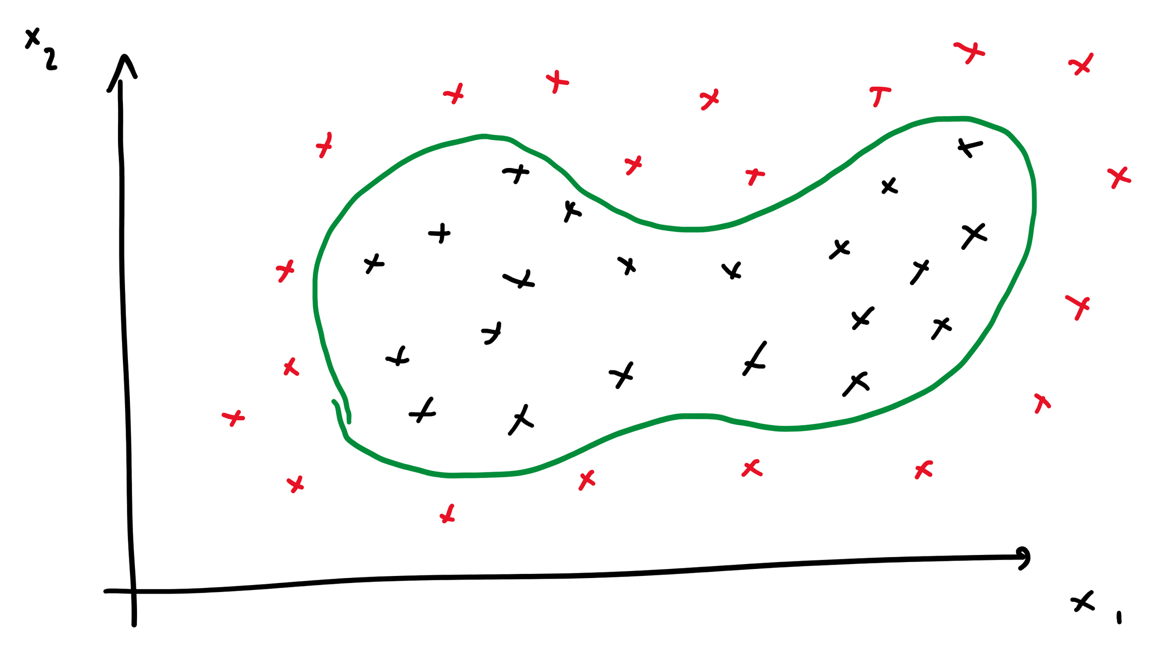
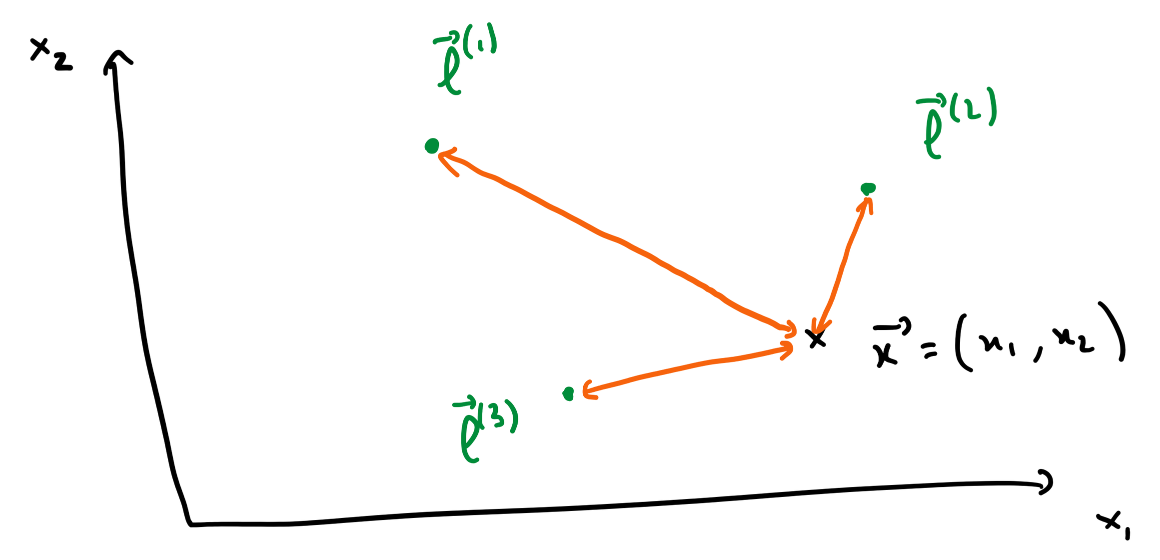
Video 14

