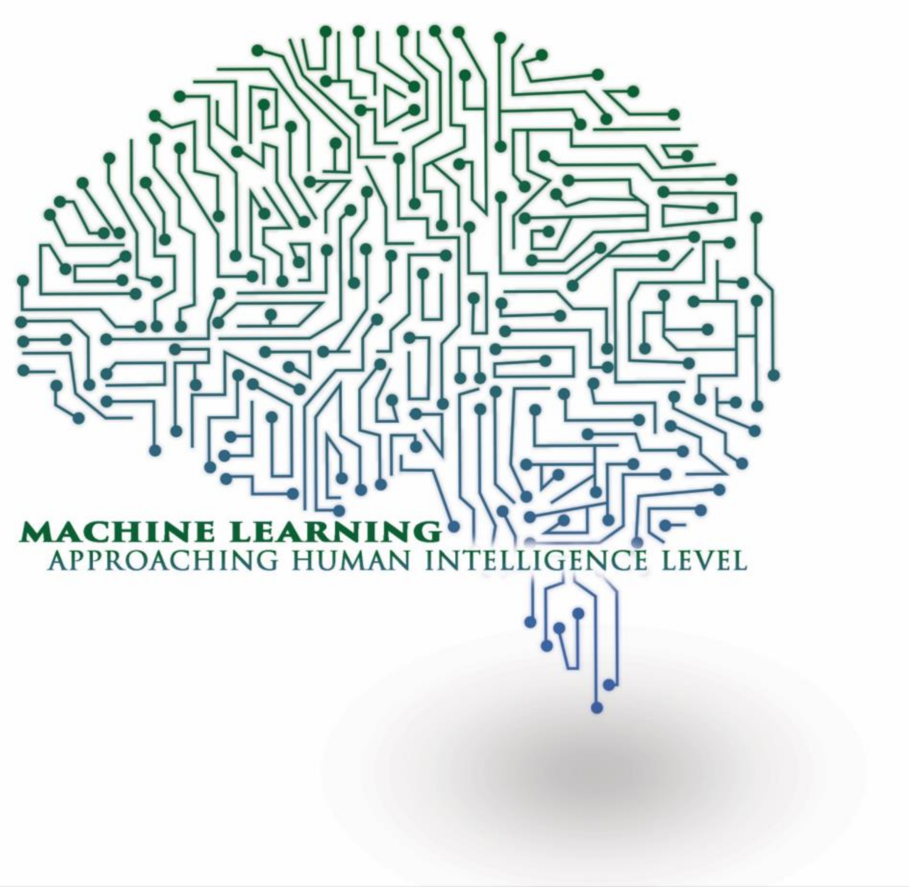
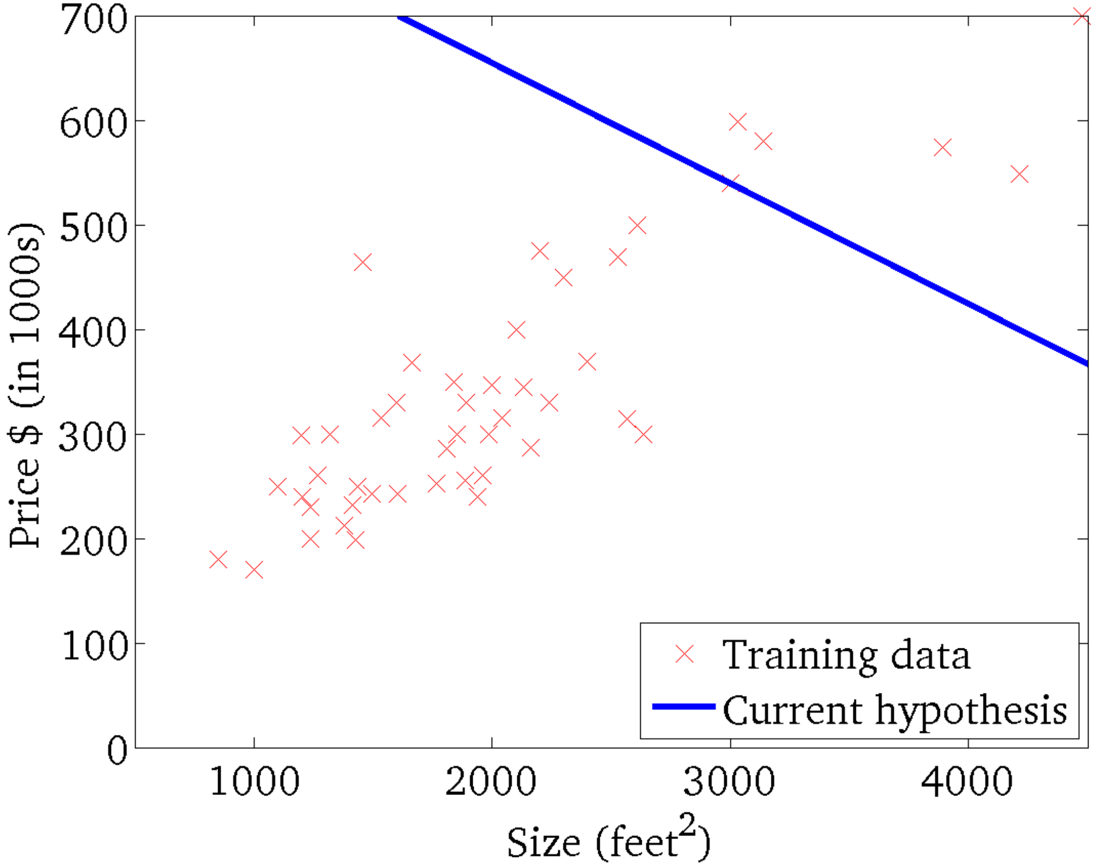


Logistic Regression & Introduction to Artificial neural networks.

