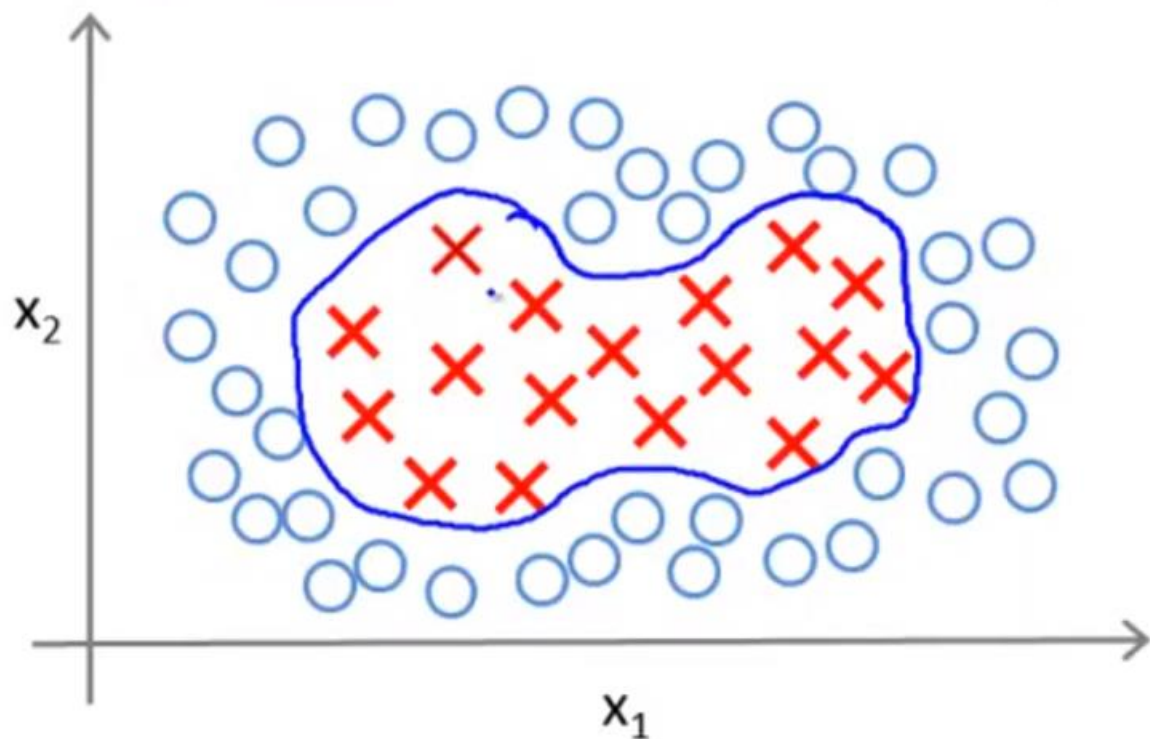


# Kernels-1

*Kernels*

Support Vector Machines

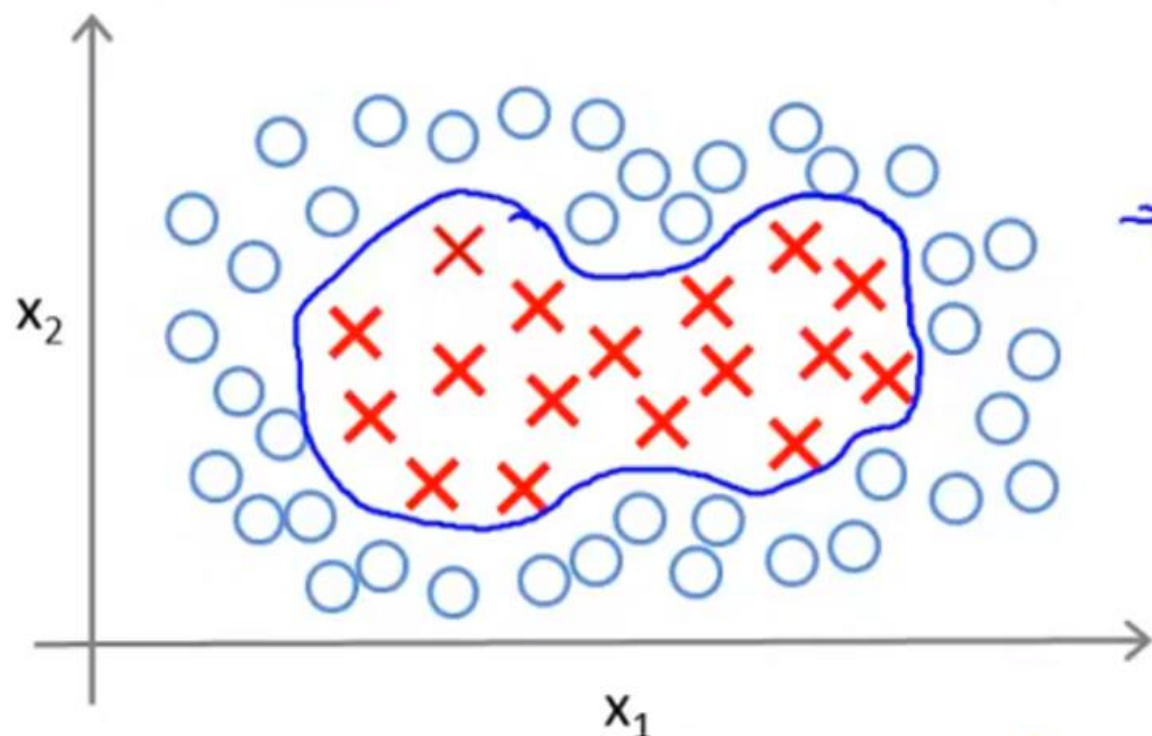
## Non-linear Decision Boundary



Predict  $y = 1$  if

$$\begin{aligned} \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 \\ + \theta_4 x_1^2 + \theta_5 x_2^2 + \dots \geq 0 \end{aligned}$$

