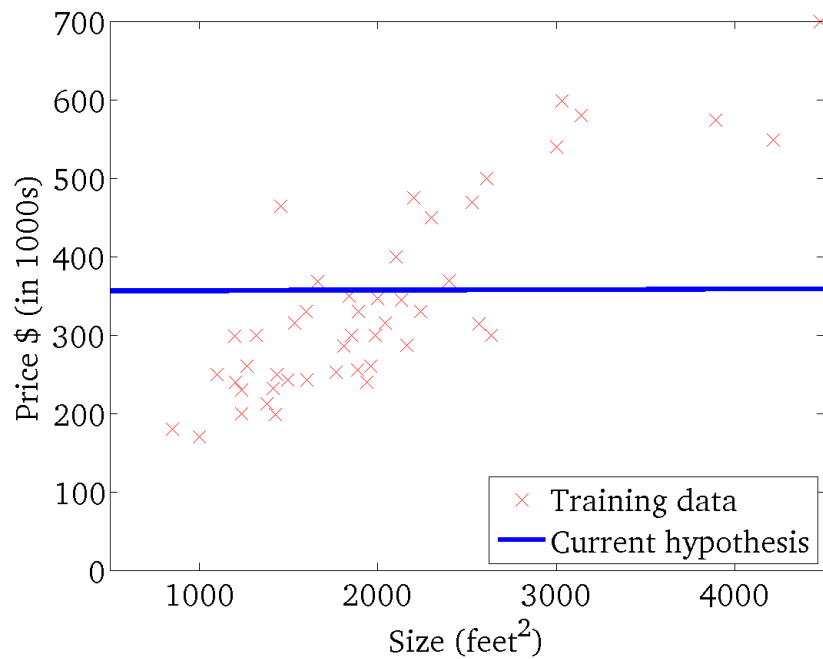
$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)



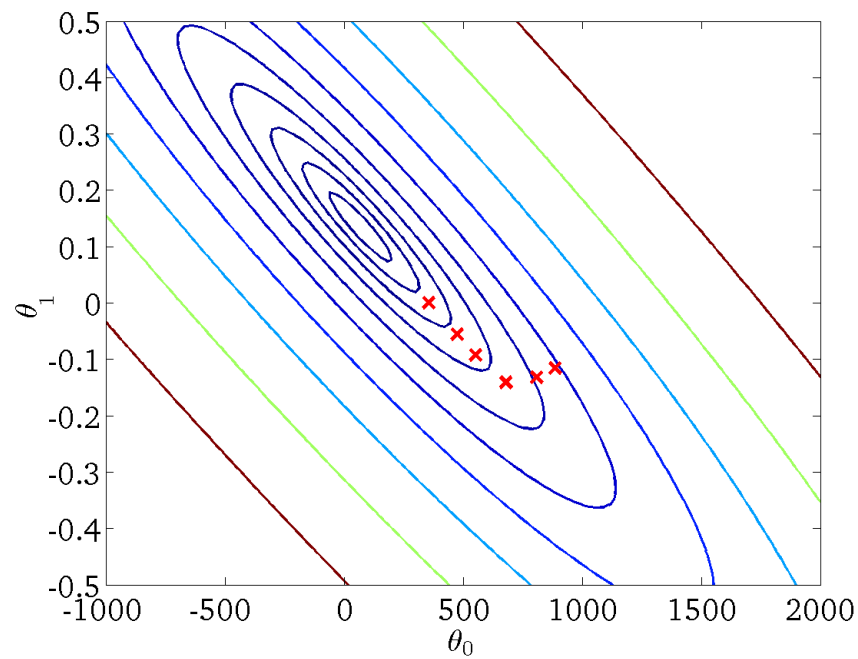
$$h_{\theta}(x)$$

(for fixed θ_0, θ_1 , this is a function of x)



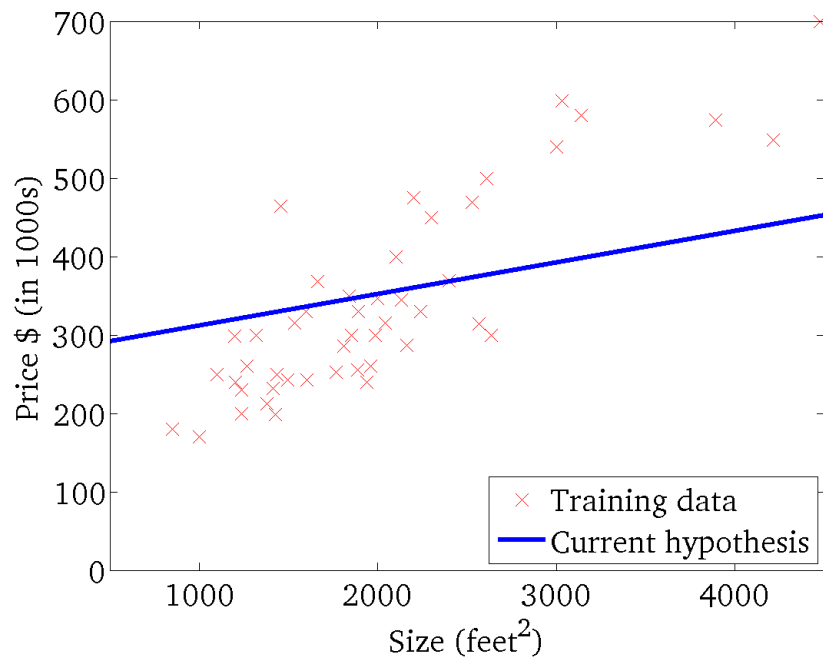
$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)