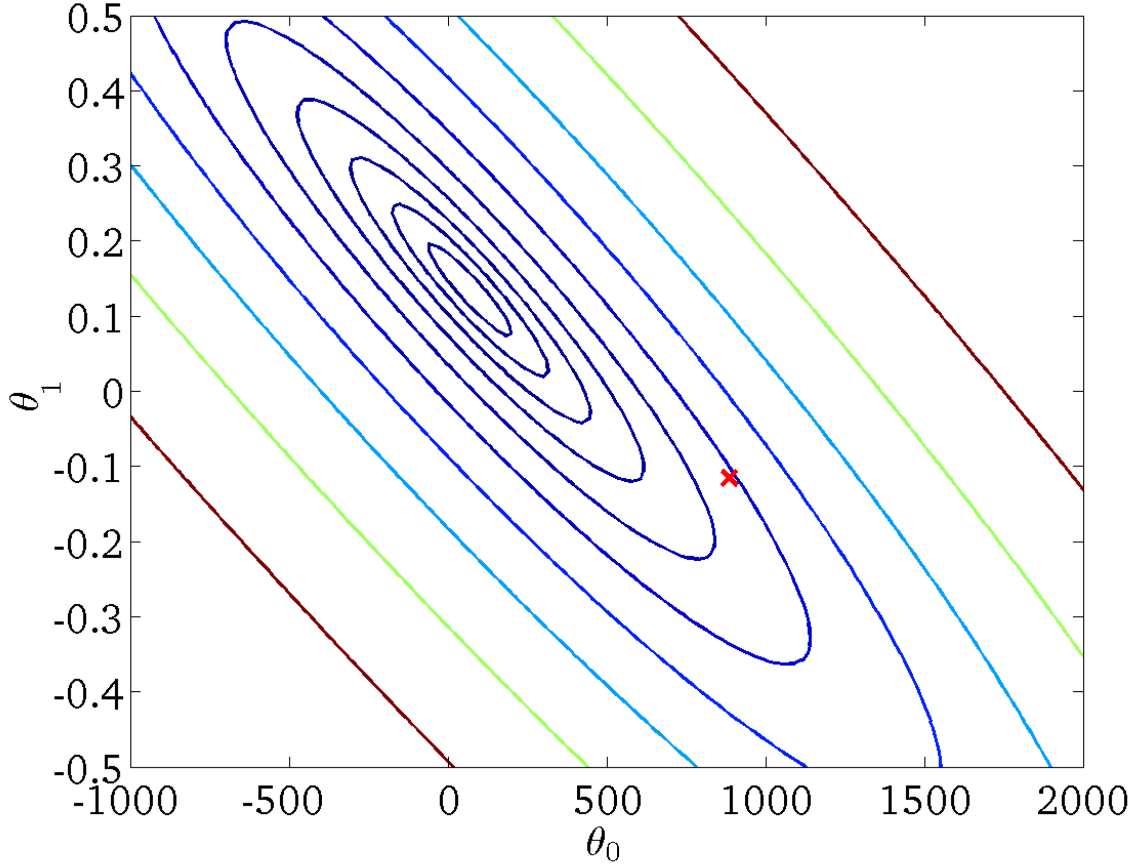
$$h_\theta(x)$$

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



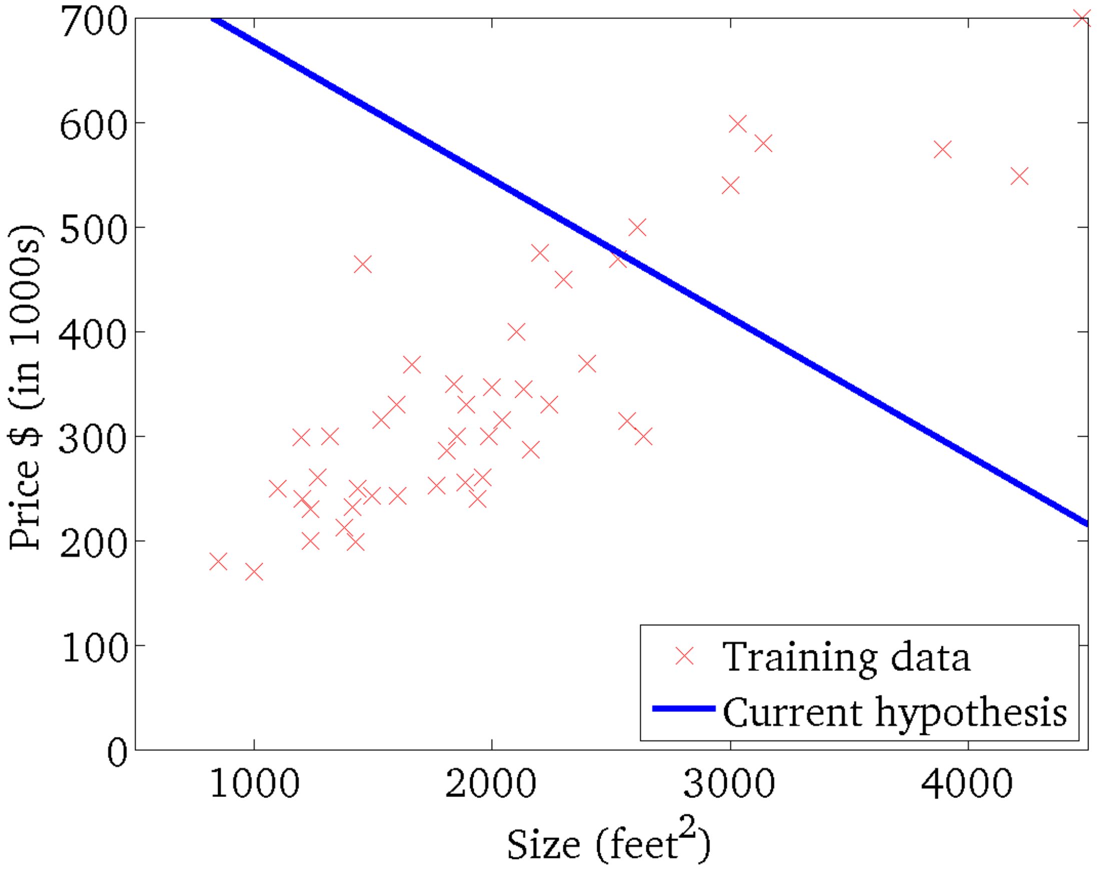
$$J(\theta_0, \theta_1)$$

(function of the parameters θ_0, θ_1)



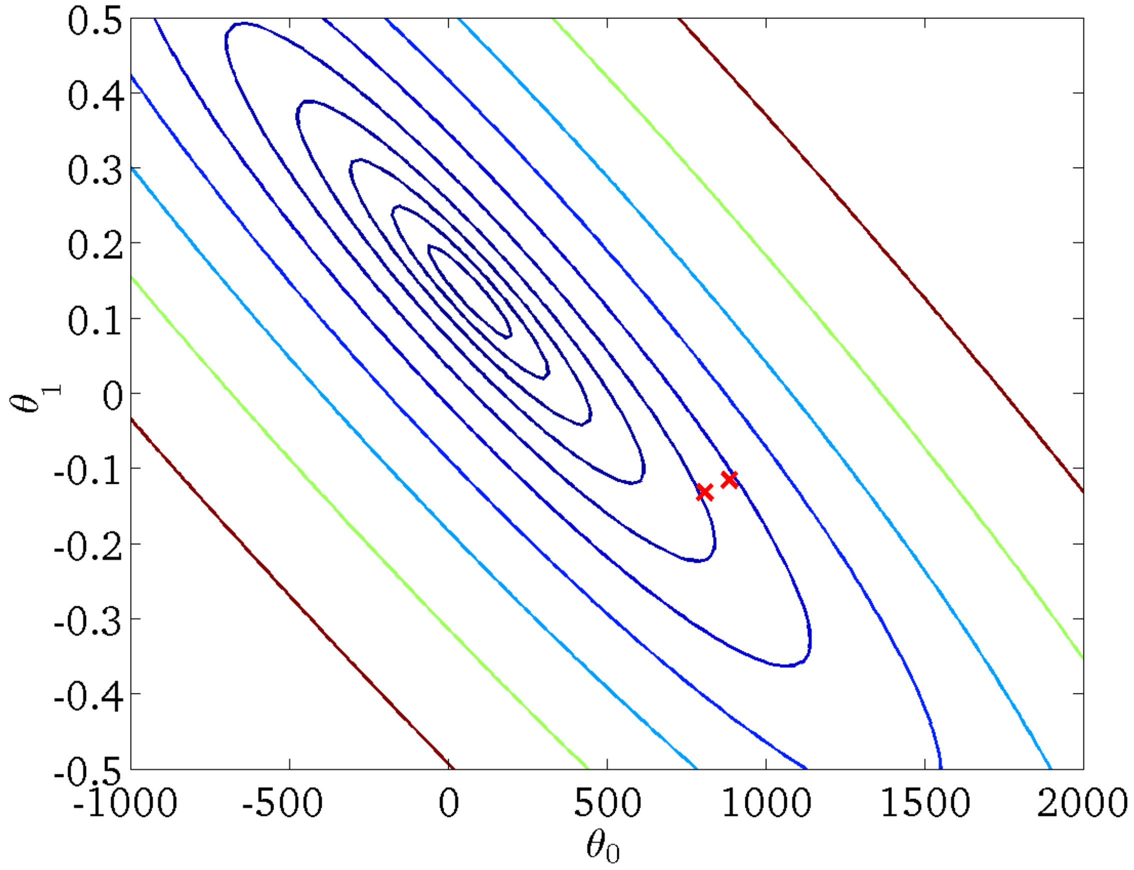
$h_\theta(x)$

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



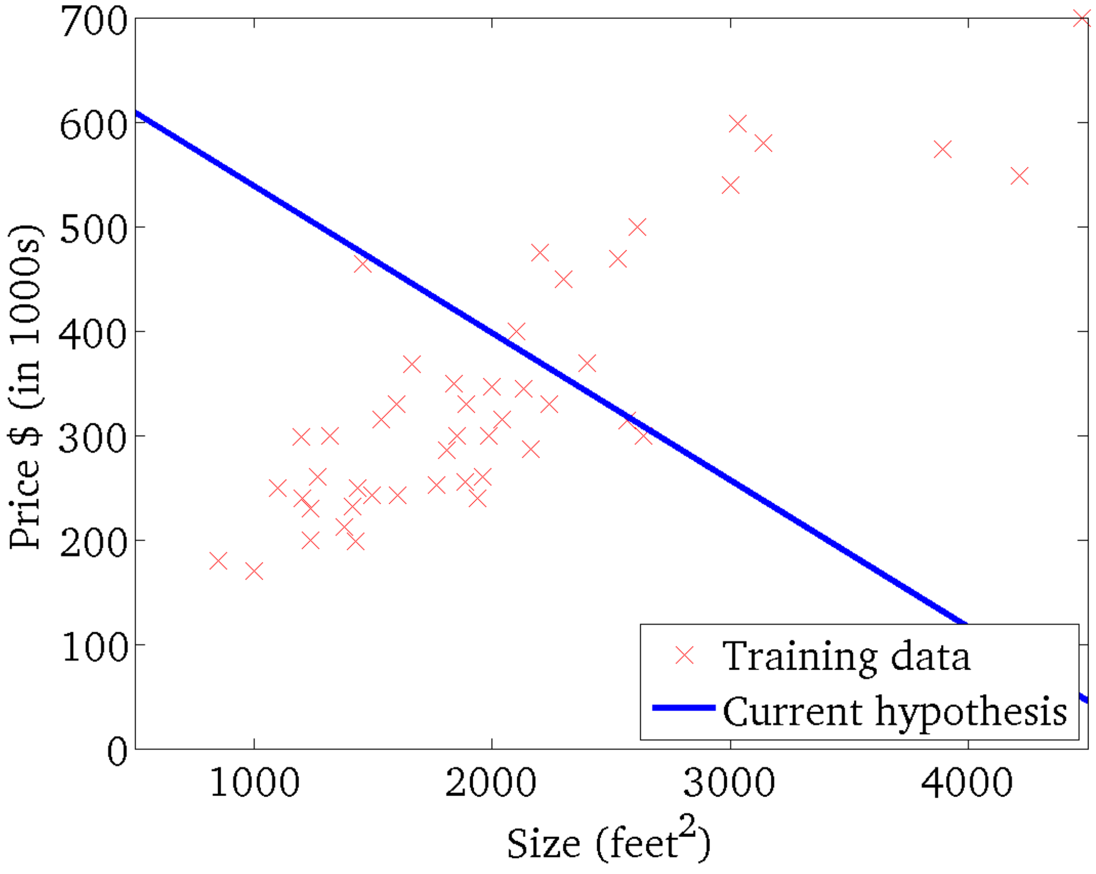
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