## Non-linear Decision Boundary



Predict  $y = 1$  if

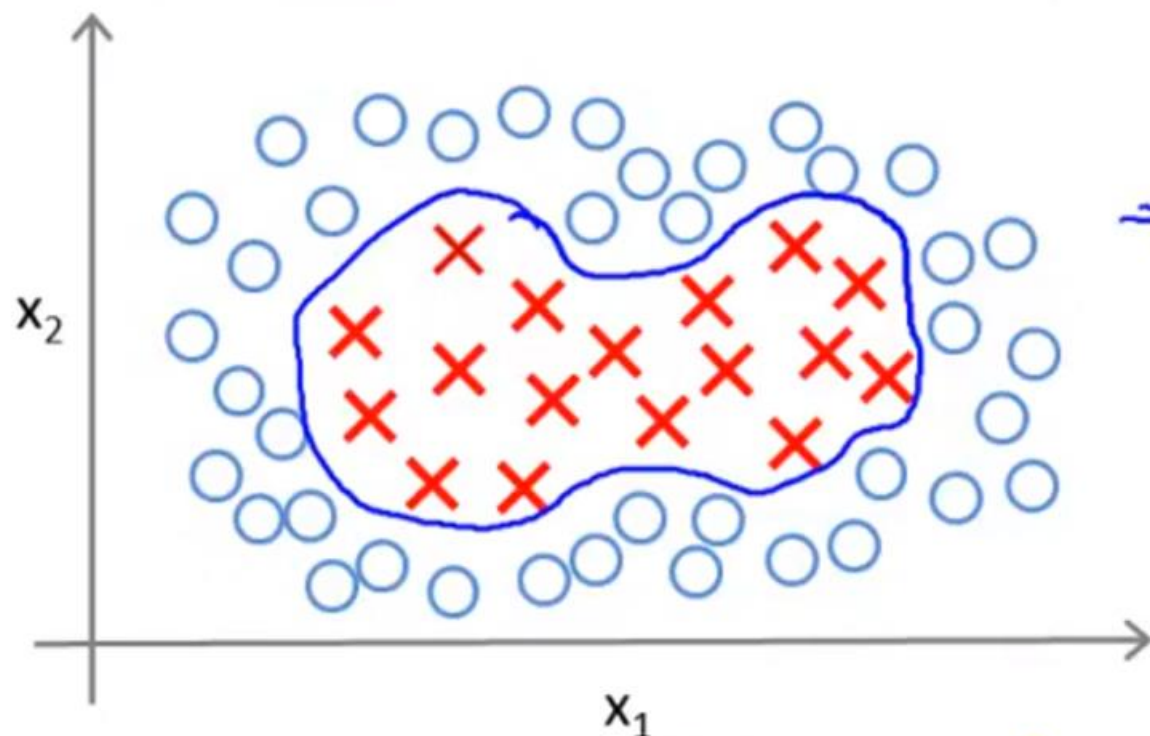
$$\rightarrow \theta_0 + \theta_1 \underline{x_1} + \theta_2 \underline{x_2} + \theta_3 \underline{x_1 x_2} \\ + \theta_4 \underline{x_1^2} + \theta_5 \underline{x_2^2} + \dots \geq 0$$

$$h_{\theta}(x) = \begin{cases} 1 & \text{if } \theta_0 + \theta_1 x_1 + \dots \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$\theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 + \dots$$

$$f_1 = x_1, \quad f_2 = x_2, \quad f_3 = x_1 x_2, \quad f_4 = x_1^2, \quad f_5 = x_2^2, \dots$$

## Non-linear Decision Boundary



Predict  $y = 1$  if

$$\rightarrow \theta_0 + \theta_1 \underline{x_1} + \theta_2 \underline{x_2} + \theta_3 \underline{x_1 x_2} \\ + \theta_4 \underline{x_1^2} + \theta_5 \underline{x_2^2} + \dots \geq 0$$

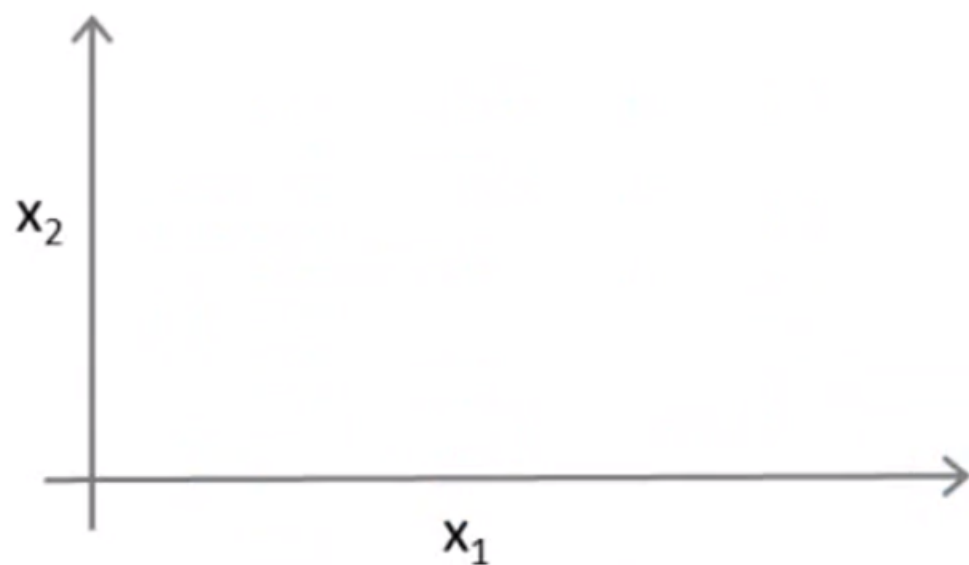
$$h_{\theta}(x) = \begin{cases} 1 & \text{if } \theta_0 + \theta_1 x_1 + \dots \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$\theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 + \dots$$

$$f_1 = x_1, \quad f_2 = x_2, \quad f_3 = x_1 x_2, \quad f_4 = x_1^2, \quad f_5 = x_2^2, \dots$$

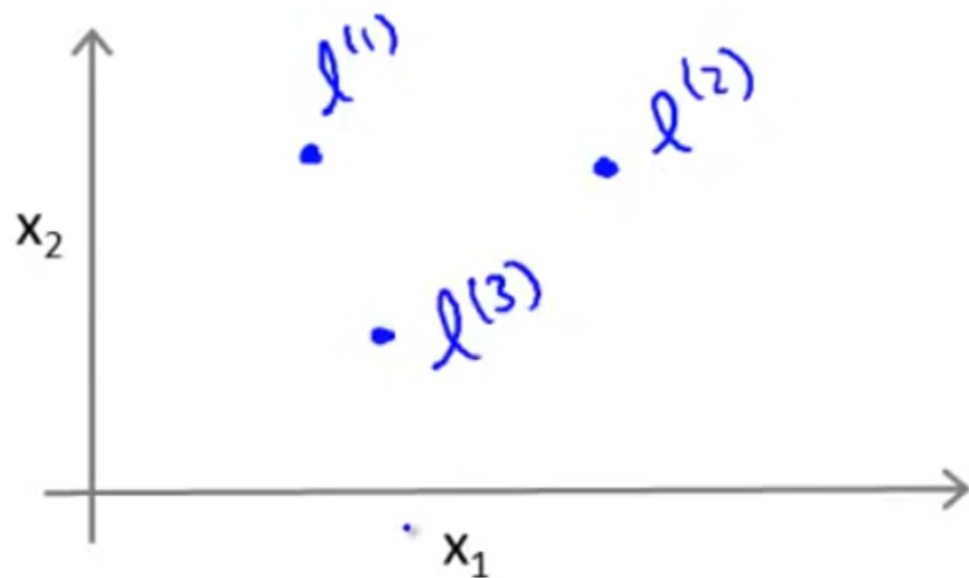
Is there a different / better choice of the features  $f_1, f_2, f_3, \dots$ ?

