Recall

For data $y_i \in \{-1,+1\}$ & p-variate features $\{\underline{x}_i\}_{i=1}^n$

the SVM decision function is:

$$f(\underline{x}) = \beta_0 + \sum_{i=1}^n \alpha_i y_i k(\underline{x}_i, \underline{x})$$

$k(\underline{x}_i,\underline{x}) = \underline{x}_i^T \underline{x}$ usual "linear"

$k(\underline{x}_i,\underline{x}) = (1+\underline{x}_i^T\underline{x})^d$ "polynomial"

$k(\underline{x}_i,\underline{x}) = \exp(-a\|\underline{x}_i-\underline{x}\|^2)$ "radial" "squared exp" "Gaussian"

Decision rule:

$$\hat{y} = \begin{cases} +1 & f(\underline{x}) > 0 \\ -1 & f(\underline{x}) < 0 \end{cases}$$

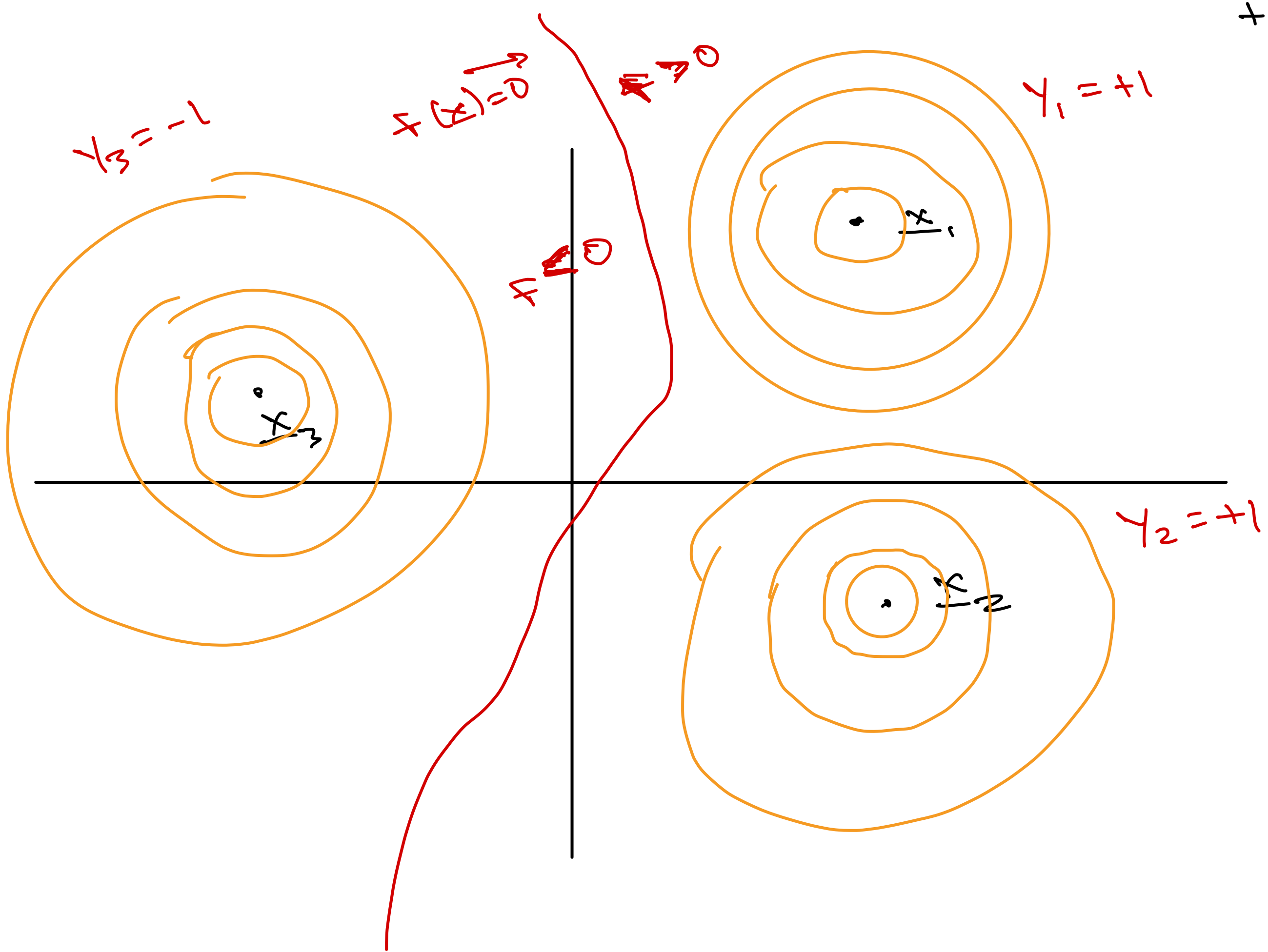
$f(\underline{x}) = 0$ = boundary

Note size of f can be interpreted as confidence in prediction

[$f>0$ but small $\Rightarrow$ close to boundary, $f \gg 0$ really far from boundary]

$\boxed{\text{Ex}}$ $p = 2$ $\quad \underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$$f(\underline{x}) = \beta_0 + \underbrace{\alpha_1 y_1}\, \underbrace{e^{-a\|\underline{x}_1 - \underline{x}\|^2}} + \underbrace{\alpha_2 y_2}\, \underbrace{e^{-a\|\underline{x}_2 - \underline{x}\|^2}} + \underbrace{\alpha_3 y_3}\, \underbrace{e^{-a\|\underline{x}_3 - \underline{x}\|^2}}$$



Remark How come we can just go

$$f(\underline{x}) = \beta_0 + \sum_{i=1}^{n} \alpha_i y_i \underbrace{\underline{x}_i^T \underline{x}} \longrightarrow \beta_0 + \sum_{i=1}^{n} \alpha_i y_i \underbrace{k(\underline{x}_i, \underline{x})}$$

and what is the significance of the mysterious "positive definite function" assumption?

Key **Mercer's Theorem**

If k is a "positive definite function" then

$$k(\underline{u}, \underline{v}) = \sum_{j=1}^{\infty} \lambda_j \phi_j(\underline{u}) \phi_j(\underline{v})$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ and $\{\phi_j(\underline{u})\}$ are some functions

$\boxed{\text{Link}}$

$$\underline{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \qquad \underline{x}_i^T \underline{x} \qquad \text{plane} \quad \begin{pmatrix} x_1 \\ x_1^2 \\ x_1^3 \\ \vdots \\ x_p^3 \end{pmatrix}$$

$$\underline{x}_{i,new} = \begin{pmatrix} \sqrt{\lambda_1}\,\phi_1(\underline{x}) \\ \sqrt{\lambda_2}\,\phi_2(\underline{x}) \\ \vdots \\ \sqrt{\lambda_N}\,\phi_N(\underline{x}) \end{pmatrix} \qquad \text{for } N >> 0$$

$$f(\underline{x}) = \beta_0 + \sum_{i=1}^{n} \alpha_i y_i \, \underline{x}_{i,new}^T \, \underline{x}_{new}$$

$$= \beta_0 + \sum_{i=1}^{n} \alpha_i y_i \sum_{j=1}^{N} \lambda_j \phi_j(\underline{x}_i) \phi_j(\underline{x})$$

$$\approx \beta_0 + \sum_{i=1}^{n} \alpha_i y_i \, K(\underline{x}_i, \underline{x})$$

$\approx$ b/c $N \neq \infty$