Saturday, October 10, 2020 9:01 PM



Idea is similar to feature maps

Introduce the concept of a landmark point:



Given $\vec{x}$:

$$f_1 = \underbrace{\text{sim}(\vec{x}, \vec{\ell}^{(1)})}_{\text{similarity metric between } \vec{x} \text{ and landmark } \vec{\ell}^{(1)}} = \underbrace{\exp\left(-\frac{\|\vec{x} - \vec{\ell}^{(1)}\|^2}{2\underbrace{\sigma^2}_{\text{scale}}}\right)}_{\text{we've chosen a Gaussian kernel to serve as the similarity metric}}$$

if $\vec{x}$ is close to $\vec{\ell}^{(1)}$: $\|\vec{x} - \vec{\ell}^{(1)}\|^2 \to 0$

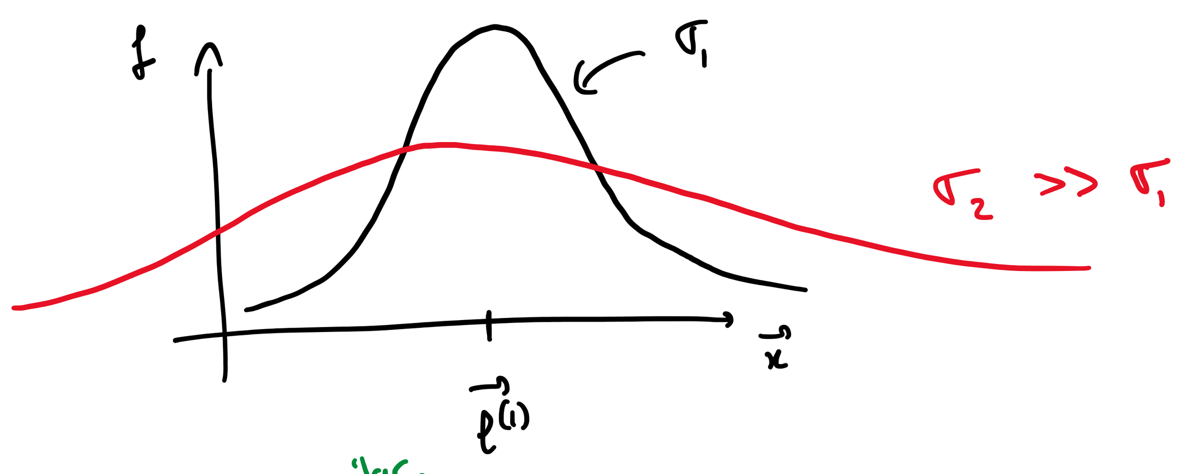
$\exp(\dots) \to 1$

$f_1 \to 1$

if $\vec{x}$ is far from $\vec{\ell}^{(1)}$: $\|\vec{x} - \vec{\ell}^{(1)}\|^2 \to \infty$

$\exp(\dots) \to 0$

$f_1 \to 0$



Given $\vec{x}$:

$$\vec{f}(\vec{x}) = \begin{pmatrix} f_1 \\ f_2 \\ f_3 \end{pmatrix} = \begin{pmatrix} \exp\left(-\frac{\|\vec{x} - \vec{\ell}^{(1)}\|^2}{2\sigma^2}\right) \\ \exp\left(-\frac{\|\vec{x} - \vec{\ell}^{(2)}\|^2}{2\sigma^2}\right) \\ \exp\left(-\frac{\|\vec{x} - \vec{\ell}^{(3)}\|^2}{2\sigma^2}\right) \end{pmatrix}$$

how similar is $\vec{x}$ to $\vec{\ell}^{(1)}$

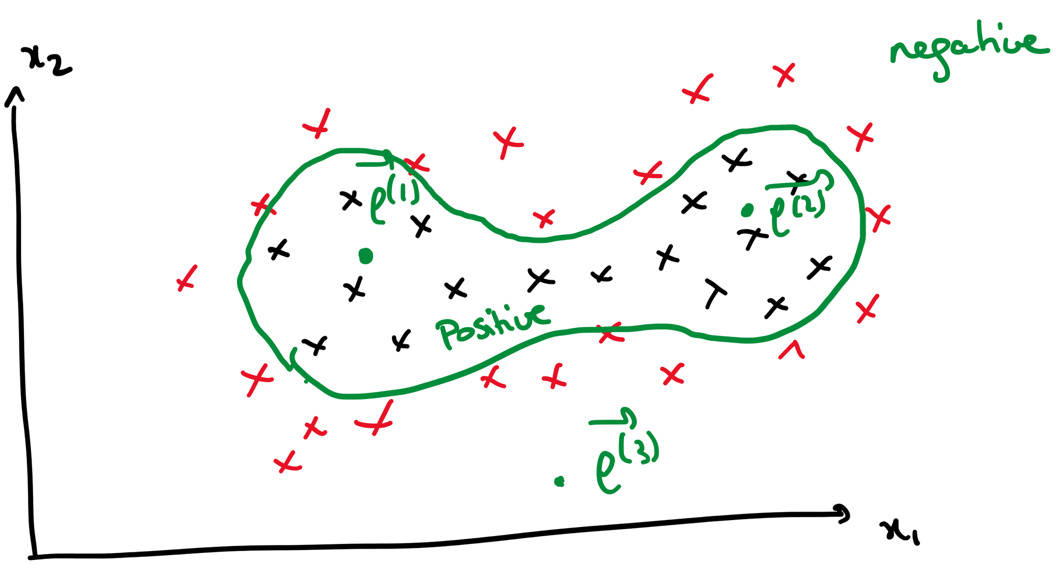
how similar is $\vec{x}$ to $\vec{\ell}^{(2)}$