## Kernel



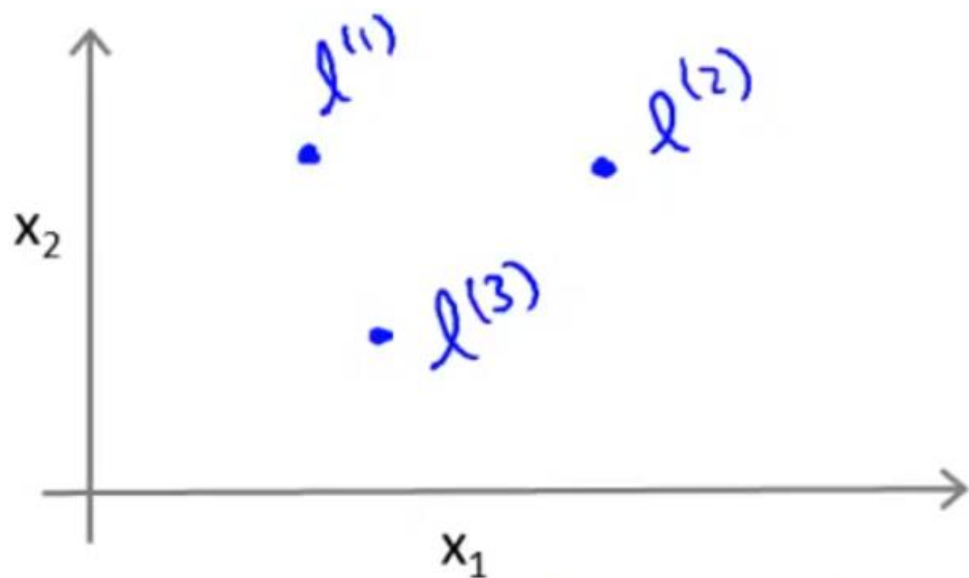
Given  $x$ , compute new feature depending on proximity to landmarks  $l^{(1)}, l^{(2)}, l^{(3)}$

## Kernel



Given  $x$ , compute new feature depending on proximity to landmarks  $l^{(1)}, l^{(2)}, l^{(3)}$

## Kernel

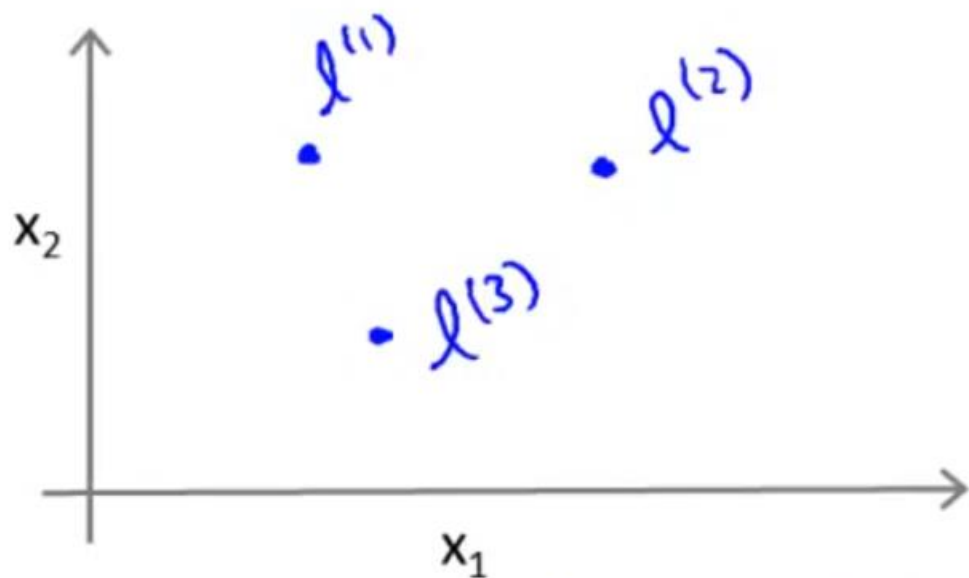


Given  $x$ , compute new feature depending on proximity to landmarks  $l^{(1)}, l^{(2)}, l^{(3)}$

Given  $x$ :  $f_1 = \text{similarity}(x, l^{(1)})$



## Kernel



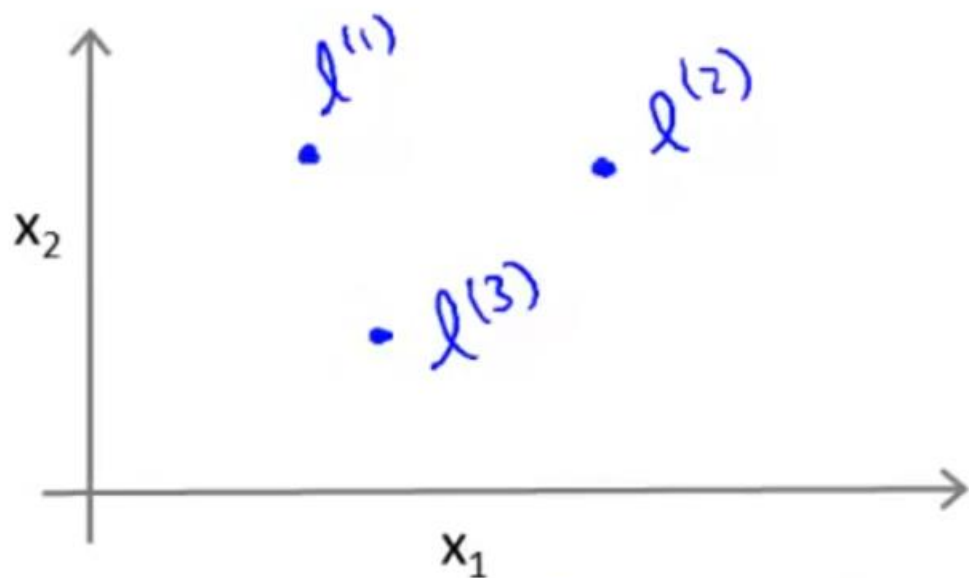
Given  $x$ , compute new feature depending on proximity to landmarks  $l^{(1)}, l^{(2)}, l^{(3)}$

Given  $x$ :

$$f_1 = \text{similarity}(x, l^{(1)}) = \exp\left(-\frac{\|x - l^{(1)}\|^2}{2\sigma^2}\right)$$



## Kernel



Given  $x$ , compute new feature depending on proximity to landmarks  $l^{(1)}, l^{(2)}, l^{(3)}$

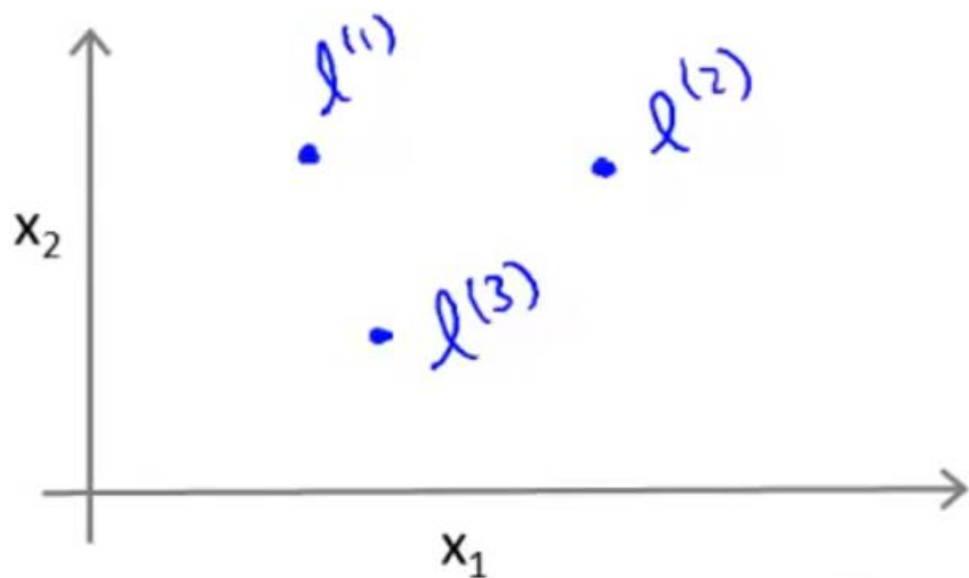
Given  $x$ :

$$f_1 = \text{similarity}(x, l^{(1)}) = \exp\left(-\frac{\|x - l^{(1)}\|^2}{2\sigma^2}\right)$$

Handwritten annotations in red:

- $\|w\|$  with an arrow pointing to the denominator  $2\sigma^2$ .
- $\|x - l^{(1)}\|^2$  with an arrow pointing to the numerator.