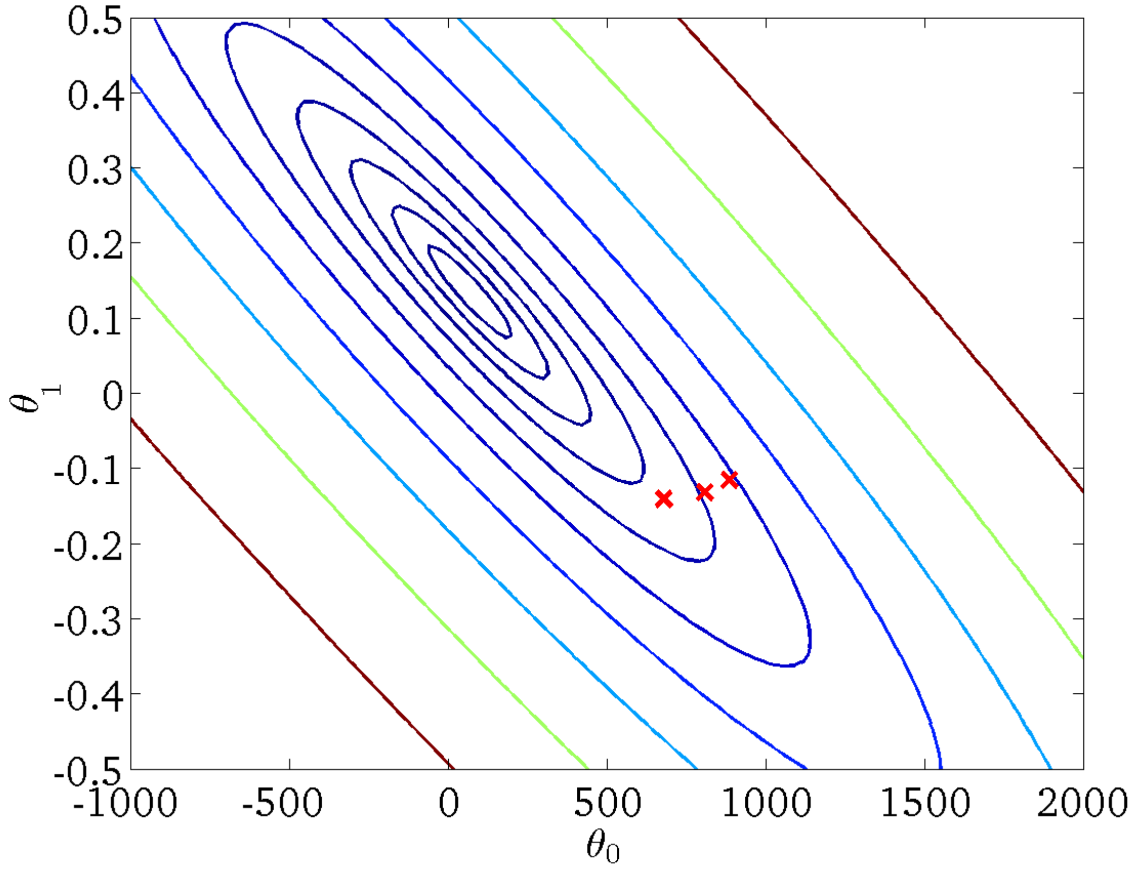
$h_\theta(x)$

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



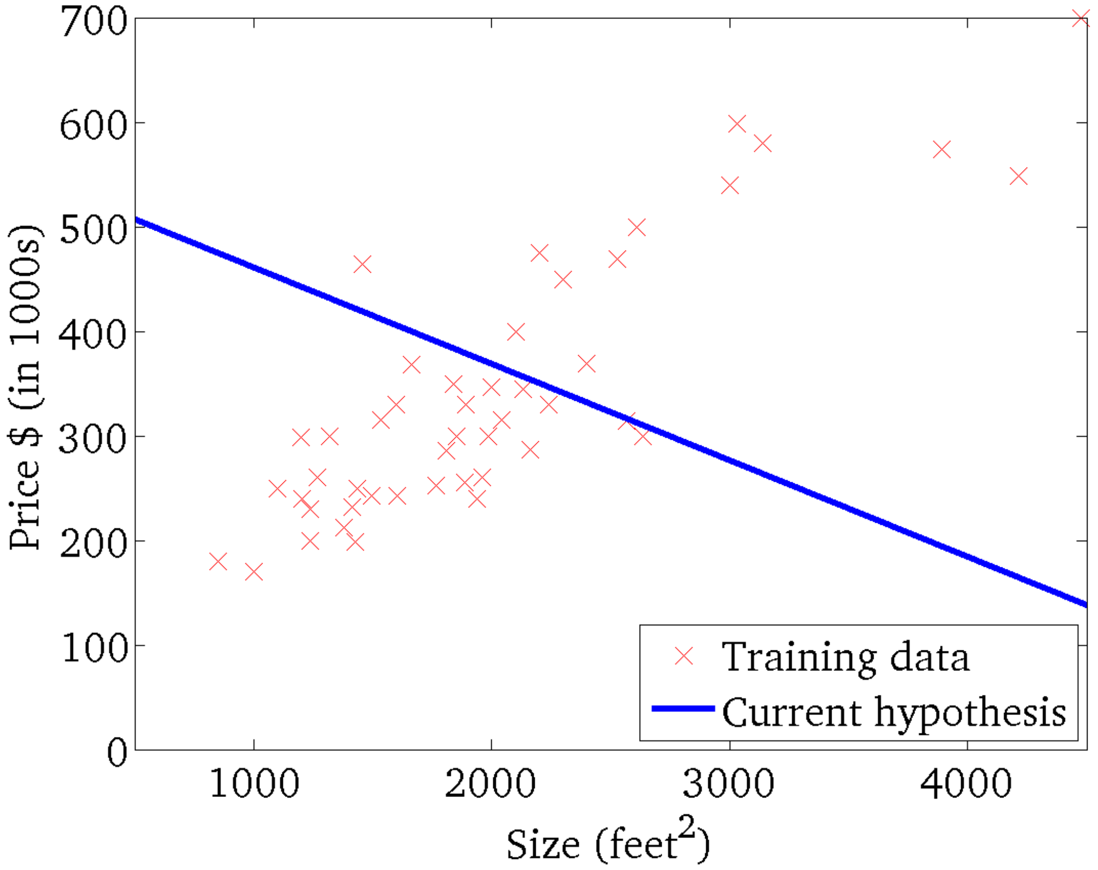
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



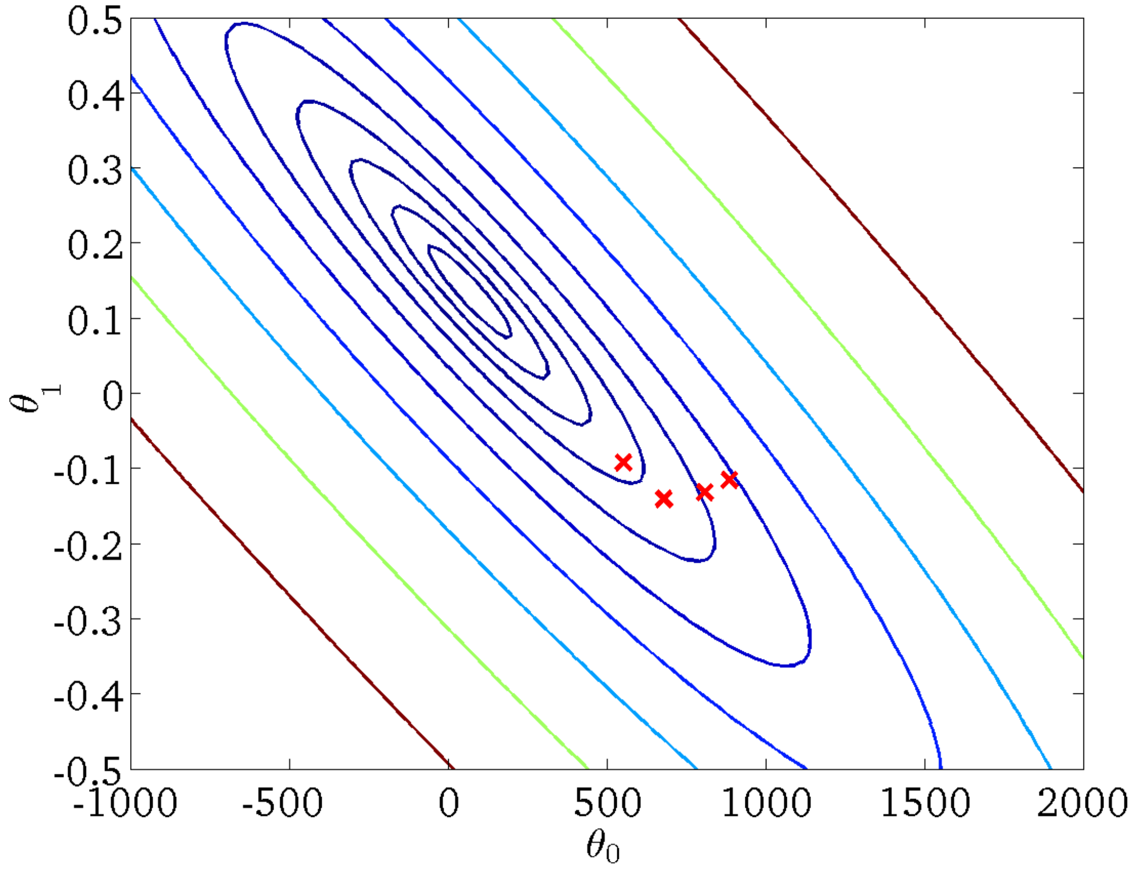
$h_\theta(x)$

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



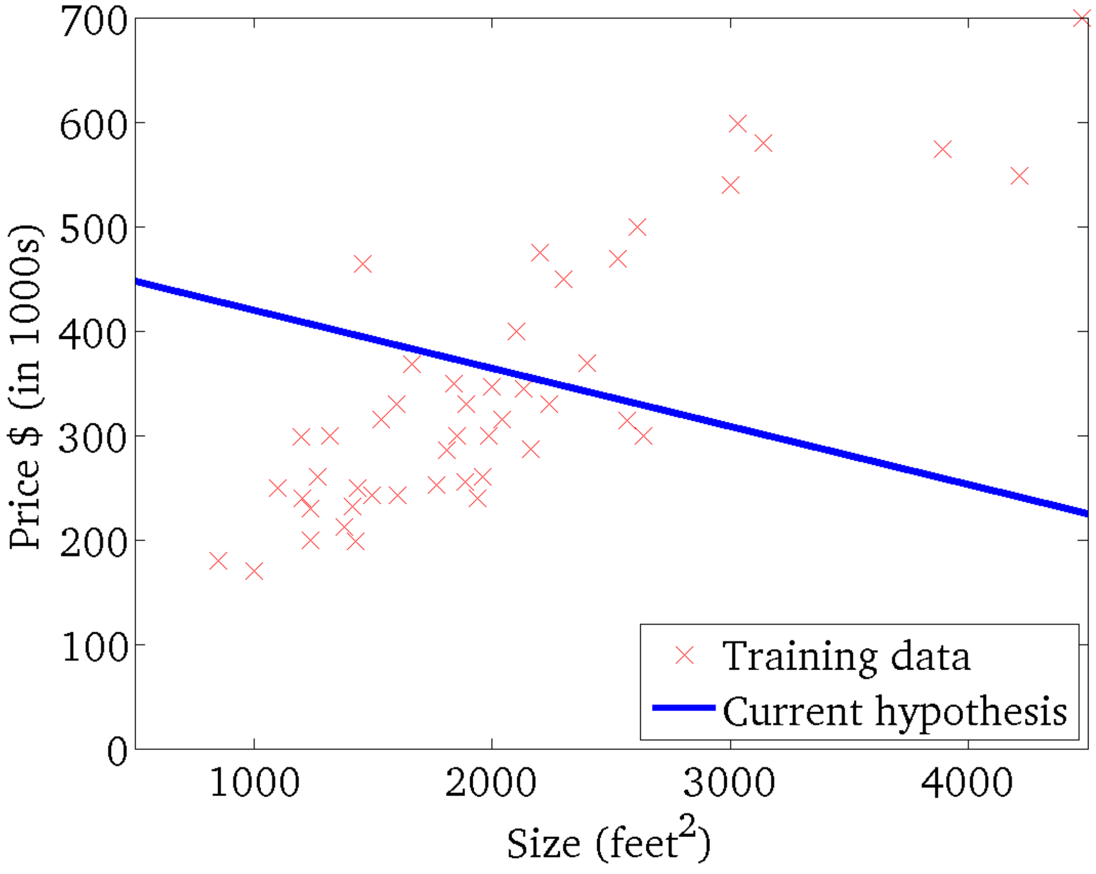
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