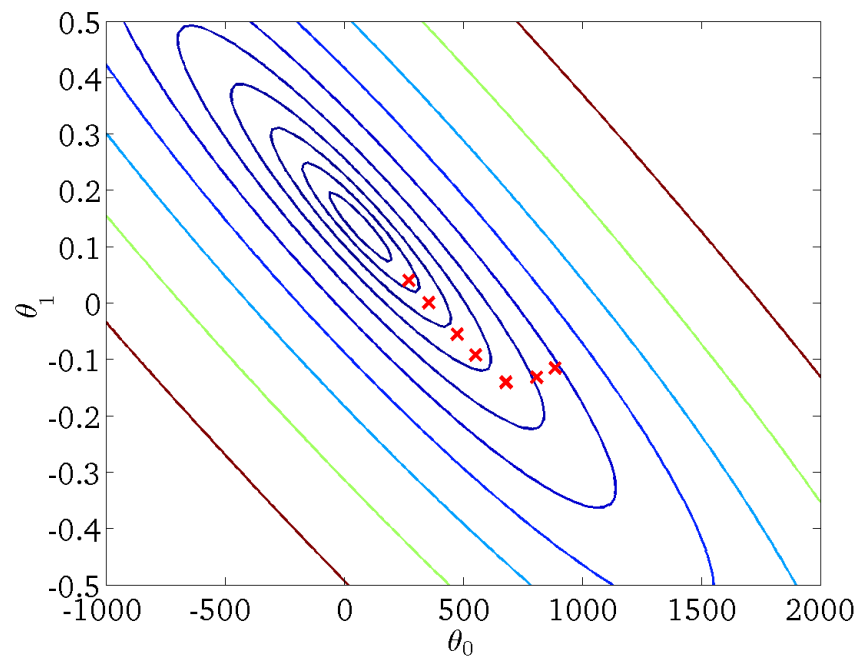
$$h_{\theta}(x)$$

(for fixed θ_0, θ_1 , this is a function of x)



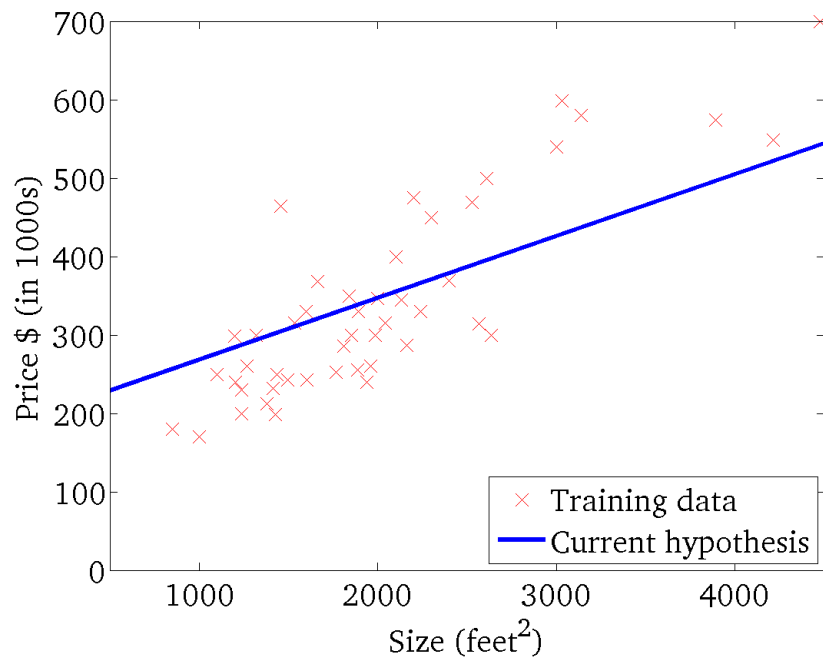
$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)