## Kernel



Given  $x$ , compute new feature depending on proximity to landmarks  $l^{(1)}, l^{(2)}, l^{(3)}$

Given  $x$ :

$$f_1 = \text{similarity}(x, l^{(1)}) = \exp\left(-\frac{\|x - l^{(1)}\|^2}{2\sigma^2}\right)$$

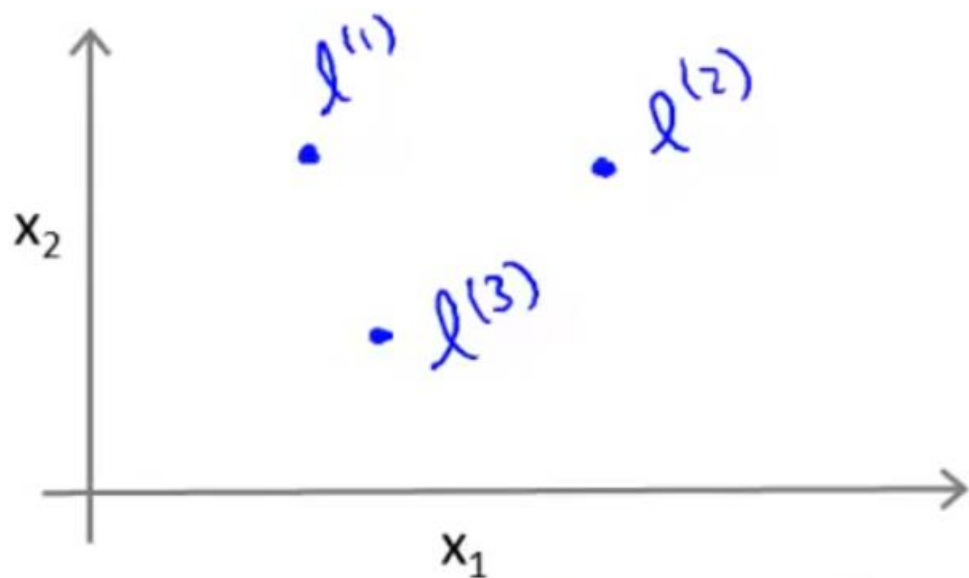
$$f_2 = \text{similarity}(x, l^{(2)}) = \exp\left(-\frac{\|x - l^{(2)}\|^2}{2\sigma^2}\right)$$

$$f_3 = \text{similarity}(x, l^{(3)}) = \exp(\dots)$$

Handwritten red annotations:

- $\|w\|$  with an arrow pointing to the denominator  $2\sigma^2$  in the first equation.
- $\|x - l^{(1)}\|^2$  with an arrow pointing to the numerator of the first equation.
- $\|x - l^{(2)}\|^2$  with an arrow pointing to the numerator of the second equation.

## Kernel



Given  $x$ , compute new feature depending on proximity to landmarks  $l^{(1)}, l^{(2)}, l^{(3)}$

Given  $x$ :

$$f_1 = \text{similarity}(x, l^{(1)}) = \exp\left(-\frac{\|x - l^{(1)}\|^2}{2\sigma^2}\right)$$

$$f_2 = \text{similarity}(x, l^{(2)}) = \exp\left(-\frac{\|x - l^{(2)}\|^2}{2\sigma^2}\right)$$

$$f_3 = \text{similarity}(x, l^{(3)}) = \exp(\dots)$$

kernel (Gaussian kernels)

$$k(x, l^{(i)})$$

Handwritten red annotations for the first equation:

- $\|w\|$  with an arrow pointing to the denominator  $2\sigma^2$ .
- $\|x - l^{(1)}\|^2$  with an arrow pointing to the numerator.
- $\|x - l^{(1)}\|^2$  with an arrow pointing to the square term in the numerator.

## Kernels and Similarity

$$f_1 = \text{similarity}(x, \underline{l^{(1)}}) = \exp\left(-\frac{\|x - l^{(1)}\|^2}{2\sigma^2}\right)$$

•

If  $x \approx l^{(1)}$  :

If  $x$  is far from  $l^{(1)}$  :

## Kernels and Similarity

$$f_1 = \text{similarity}(x, \underline{l^{(1)}}) = \exp \left( - \overset{l}{\rightarrow} \frac{\|x - l^{(1)}\|^2}{2\sigma^2} \right) = \exp \left( - \frac{\sum_{j=1}^n (x_j - \overset{\downarrow}{l_j^{(1)}})^2}{2\sigma^2} \right)$$

If  $x \approx l^{(1)}$  :

If  $x$  is far from  $l^{(1)}$  :

## Kernels and Similarity

$$f_1 = \text{similarity}(x, \underline{l}^{(1)}) = \exp \left( -\frac{\|x - l^{(1)}\|^2}{2\sigma^2} \right) = \exp \left( -\frac{\sum_{j=1}^n (x_j - l_j^{(1)})^2}{2\sigma^2} \right)$$

If  $x \approx l^{(1)}$  :

$$f_1 \approx \exp \left( -\frac{0^2}{2\sigma^2} \right) \approx 1$$

If  $x$  is far from  $l^{(1)}$  :

## Kernels and Similarity

$$f_1 = \text{similarity}(x, \underline{l^{(1)}}) = \exp\left(-\frac{\|x - l^{(1)}\|^2}{2\sigma^2}\right) = \exp\left(-\frac{\sum_{j=1}^n (x_j - l_j^{(1)})^2}{2\sigma^2}\right)$$

If  $x \approx l^{(1)}$  :

$$f_1 \approx \exp\left(-\frac{0^2}{2\sigma^2}\right) \approx 1$$

If  $x$  is far from  $l^{(1)}$  :

$$f_1 = \exp\left(-\frac{(\text{large number})^2}{2\sigma^2}\right) \approx 0.$$



## Kernels and Similarity

$$f_1 = \text{similarity}(x, \underline{l^{(1)}}) = \exp\left(-\frac{\|x - l^{(1)}\|^2}{2\sigma^2}\right) = \exp\left(-\frac{\sum_{j=1}^n (x_j - l_j^{(1)})^2}{2\sigma^2}\right)$$

If  $x \approx l^{(1)}$  :

$$f_1 \approx \exp\left(-\frac{0^2}{2\sigma^2}\right) \approx 1$$

$$\begin{aligned} l^{(1)} &\rightarrow f_1 \\ l^{(2)} &\rightarrow f_2 \\ l^{(3)} &\rightarrow f_3 \dots \end{aligned}$$

