Non-Linear Decision Boundary:

May 29, 2021

$$h_\theta(x) = \begin{cases} 1 & \text{if } \theta_0 + \theta_1 x_1 + \theta_2 x_2 \ldots \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Now let $f_1 = x_1$, $f_2 = x_2$, $f_3 = x_1 x_2$, etc

↳ $\theta_0 + \theta_1 f_1 + \theta_2 f_2 \ldots$

Is there a better choice for $f_1, f_2 \ldots$?

Kernels:

Given x, calculate f depending on proximity to landmarks $l^{(1)}, l^{(2)} \ldots$

i.e.

function

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

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

$$f_3 = \text{similarity}(x, l^{(3)}) = \exp\left(-\frac{\|x - l^{(3)}\|^2}{2\sigma^2}\right)$$

↳ Gaussian kernel, a.k.a. $k(x, l^{(i)})$

Kernels and Similarity:

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

$$= \exp\left(-\frac{\sum_{j=1}^{n}(x_j - l_j^{(1)})^2}{2\sigma^2}\right)$$

↳ cuz $x \in \mathbb{R}^n$

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

i.e. x close to first landmark.

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

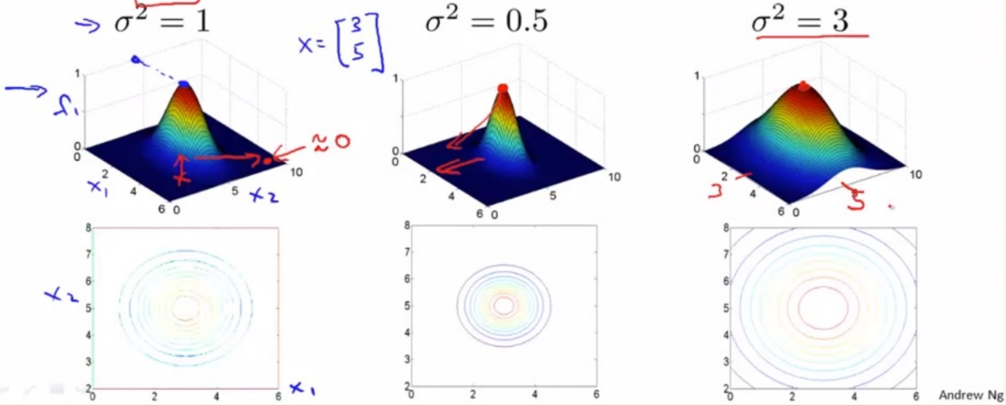
If x far from $l^{(1)}$:

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

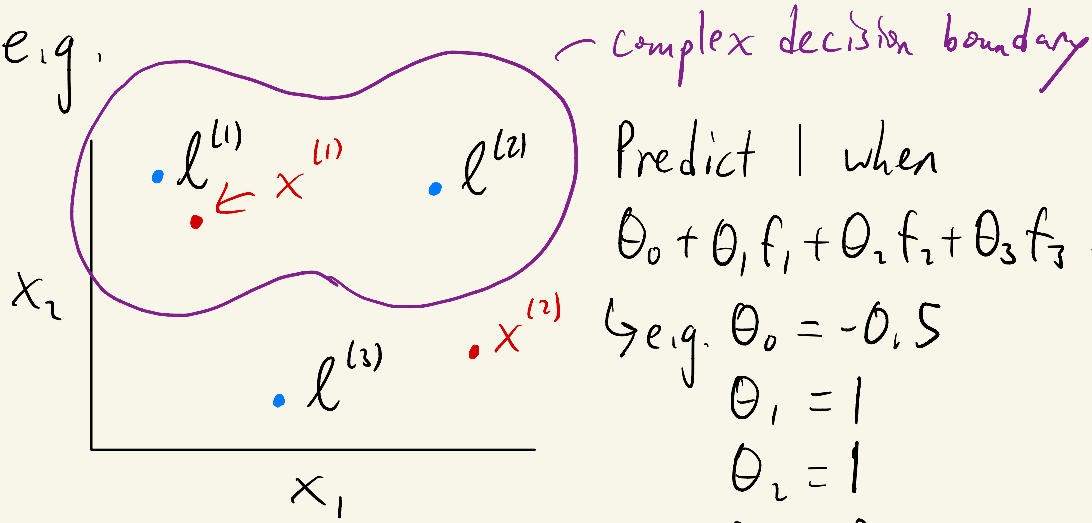
$$\left.\begin{array}{l} l^{(1)} \to f_1 \\ l^{(2)} \to f_2 \\ \vdots \end{array}\right\} x$$

Example:

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



e.g.