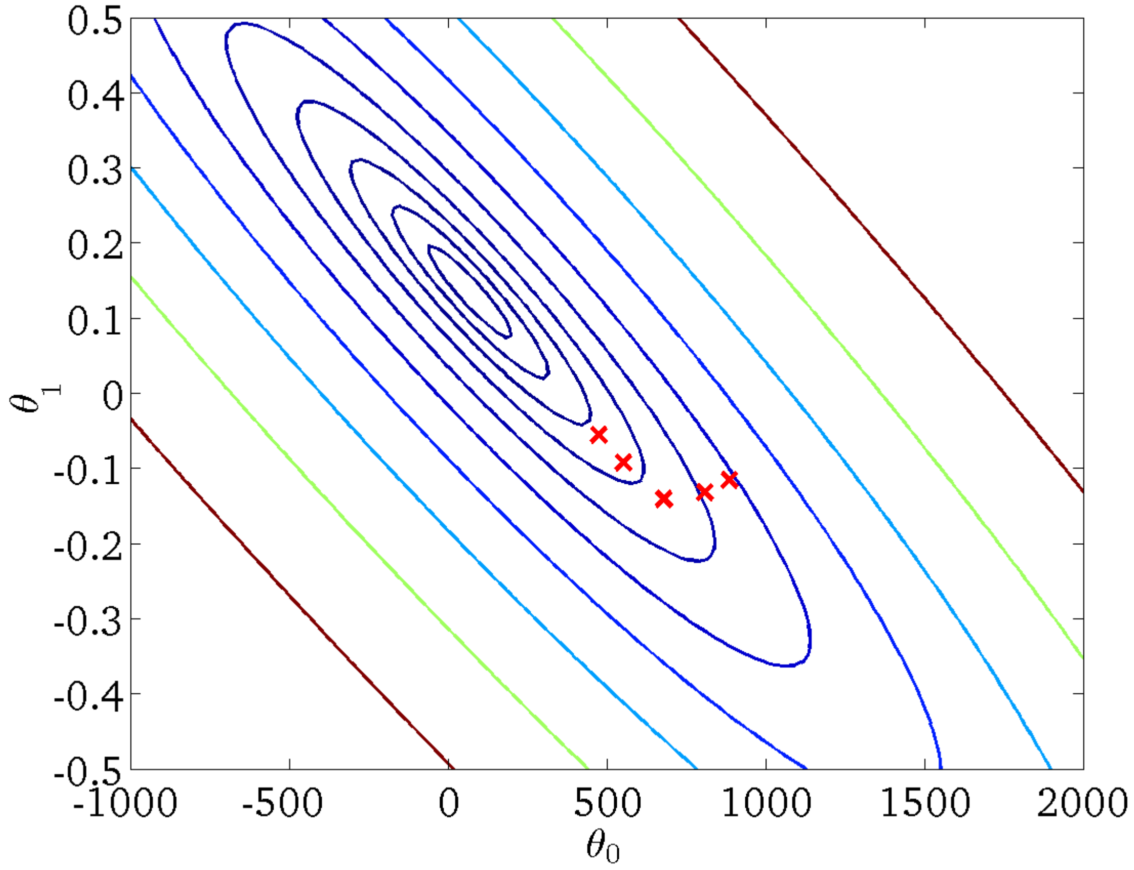
$h_\theta(x)$

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



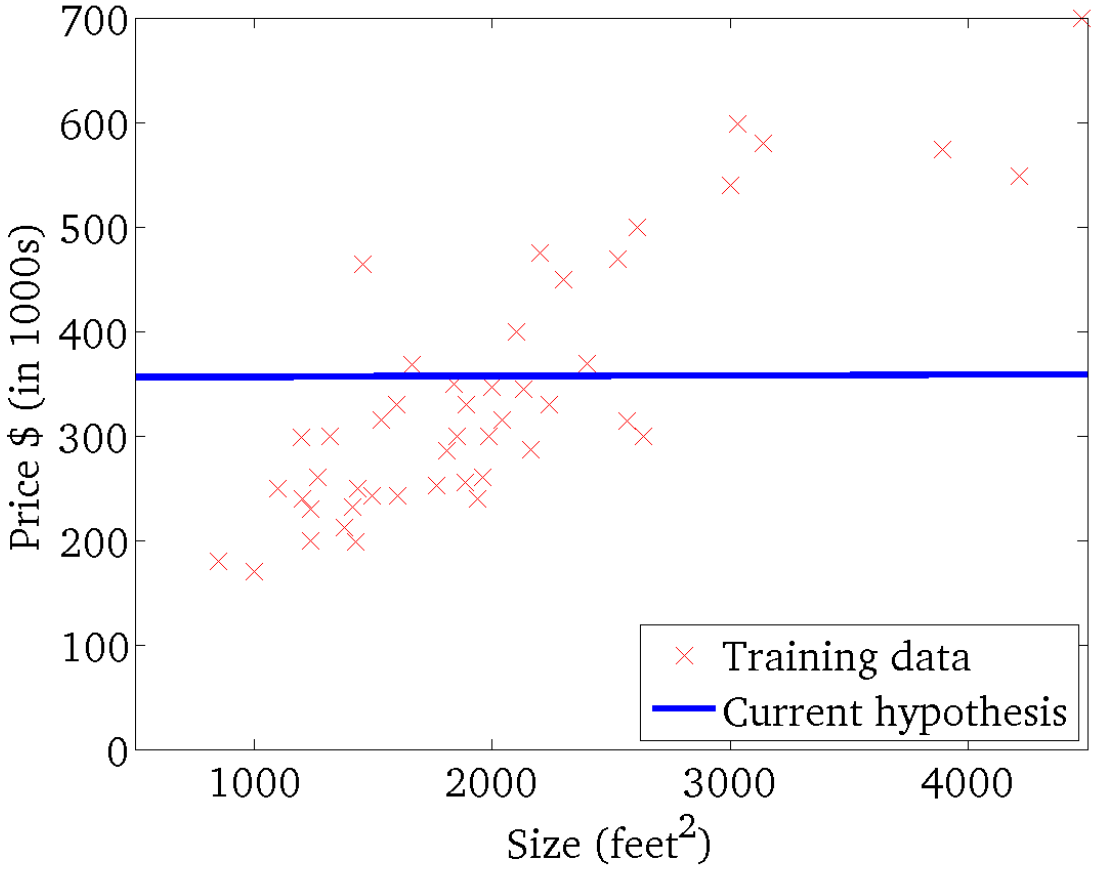
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