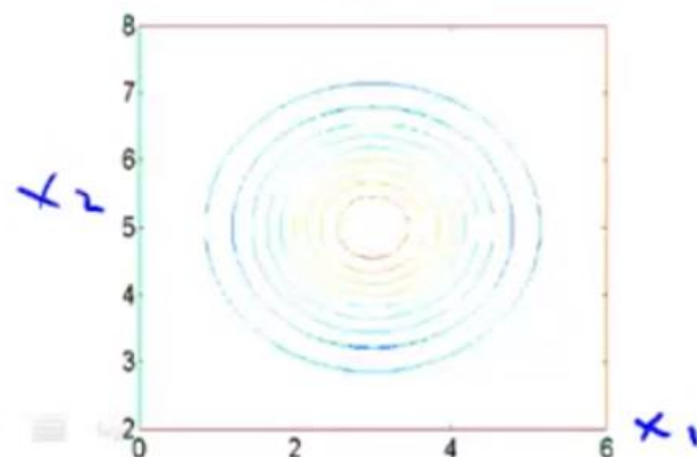
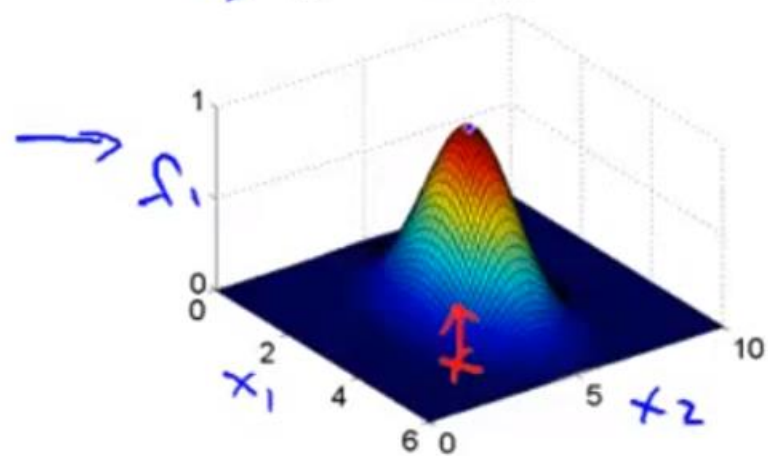
If  $x$  is far from  $l^{(1)}$  :

$$f_1 = \exp\left(-\frac{(\text{large number})^2}{2\sigma^2}\right) \approx 0.$$

## Example:

$$\rightarrow l^{(1)} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \quad \underline{f_1 = \exp\left(-\frac{\|x - l^{(1)}\|^2}{2\sigma^2}\right)}$$

$$\rightarrow \sigma^2 = 1$$

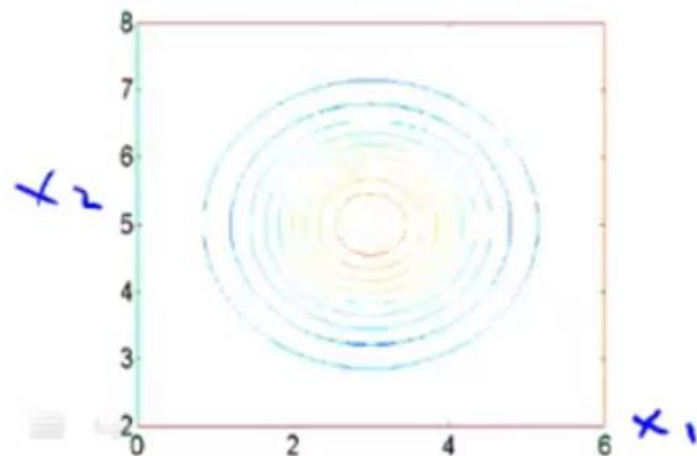
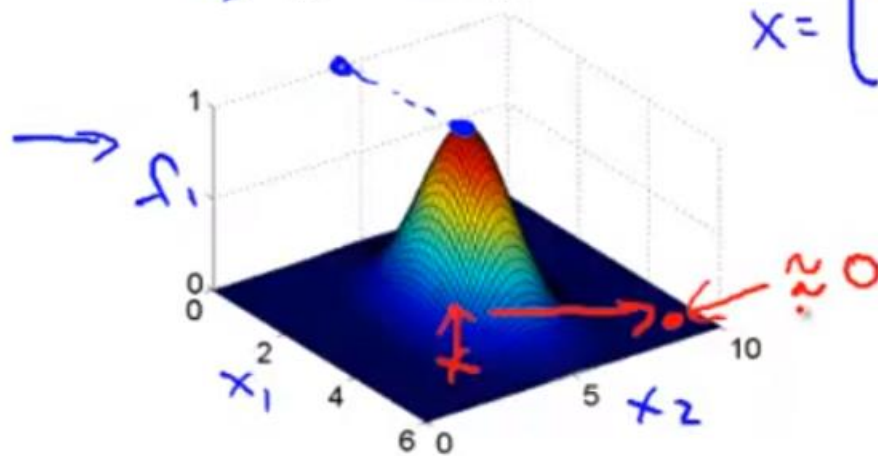


## Example:

$$\rightarrow l^{(1)} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \quad \underline{f_1 = \exp\left(-\frac{\|x - l^{(1)}\|^2}{2\sigma^2}\right)}$$

$$\rightarrow \sigma^2 = 1$$

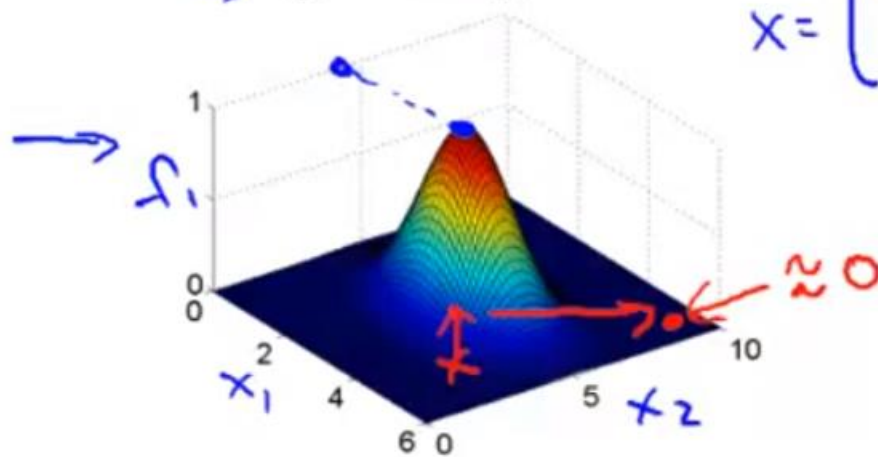
$$x = \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$



