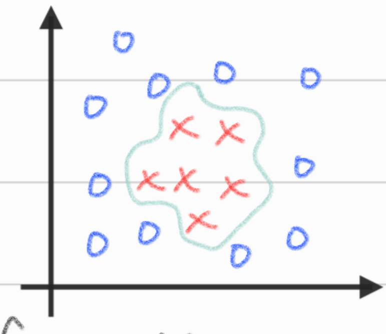
Kernels

Non-linear Decision Boundary



Given x:

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

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

If x is far from $l^{(1)}$: $f_1 = \exp\left(-\frac{(\text{큰수})^2}{2\sigma^2}\right) \approx 0$

σ^2 (sigma square) 크면 , 작으면

Define extra features using landmarks and similarity functions to learn more complex nonlinear classifier

SVM with Kernels

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

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

Given example x:

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

$$f_2 = \text{similarity}(x, l^{(2)})$$

$$f = \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} \quad f_0 = 1$$

For training example $(x^{(i)}, y^{(i)})$:

$$f_1^{(i)} = \text{sim}(x^{(i)}, l^{(1)})$$

$$x^{(i)} \rightarrow f_2^{(i)} = \text{sim}(x^{(i)}, l^{(2)})$$

$$\vdots \quad \leftarrow f_i^{(i)} = \text{sim}(x^{(i)}, l^{(i)}) = \exp\left(-\frac{0}{2\sigma^2}\right) = 1$$

$$f_m^{(i)} = \text{sim}(x^{(i)}, l^{(i)})$$

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

Predict "$y=1$" if $\theta^T f \geq 0$

Training:

$$\min_\theta C \sum_{i=1}^{m} y^{(i)} \text{cost}_1(\theta^T f^{(i)}) + (1-y^{(i)}) \text{cost}_0(\theta^T f^{(i)}) + \frac{1}{2} \sum_{j=1}^{m} \theta_j^2$$

$C \left(= \frac{1}{\lambda}\right)$ $\begin{cases} \text{Large } C: \text{Lower Bias, Higher Variance} \; (= \text{small } \lambda) \text{ 규제}\downarrow \\ \text{small } C: \text{Higher Bias, Lower Variance} \; (= \text{large } \lambda) \text{ 규제}\uparrow \end{cases}$

σ^2 $\begin{cases} \text{Large } \sigma^2: \text{Higher Bias, Lower Variance} \; \text{features vary smoothly} \text{ 차}\uparrow \\ \text{small } \sigma^2: \text{Lower Bias, Higher Variance} \; \text{" less smoothly} \text{ 차}\downarrow \end{cases}$