how similar is $\vec{x}$ to $\vec{\ell}^{(3)}$

$y = \vec{w}^T \vec{x} = w_1 x_1 + w_2 x_2$

With kernel: $y = w_1 f_1 + w_2 f_2 + w_3 f_3$

here the new features computed with the kernel become the inputs to our linear classifier / SVM.



large weight w_1 w_2

small weight w_3

⇓

asking for the classifier to predict 1 when $\vec{x}$ is close to $\vec{\ell}^{(1)}$ and/or $\vec{\ell}^{(2)}$

or 0 when $\vec{x}$ is far from these two landmarks

How to choose the landmarks?

↳ use the data points themselves as landmarks.

$\vec{x} = (x_1, \dots, x_m)$

each of these components is a feature

m is the number of dimensions

$\vec{f}(\vec{x}) = (f_1, \dots, f_N)$

N is the number of training points

for $\vec{x}^{(1)}$: $\vec{f}(\vec{x}^{(1)}) = \begin{pmatrix} \text{sim}(\vec{x}^{(1)}, \vec{x}^{(1)}) \\ \text{sim}(\vec{x}^{(1)}, \vec{x}^{(2)}) \\ \vdots \\ \text{sim}(\vec{x}^{(1)}, \vec{x}^{(N)}) \end{pmatrix}$

for $\vec{x}^{(2)}$: $\vec{f}(\vec{x}^{(2)}) = \begin{pmatrix} \text{sim}(\vec{x}^{(2)}, \vec{x}^{(1)}) \\ \text{sim}(\vec{x}^{(2)}, \vec{x}^{(2)}) \\ \vdots \\ \text{sim}(\vec{x}^{(2)}, \vec{x}^{(N)}) \end{pmatrix}$

$\vdots$

for $\vec{x}^{(N)}$: $\vec{f}(\vec{x}^{(N)}) = \begin{pmatrix} \vdots \end{pmatrix}$

We need to compute $N \times N$ similarity matrix.

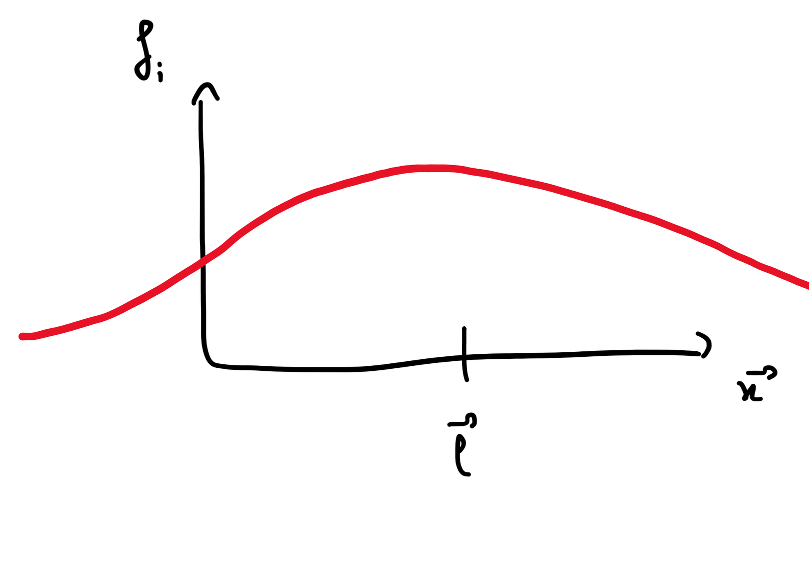
Choice of scale σ:

hyperparameter

Large σ:

underfit

(high bias, lower variance)



Low σ:

overfit

(low bias, high variance)