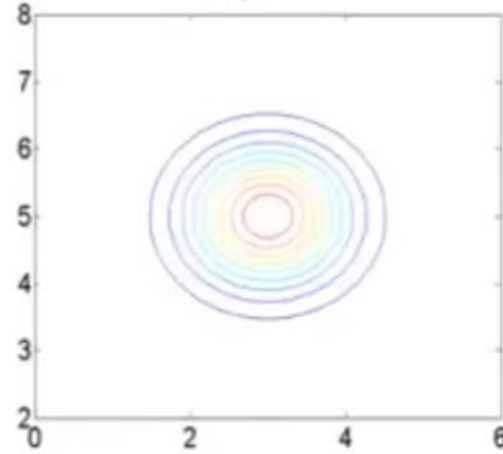
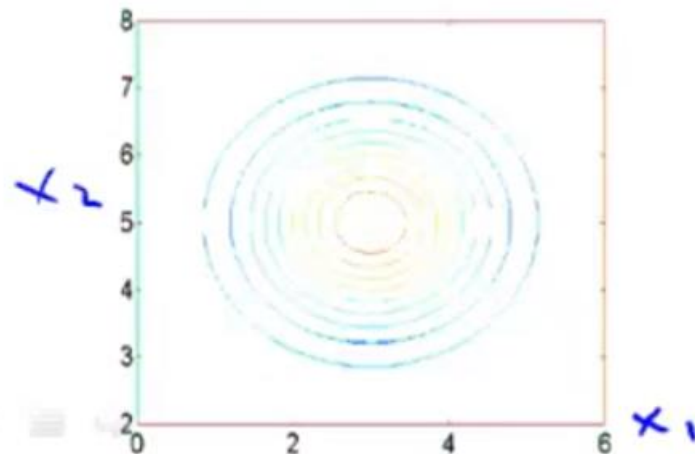
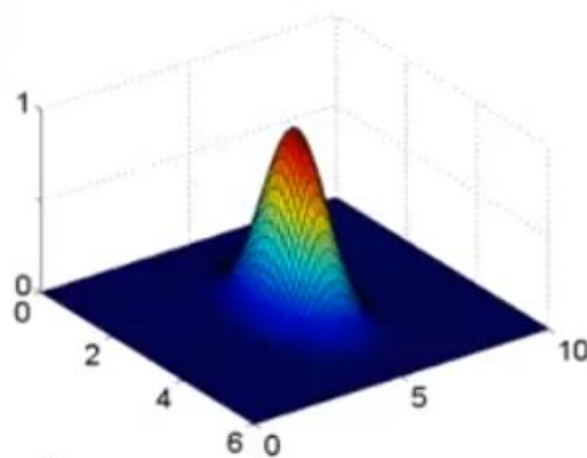
## Example:

$$\rightarrow l^{(1)} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \quad f_1 = \exp\left(-\frac{\|x - l^{(1)}\|^2}{2\sigma^2}\right)$$

$$\rightarrow \sigma^2 = 1$$



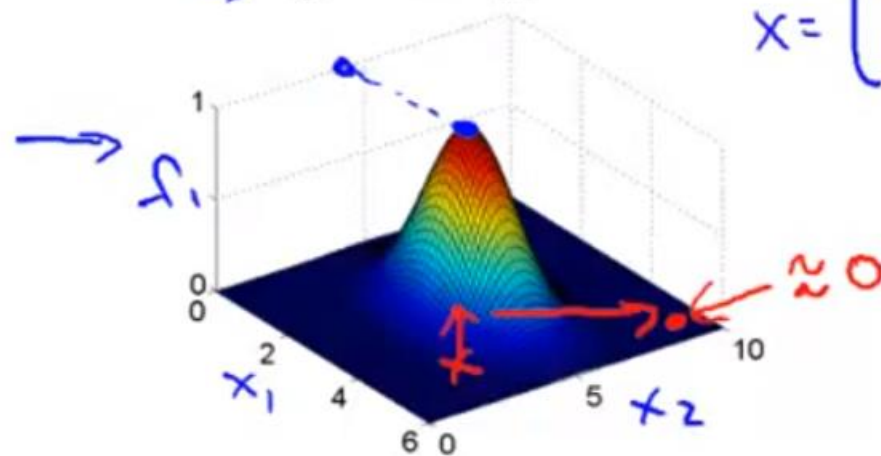
$$\sigma^2 = 0.5$$



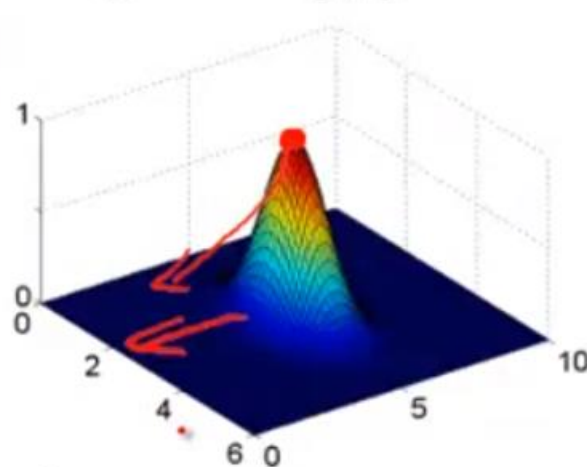
## Example:

$$\rightarrow l^{(1)} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \quad f_1 = \exp\left(-\frac{\|x - l^{(1)}\|^2}{2\sigma^2}\right)$$

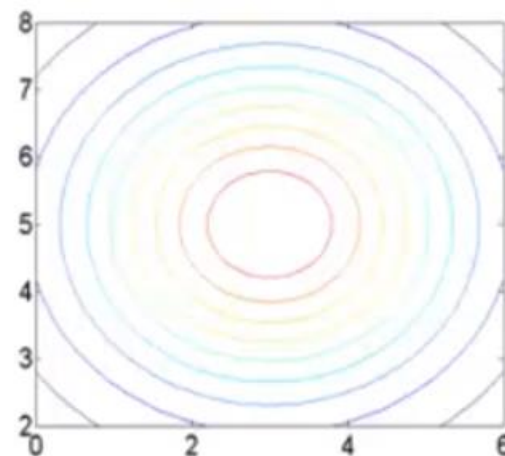
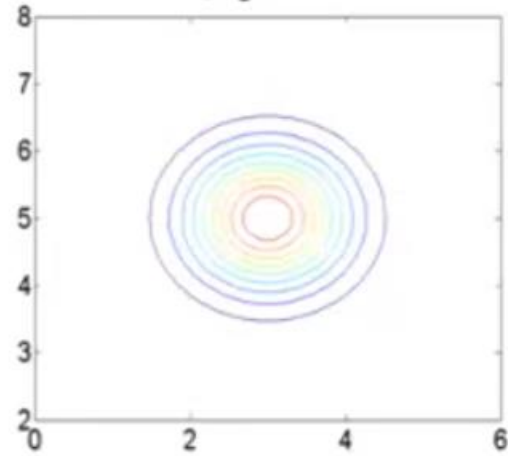
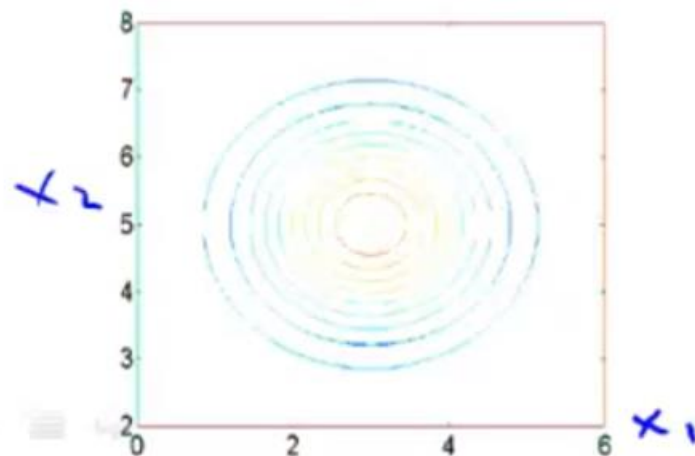
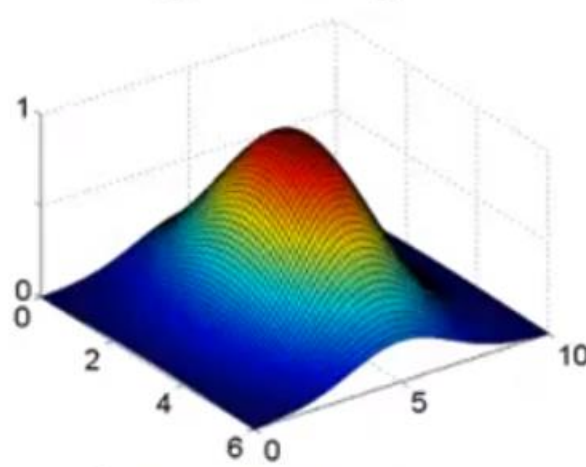
$$\rightarrow \sigma^2 = 1$$