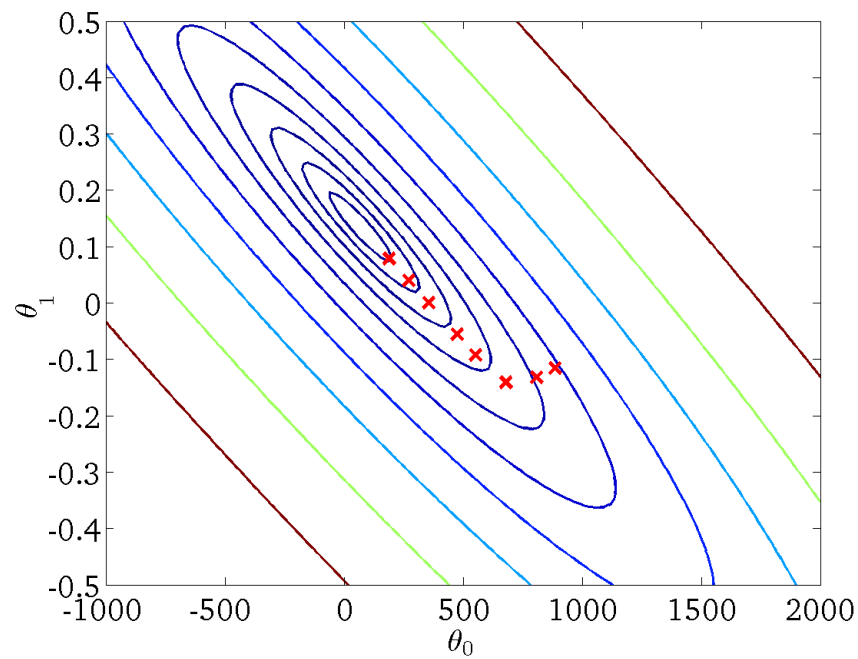
$$h_{\theta}(x)$$

(for fixed θ_0, θ_1 , this is a function of x)



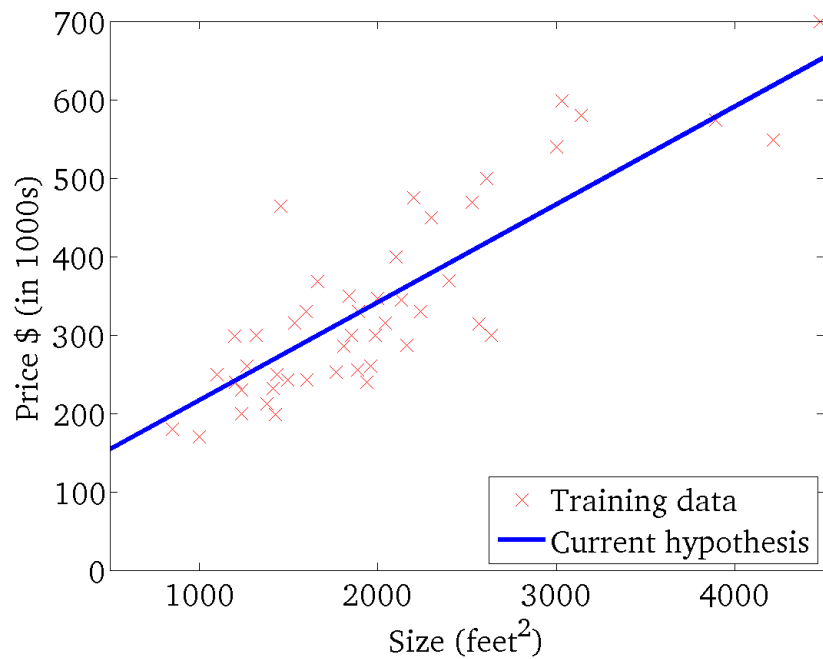
$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)