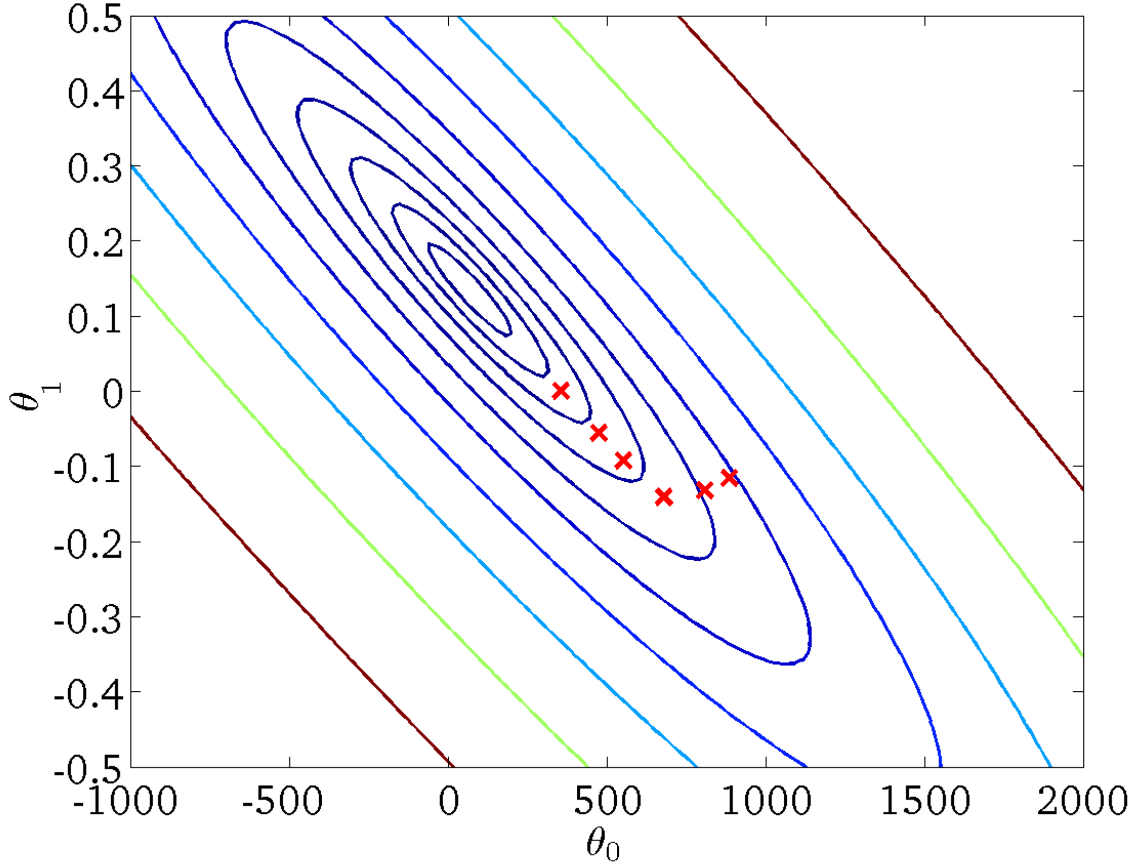
$h_\theta(x)$

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



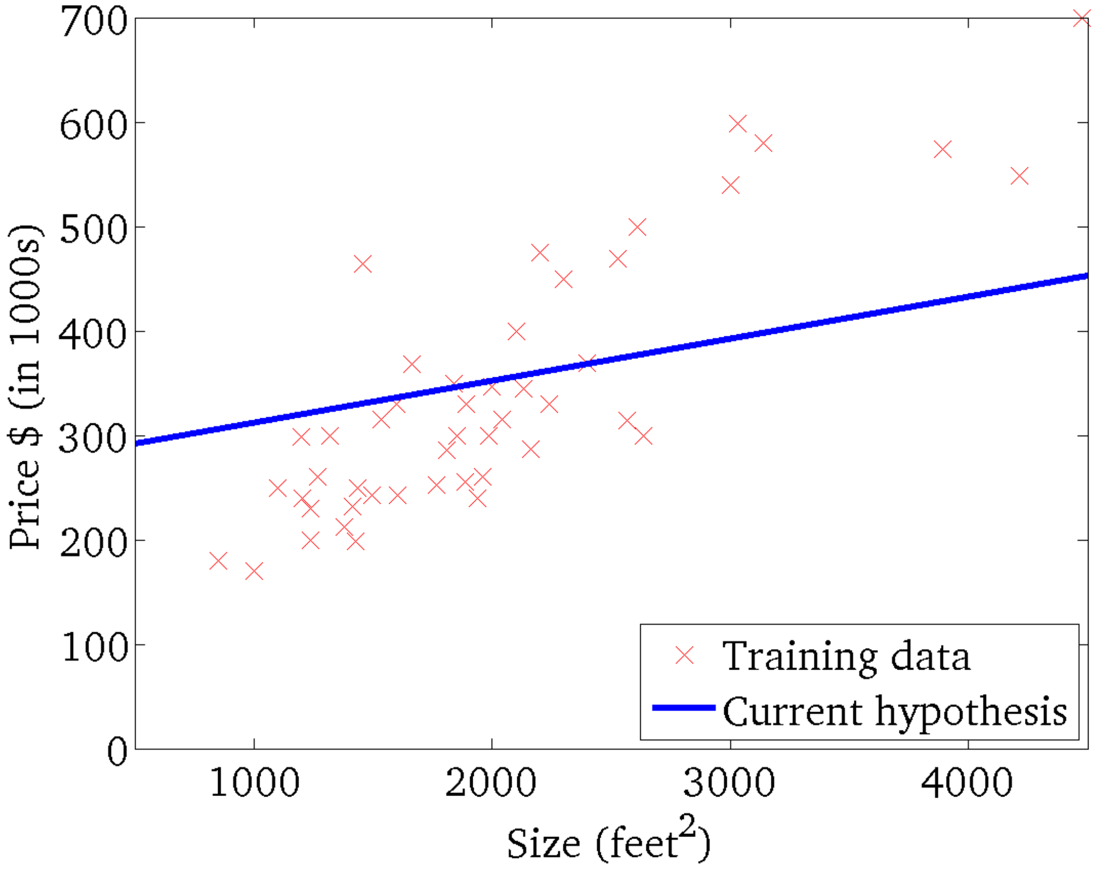
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



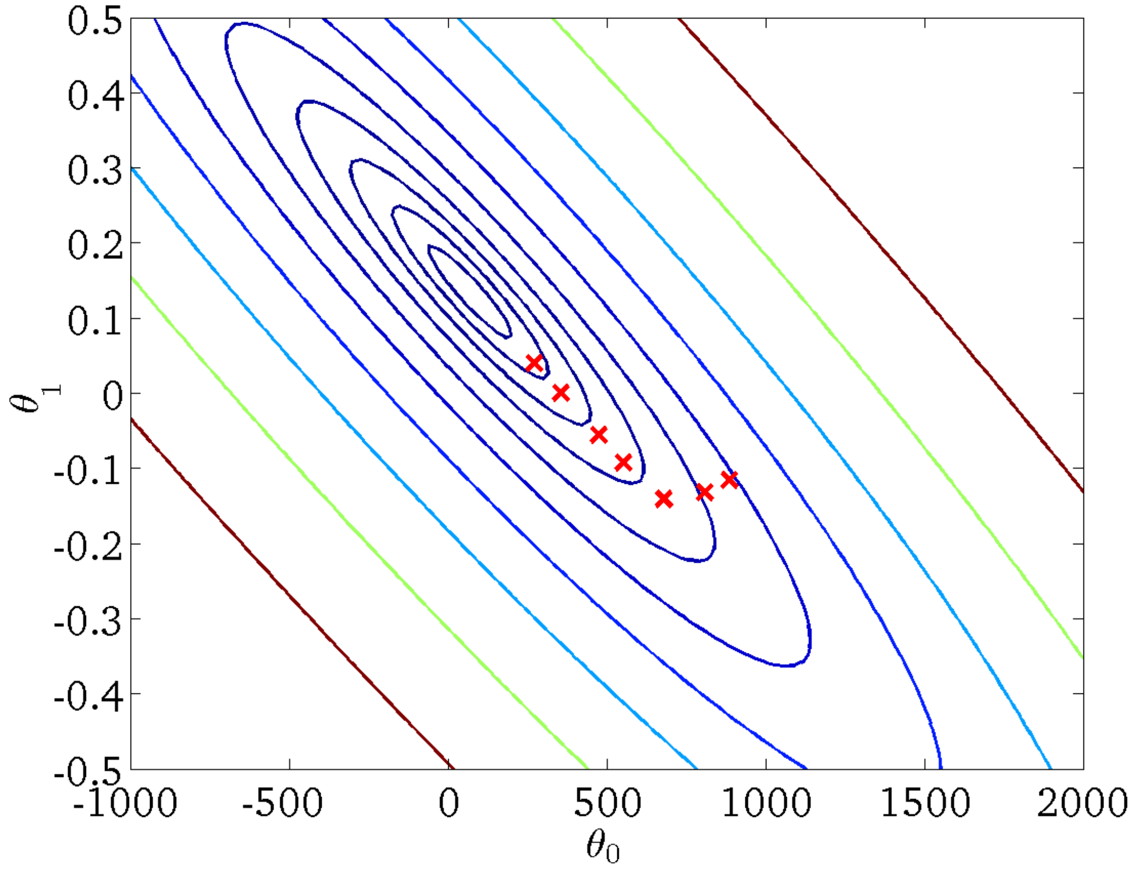
$h_\theta(x)$

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



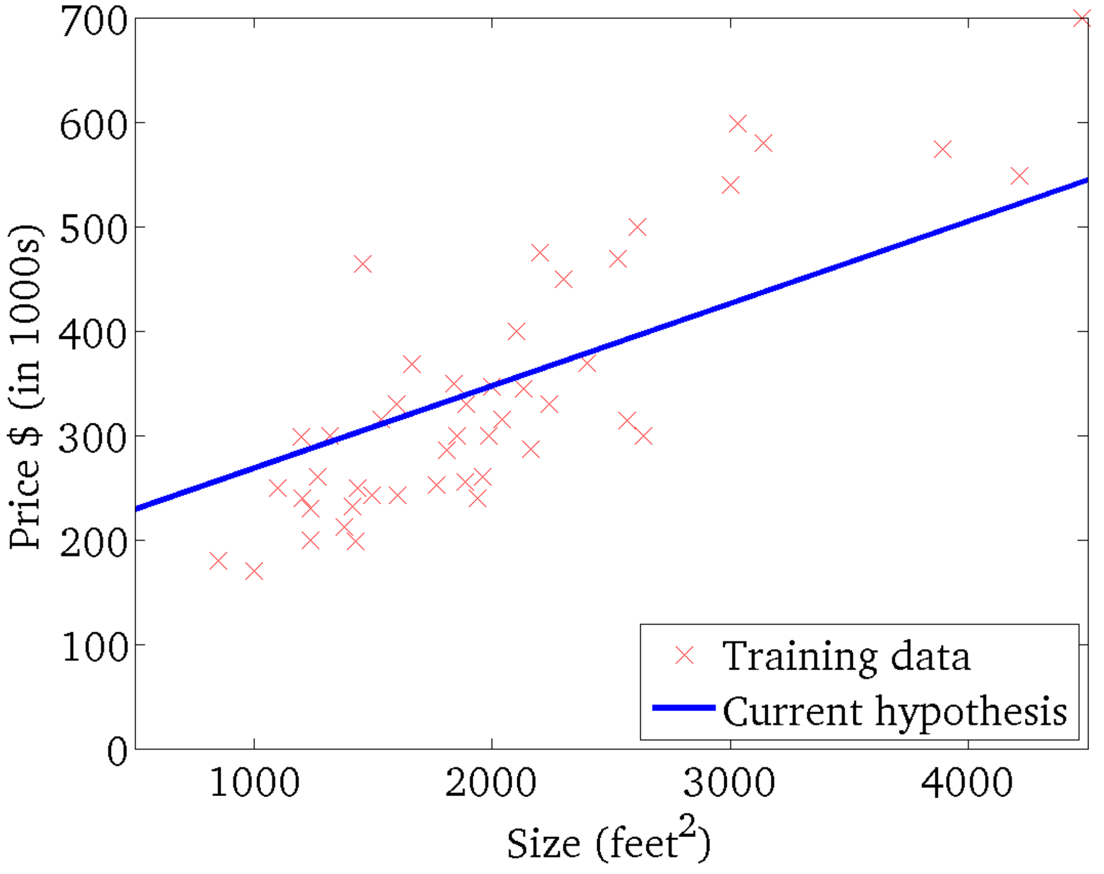
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