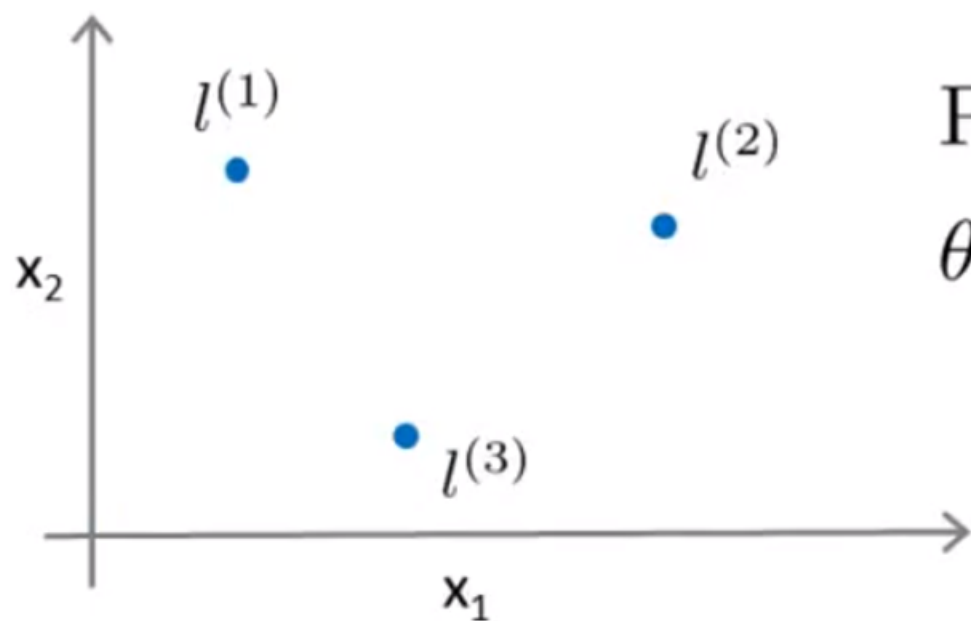


$$\sigma^2 = 0.5$$



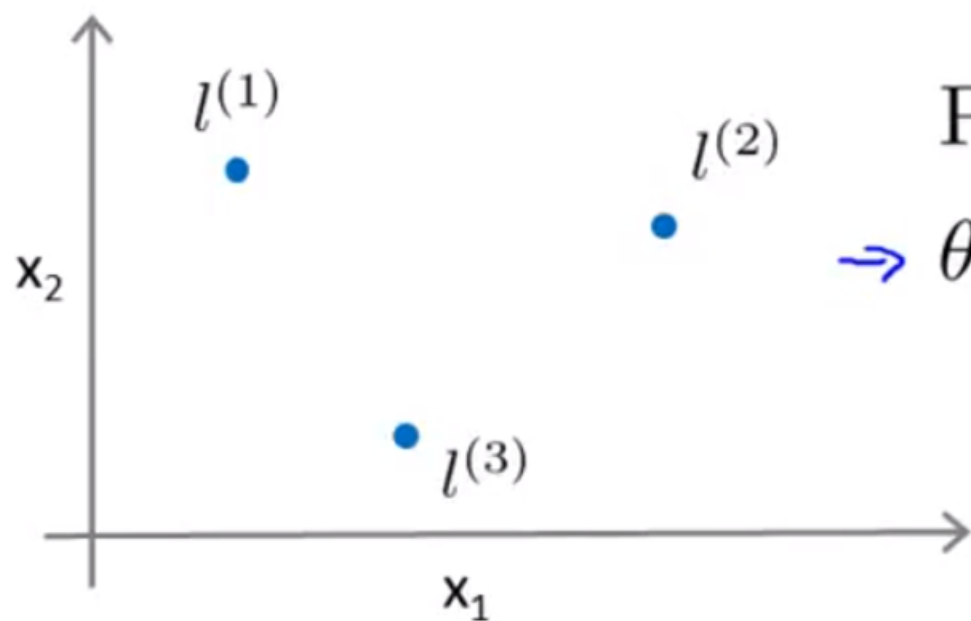
$$\sigma^2 = 3$$





Predict “1” when

$$\theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 \geq 0$$



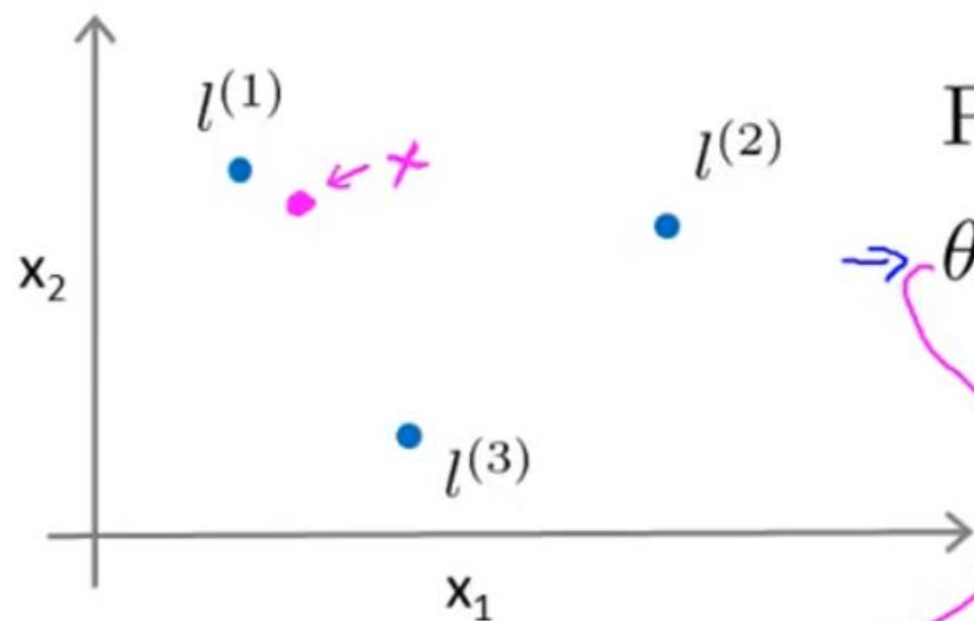
Predict “1” when

$$\rightarrow \theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 \geq 0$$



$$\Theta_0 = -0.5, \quad \Theta_1 = 1, \quad \Theta_2 = 1, \quad \Theta_3 = 0$$





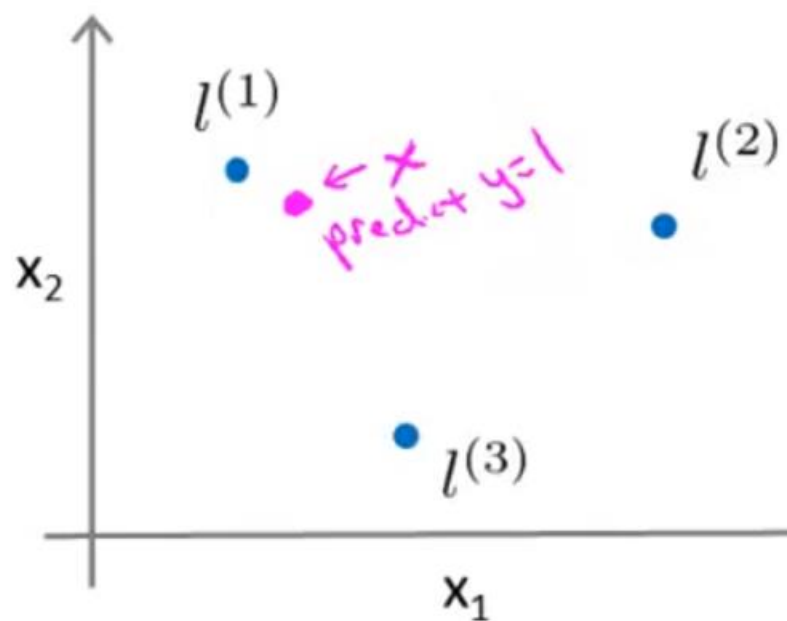
Predict "1" when

$$\rightarrow \theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 \geq 0$$



$$\theta_0 = -0.5, \quad \theta_1 = 1, \quad \theta_2 = 1, \quad \theta_3 = 0$$

$$f_1 \approx 1, \quad f_2 \approx 0, \quad f_3 \approx 0.$$



Predict "1" when

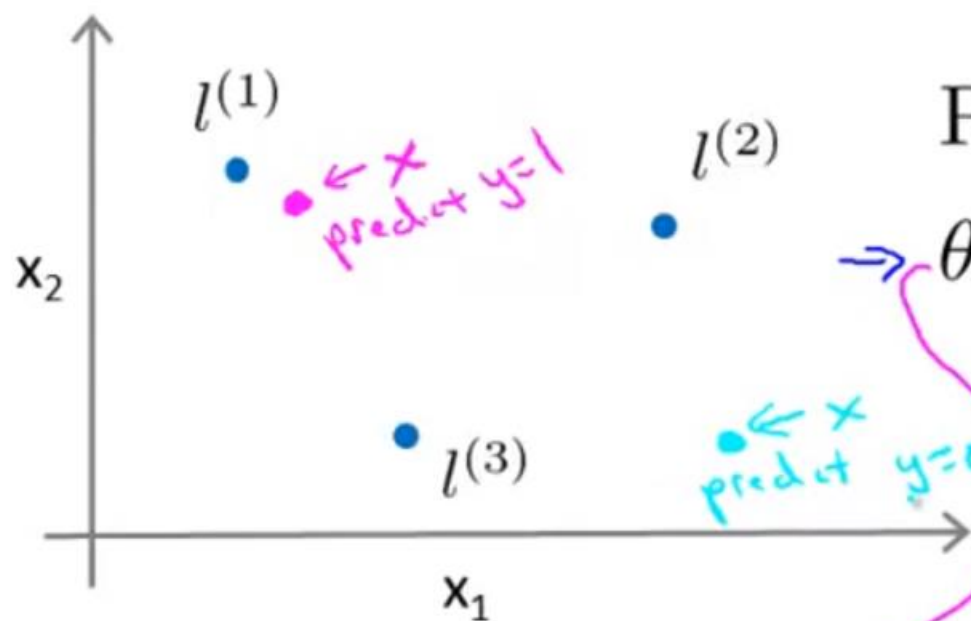
$$\rightarrow \theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 \geq 0$$

↑  
X

$$\underline{\theta_0 = -0.5, \theta_1 = 1, \theta_2 = 1, \theta_3 = 0}$$

$$f_1 \approx 1, f_2 \approx 0, f_3 \approx 0.$$