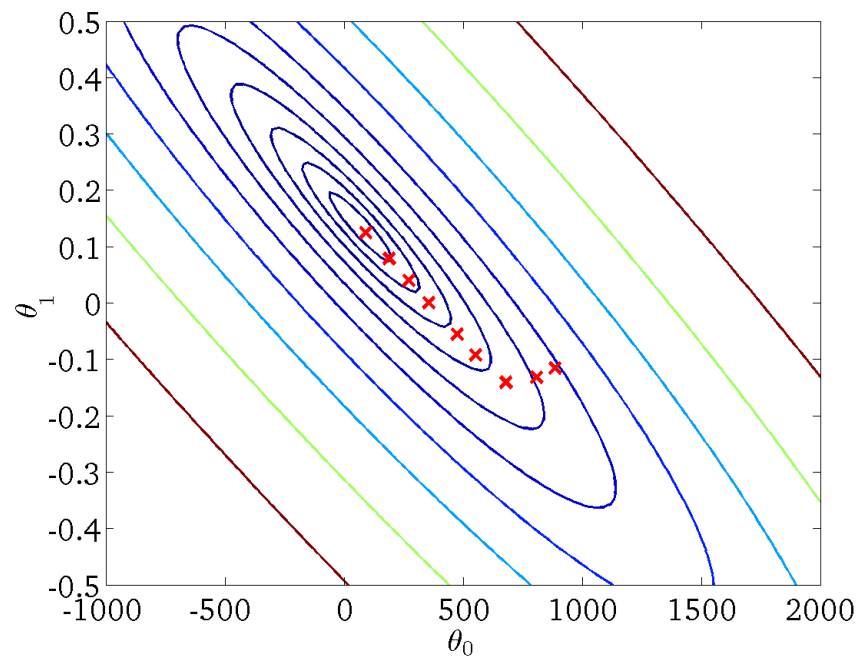
$$h_{\theta}(x)$$

(for fixed θ_0, θ_1 , this is a function of x)



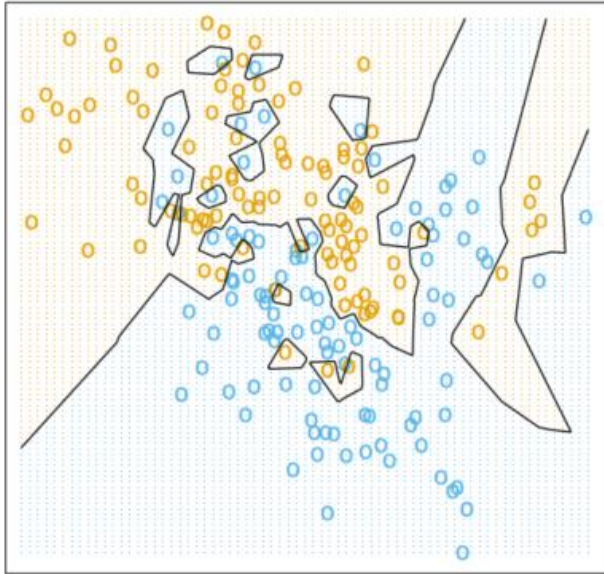
$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)

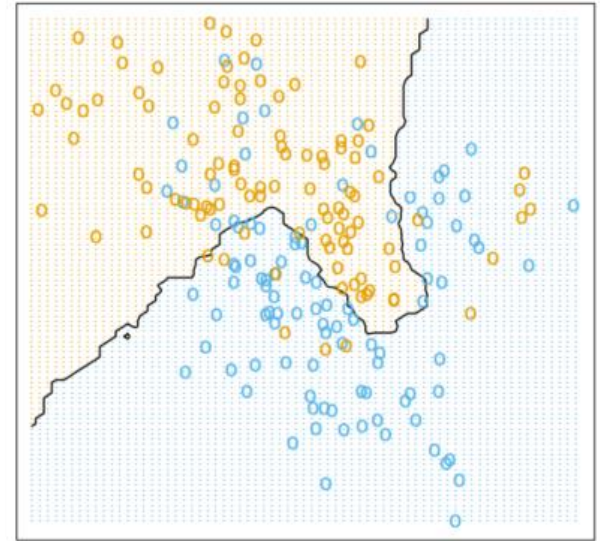


Linear Regression vs. k-Nearest Neighbours

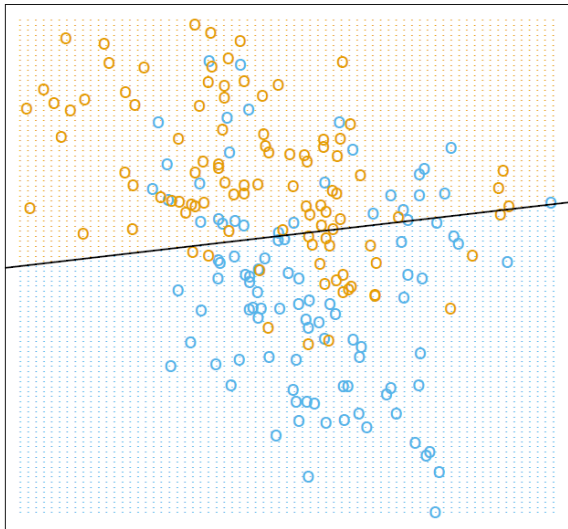
1-Nearest Neighbor Classifier



15-Nearest Neighbor Classifier



Linear Regression of 0/1 Response



Orange: $y = 1$
Blue: $y = 0$

Linear Regression vs. k-Nearest Neighbours

- Linear Regression: the boundary can only be linear
- Nearest Neighbours: the boundary can more complex
- Which is better?
 - Depends on what the *actual boundary* looks like
 - Depends on whether we have enough data to figure out the *correct* complex boundary