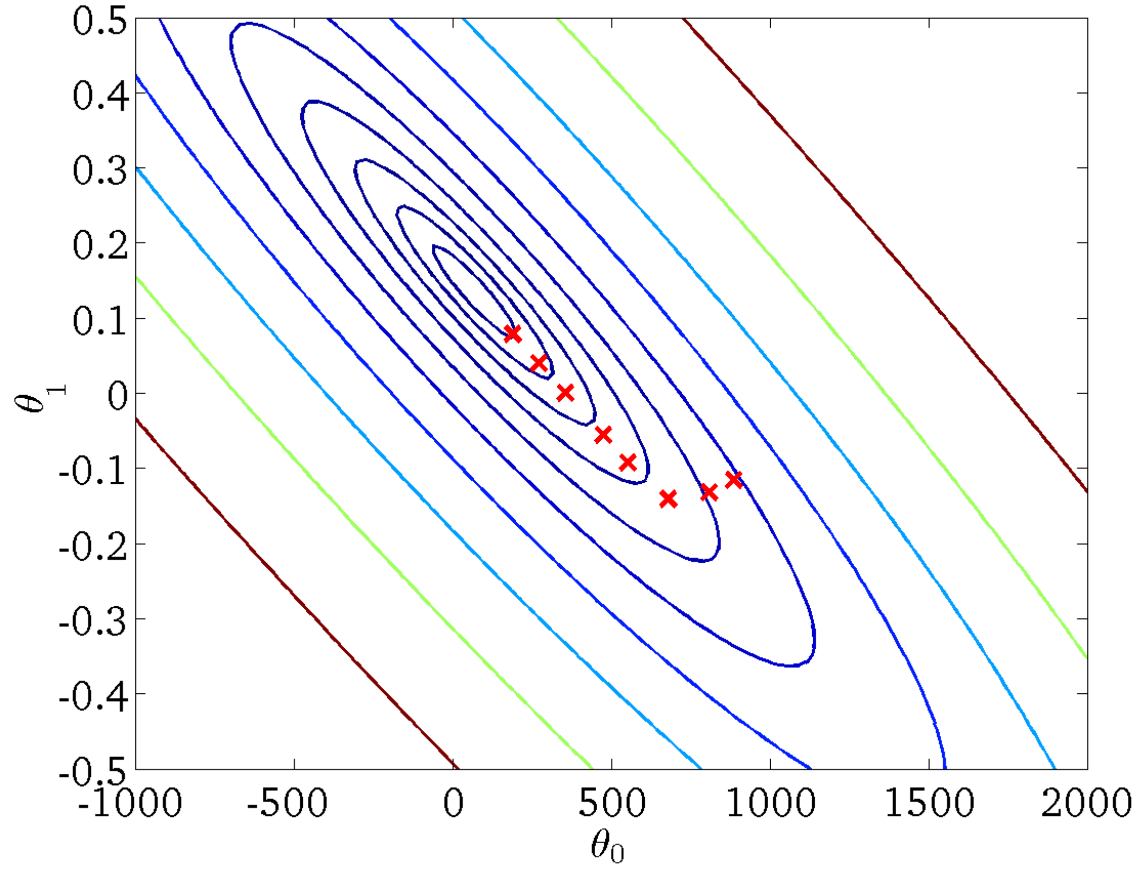
$h_\theta(x)$

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



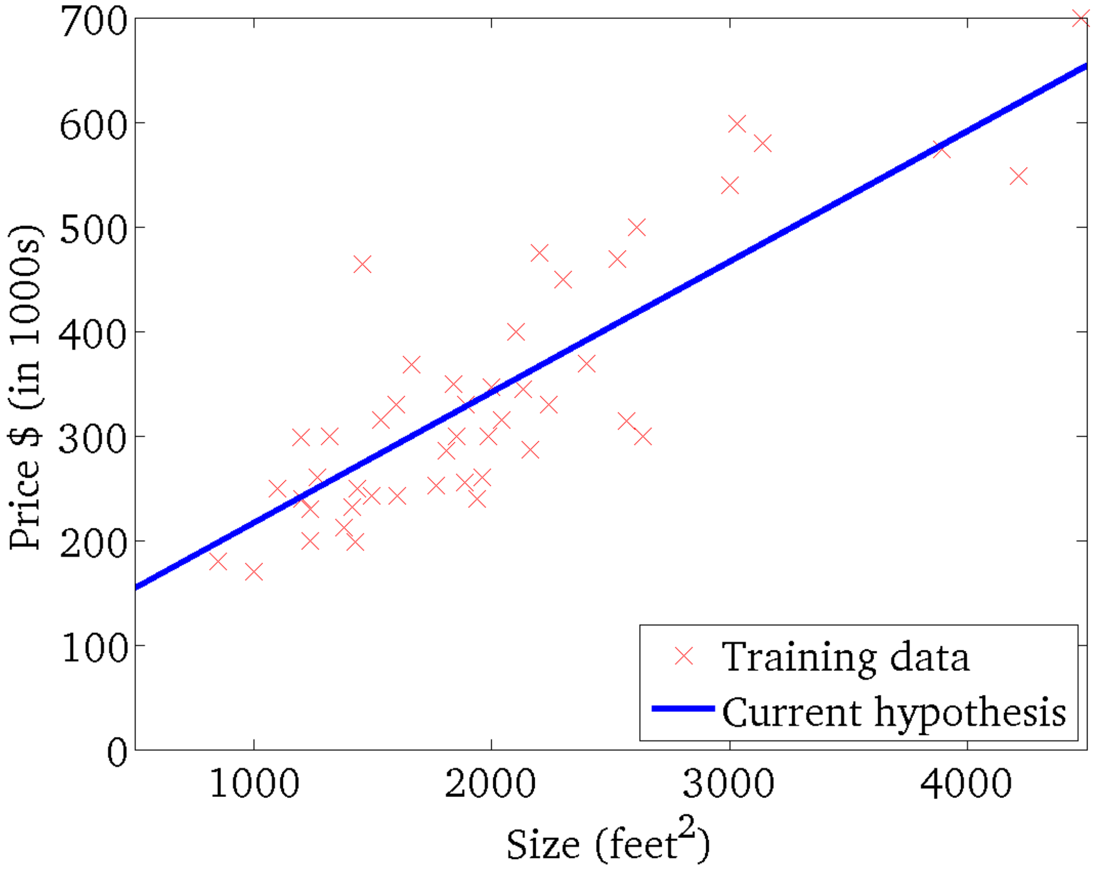
$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