$$\begin{aligned} \rightarrow & \theta_0 + \theta_1 \times 1 + \theta_2 \times 0 + \theta_3 \times 0 \\ & = -0.5 + 1 = 0.5 \geq 0 \end{aligned}$$



Predict "1" when

$$\rightarrow \theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 \geq 0$$



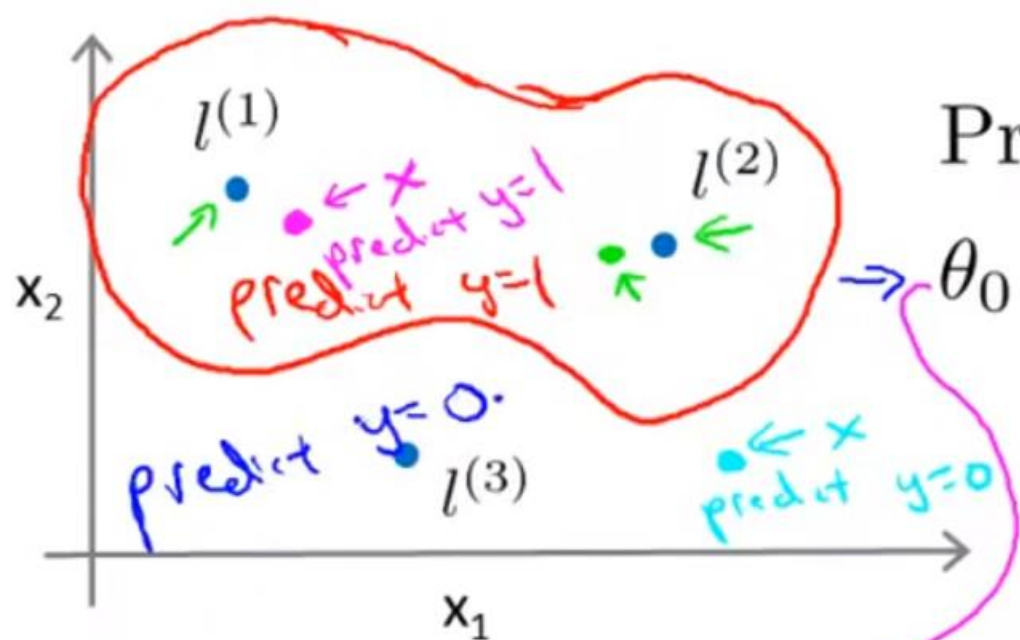
$$\underline{\theta_0 = -0.5, \theta_1 = 1, \theta_2 = 1, \theta_3 = 0}$$

$$f_1 \approx 1, f_2 \approx 0, f_3 \approx 0.$$

$$\begin{aligned} \rightarrow \theta_0 + \theta_1 \times 1 + \theta_2 \times 0 + \theta_3 \times 0 \\ = -0.5 + 1 = 0.5 \geq 0 \end{aligned}$$

$$f_1, f_2, f_3 \approx 0$$

$$\rightarrow \underline{\theta_0} + \theta_1 \underline{f_1} + \dots \approx -0.5 < 0$$



Predict "1" when

$$\theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 \geq 0$$

$\uparrow$   
 $\times$

$$\theta_0 = -0.5, \theta_1 = 1, \theta_2 = 1, \theta_3 = 0$$

$$f_1 \approx 1, f_2 \approx 0, f_3 \approx 0.$$

$$\begin{aligned} \rightarrow \theta_0 + \theta_1 \times 1 + \theta_2 \times 0 + \theta_3 \times 0 \\ = -0.5 + 1 = 0.5 \geq 0 \end{aligned}$$

$$f_1, f_2, f_3 \approx 0$$

$$\rightarrow \theta_0 + \theta_1 f_1 + \dots \approx -0.5 < 0$$