Predict 1 when

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

↳ e.g. $\theta_0 = -0.5$

$\theta_1 = 1$

$\theta_2 = 1$

$\theta_3 = 0$

↳ $f_1^1 \approx 1,\ f_2^1 \approx 0,\ f_3^1 \approx 0$

$$\theta_0 + \theta_1 + 0 + 0$$

$$= -0.5 + 1 = 0.5 \geq 0$$

$\therefore$ Predict $y = 1$

↳ $f_1^2 \approx 0,\ f_2^2 \approx 0,\ f_3^2 \approx 0$

$$\theta_0 + 0 + 0 + 0$$

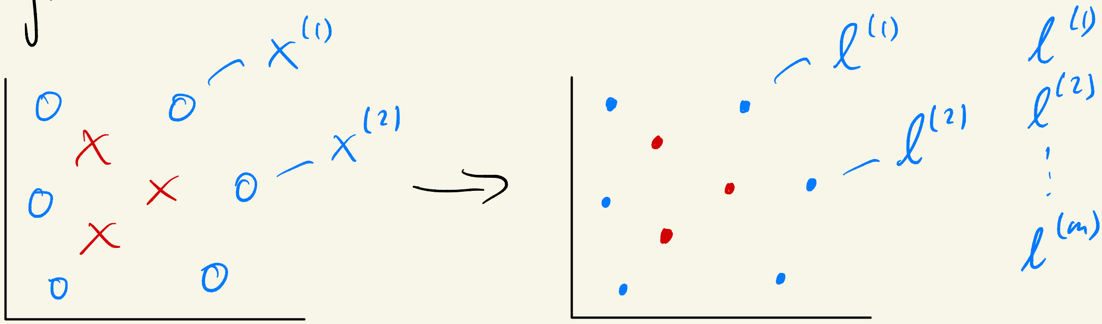
$$= -0.5 < 0$$

$\therefore$ Predict $y = 0$

How Do We Choose Landmarks?

Put a landmark where every training example is.

e.g.



i.e. Given $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}) \ldots (x^{(m)}, y^{(m)})$

Choose $l^{(1)} = x^{(1)}, l^{(2)} = x^{(2)} \ldots l^{(m)} = x^{(m)}$

For some new example $x^{(i)}$:

$$\left.\begin{aligned} f_1^{(i)} &= \text{similarity}(x^{(i)}, l^{(1)}) \quad x^{(1)} \\ f_2^{(i)} &= \text{similarity}(x^{(i)}, l^{(2)}) \\ &\vdots \\ f_m^{(i)} &= \text{similarity}(x^{(i)}, l^{(m)}) \end{aligned}\right\} \therefore f^{(i)} = \begin{bmatrix} f_0^{(i)} = 1 \\ \vdots \\ f_m^{(i)} \end{bmatrix}$$

$\therefore$ Given x, compute features $f \in \mathbb{R}^{m+1}$.

Predict $y=1$ if $\theta^T f \geq 0$, $\theta \in \mathbb{R}^{m+1}$

$\hookrightarrow = \theta_0 f_0 + \theta_1 f_1 \ldots \theta_m f_m$

Training:

$$\min_\theta C \sum_{i=1}^{m} y^{(i)} \text{cost}_1(\underbrace{\theta^T f^{(i)}}_{\theta^T x^{(i)}}) + (1-y^{(i)}) \text{cost}_0(\theta^T f^{(i)})$$

$$+ \frac{1}{2} \sum_{j=1}^{n=m} \theta_j^2$$

SVM Parameters:

$C = \frac{1}{\lambda}$: Large C = lower bias, higher variance

Small C = higher bias, lower variance

σ^2: Large σ^2 = features $f^{(i)}$ vary more smoothly

$\hookrightarrow$ higher bias, lower variance

Small σ^2 = features $f^{(i)}$ vary less smoothly

$\hookrightarrow$ lower bias, higher variance