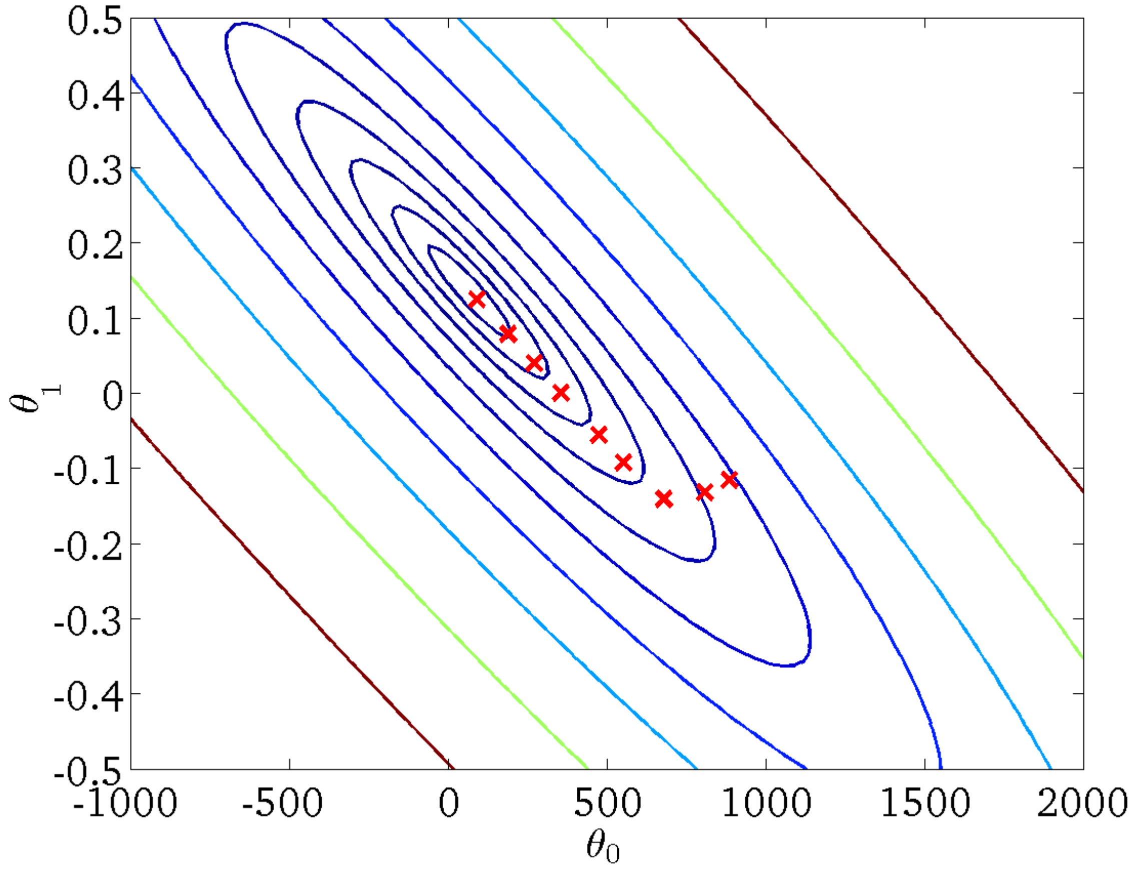
$h_\theta(x)$

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

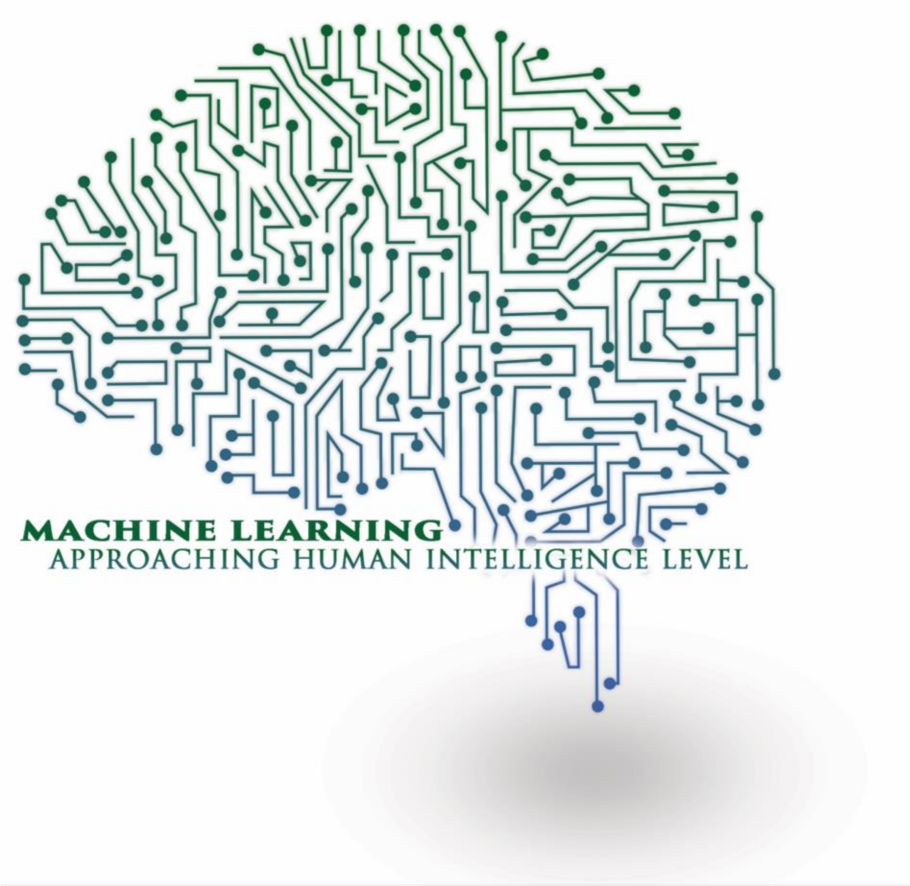


$J(\theta_0, \theta_1)$

(function of the parameters θ_0, θ_1)



Classification



Classification

- Reviews (sentiment): positive / negative
- Email: Spam / Not Spam
- User / customer type:
- Online Transactions: Fraudulent (Yes / No)
- Tumor: Malignant / Benign
- User Activity: laying / walking / setting / standing

$y \in \{0,1\}$

0: "*negative class*"

1: "*positive class*"

$y \in \{0,1,2,\ldots,k\}$

0: "*first class*"

1: "*second class*"

k: "*kth class*"

Classification $y \in \{0,1\}$

Hypothesis $h_\theta(x)$ *should be ether* 0 *or* 1

Simplification

$h_\theta(x)$ *should be between* 0 *and* 1

$h_\theta(x) \geq 0.5$ *then we predict* y *to be* 1

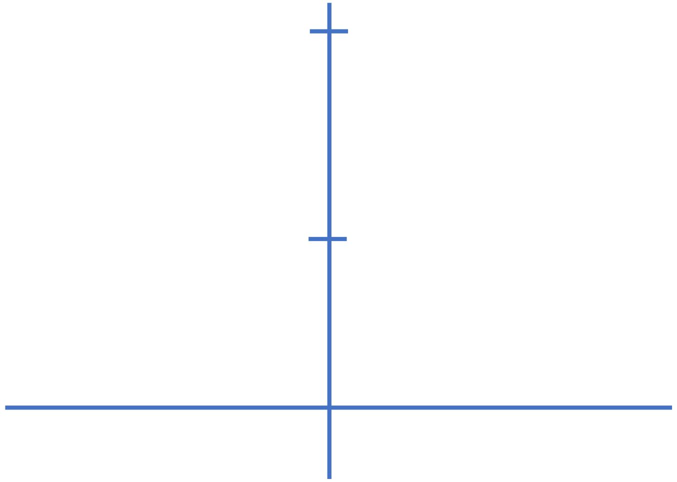
$h_\theta(x) < 0.5$ *then we predict* y *to be* 0

$$Logistic\ Regression:\ 0 \leq h_\theta(x) \leq 1$$

Logistic Regression Model

Sigmoid function
Logistics function

$$g(z) = \frac{1}{1+e^{-z}}$$



$$h_\theta(x) = g(\ \theta^T x\ )$$

Interpretation of Hypothesis Output

$h_\theta(x)$ = estimated probability that y = 1 on input x

$Example\ (tumor\ classification)\ if\!: X = \begin{bmatrix} 1 \\ tumor\ size \end{bmatrix}$

$if\ we\ get\ h_\theta(x) = 0.7\ then\ the\ tumor\ has\ 70\%\ probability\ to\ be\ mailgnent$

"probability that y = 1, given x, parameterized by θ"

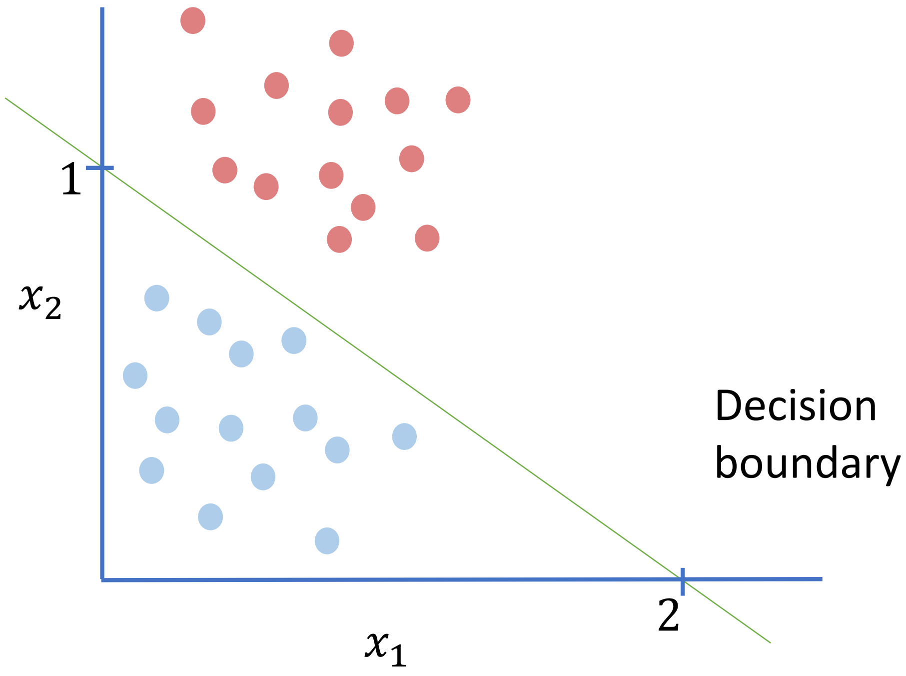
$P(y = 0\ |x; \theta) \equiv probability\ that\ the\ given\ x, \theta\ is\ in\ calss\ 0$

$P(y = 1\ |x; \theta) \equiv probability\ that\ the\ given\ x, \theta\ is\ in\ calss\ 0$

$$P(y=0|x;\theta) + P(y=1|x;\theta) = 1$$

$$P(y=0|x;\theta) = 1 - P(y=1|x;\theta)$$

What Logistic Regression really do?

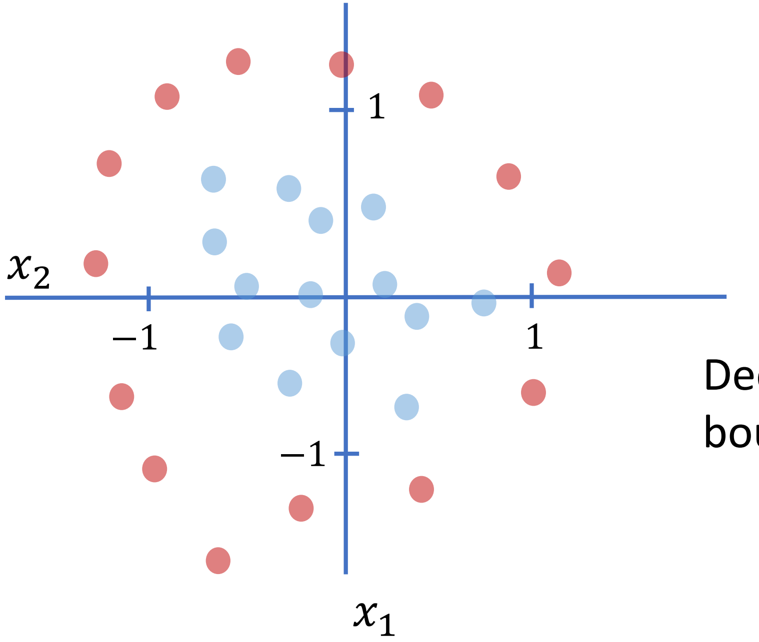


$$h_\theta(z) = h(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

$$Predict\ y = 1\ if\ -2 + x_1 + 2\ x_2 \geq 0$$

In non-linear case

$$h_\theta(z) = h(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2 + \theta_5 x_1 x_2)$$



Decision boundary

$$Predict\ y = 1\ if\ -1 + x_1^2 + x_2^2 \geq 0$$

Logistics Regression Cost function

Recall from linear regression:

$$J(\bar{\theta}\ ) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta(x_i)\ - y_i)^2$$

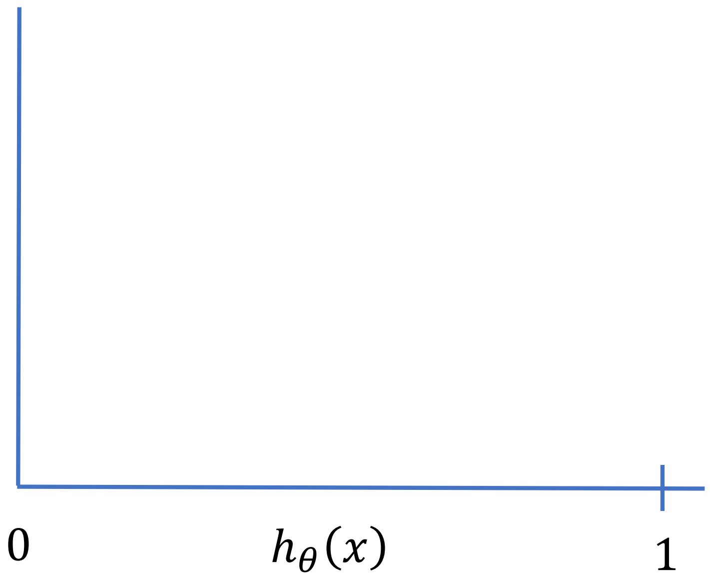
In logistic regression we have to cases:

- when $\mathrm{y} = 1\ cost(h_\theta(x_i), y)\ should\ be\ low\ if\, h_\theta(x_i) \geq 0.5$ and height if $h_\theta(x_i) < 0.5$

- when $\mathrm{y} = 0\ cost(h_\theta(x_i), y)\ should\ be\ low\ if\, h_\theta(x_i) < 0.5$ and height if $h_\theta(x_i) \geq 0.5$

From calculus we do have a function with similar behaviour :

$$Cost(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & , if\ y = 1 \\ -\log(1 - h_\theta(x)) & , if\ y = 0 \end{cases}$$



Cost $= 0$ if $y = 1, h_\theta(x) = 1$

But as $\quad h_\theta(x) \to 0$

$\qquad\qquad Cost \to \infty$

Captures intuition that if $h_\theta(x) = 0$, (predict $P(y = 1|x;\theta) = 0$), but $y = 1$, we'll penalize learning algorithm by a very large cost.

Logistic regression cost function

$$J(\theta\ ) = \frac{1}{m}\sum_{i=1}^{m} Cost(h_\theta(x_i), y_i)$$

$$J(\theta\ ) = -\frac{1}{m}\sum_{i=1}^{m} y_i(\log\left(h_\theta(x_i)\right) + (1 - y_i)log(1 - h_\theta(x_i))$$

$$\min_\theta J(\theta\ )$$

Gradient decent:

$repeat\ until\ convergence\ \{$

$$\theta_i \coloneqq \theta_i\ -\ \alpha\frac{\partial J(\theta\ )}{\partial\theta_i}$$

$$(for\ i = 1,0)$$

$\}$

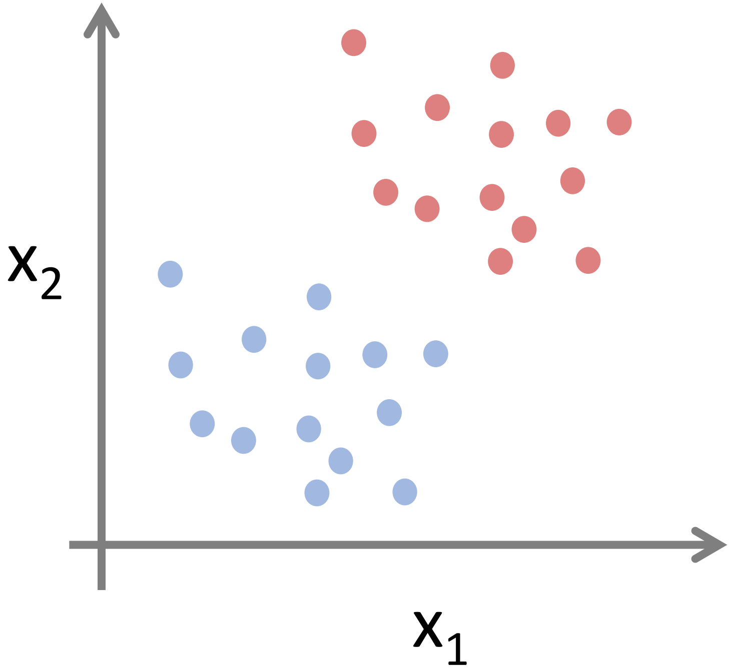
Multi-class classification: One-vs-all

Weather: Sunny, Cloudy, Rain, Snow

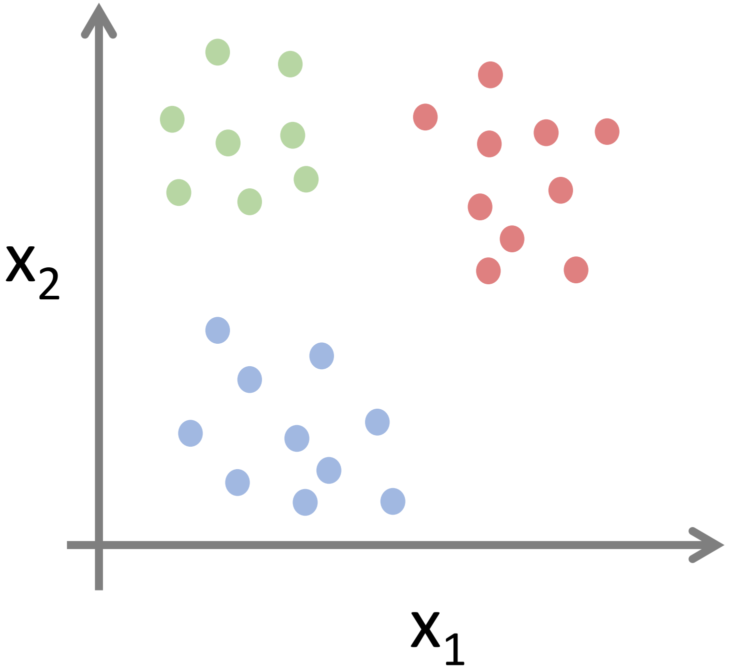
User Activity: laying / walking / setting / standing

Medical diagrams: Not ill, Cold, Flu

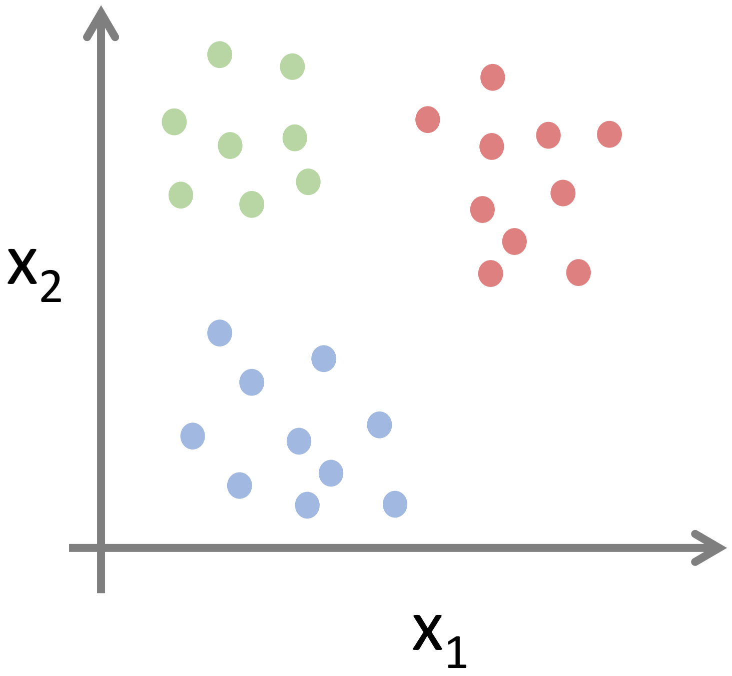
Binary-class classification:

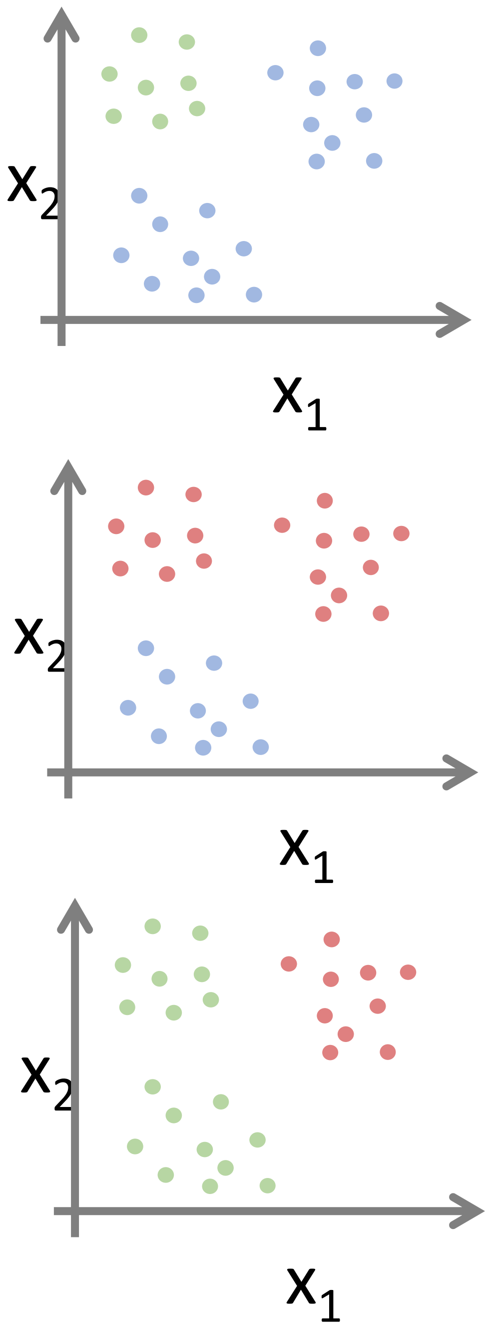


Multi-class classification:



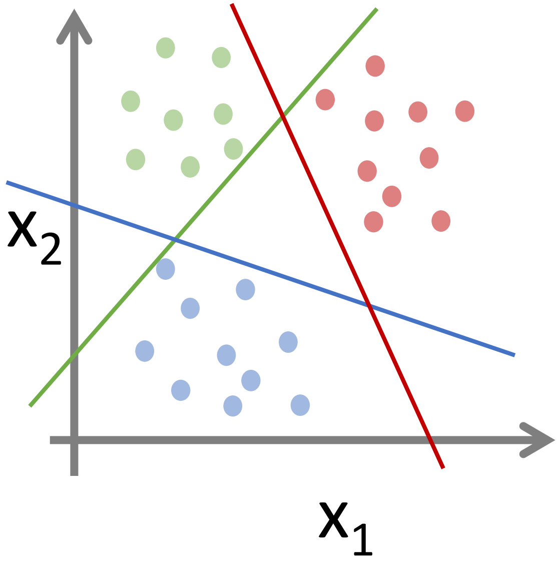
One-vs-all (one-vs-rest):





One-vs-all

Train a logistic regression classifier $h_\theta{}^{(i)}(x)$ for each class i to predict the probability that $y = i$.



On a new input x, to make a prediction, pick the class i that maximizes

$$\max_i h_\theta^{(i)}(x)$$

ANACONDA®

jupyter