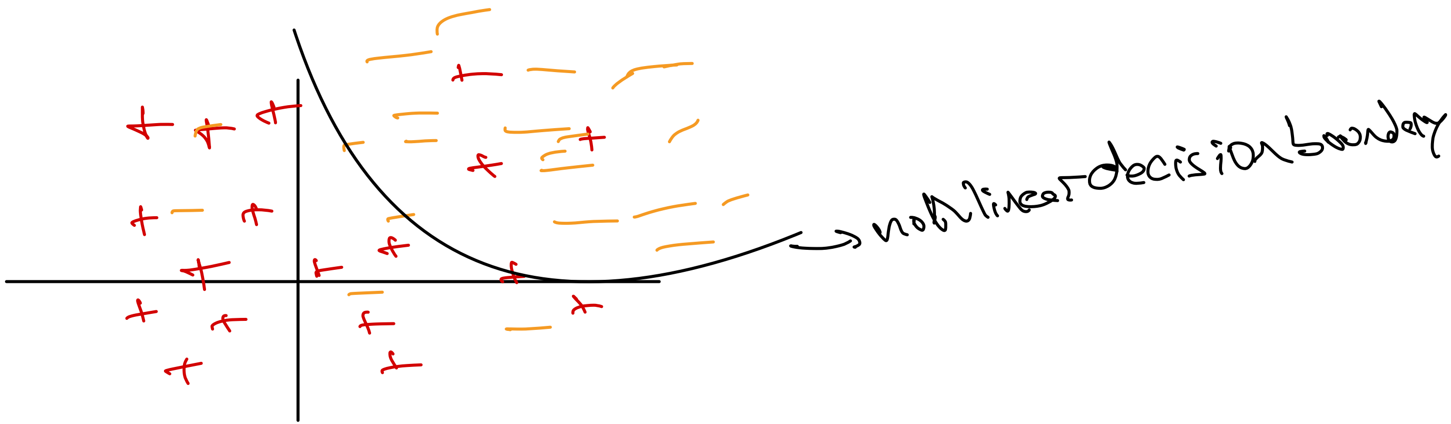
7.3 Support Vector Machines (SVMs)

So far our maximum margin/SV classifier results in a linear decision boundary. A <u>SVM</u> allows for nonlinear boundaries!



Transformations

In linear regression we use transformations to capture nonlinear relationships btwn a response/feature.

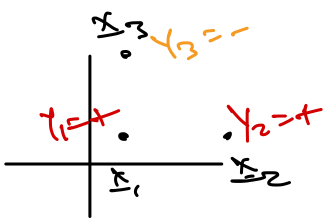
Instead of just

$$x_1, x_2, \dots, x_p$$

could use

$$x_1, x_2, \dots, x_p, x_1^2, x_2^2, \dots, x_p^2, \log(x_1), x_1 x_2, \dots$$

Ex Suppose have data $(\underline{x}_i, y_i)$ $i = 1, 2, 3$ $p = 2$



$$\underline{x}_1 = \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} \quad \underline{x}_2 = \begin{pmatrix} x_{21} \\ x_{22} \end{pmatrix} \quad \underline{x}_3 = \begin{pmatrix} x_{31} \\ x_{32} \end{pmatrix}$$

The classification function at a new $\underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ is:

$$f(\underline{x}) = \beta_0 + \alpha_1 y_1 \underline{x}_1^T \underline{x} + \alpha_2 y_2 \underline{x}_2^T \underline{x} + \alpha_3 y_3 \underline{x}_3^T \underline{x}$$

$$= \beta_0 + \alpha_1 y_1 (x_{11} x_1 + x_{12} x_2) + \alpha_2 y_2 (x_{21} x_1 + x_{22} x_2)$$
$$+ \alpha_3 y_3 (x_{31} x_1 + x_{32} x_2)$$

$$= \beta_0 + [\alpha_1 y_1 x_{11} + \alpha_2 y_2 x_{21} + \alpha_3 y_3 x_{31}] x_1$$
$$+ [\alpha_1 y_1 x_{12} + \alpha_2 y_2 x_{22} + \alpha_3 y_3 x_{32}] x_2$$

$$= \beta_0 + [\text{some coef}] x_1 + [\text{some coef}] x_2$$

= in form of $= \beta_0 + \beta_1 x_1 + \beta_2 x_2$

= hyperplane [decision boundary is $f(\underline{x}) = 0$]

Let's include transformations in $\underline{x}$:

$$\underline{x}_1 = \begin{pmatrix} x_{11} \\ x_{12} \\ x_{11}^2 \\ x_{12}^2 \\ x_{11}x_{12} \end{pmatrix} \qquad \underline{x}_2 = \begin{pmatrix} x_{21} \\ x_{22} \\ x_{21}^2 \\ x_{22}^2 \\ x_{21}x_{22} \end{pmatrix} \qquad \underline{x}_3 = \begin{pmatrix} x_{31} \\ x_{32} \\ x_{31}^2 \\ x_{32}^2 \\ x_{31}x_{32} \end{pmatrix}$$

$$f(\underline{x}) = \beta_0 + \alpha_1 y_1 \underline{x}_1^T \underline{x} + \alpha_2 y_2 \underline{x}_2^T \underline{x} + \alpha_3 y_3 \underline{x}_3^T \underline{x}$$

$\underline{x}$ is now 5-dimensional

$$= \beta_0 + (\text{some coef})x_1 + (\text{some coef})x_2 + (\text{some coef})x_1^2 + (\text{some coef})x_2^2 + (\text{some coef})x_1 x_2$$

= eqn of form = $a + bx + cy + dx^2 + ey^2 + fxy$

= quadratic function.

decision boundary is still $f(\underline{x})=0$
which is no longer linear in x_1, x_2

Key insight: In SVC the $\underline{x}_1, \dots, \underline{x}_n$ and new feature $\underline{x}$ only interact through products like $\underline{x}_i^T \underline{x}$

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

A SVM will replace $\underline{x}_i^T \underline{x}$ with something else.

$\boxed{\text{DEF}}$ A function $k: \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$ is a positive definite function or positive definite kernel if, for any $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_n \in \mathbb{R}^p$ and any $a_1, a_2, \dots, a_n \in \mathbb{R}$

$$\underline{a}^T \Sigma \underline{a} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j k(\underline{x}_i, \underline{x}_j) \geq 0$$

$$\left(k(\underline{x}_i, \underline{x}_j)\right)_{i,j=1}^{n}$$

$\to \mathrm{Var}(\underline{a}^T \underline{z})$ where $\mathrm{Var}\, \underline{z} = \Sigma$

["positive definite" go together as term, b/c k can take on negative values

Positive function $\neq$ pos. def. fcn.]

A SVM uses

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

using $k(\underline{x}_i, \underline{x}) = \underline{x}_i^T \underline{x}$ is linear case

to define decision boundary $f(\underline{x}) = 0$.

Options for k:

- **Polynomial**: $k(\underline{x}_i, \underline{x}) = (1 + \underline{x}_i^T \underline{x})^d$

  for $d = 1, 2, \ldots$

- **Radial**: $k(\underline{x}_i, \underline{x}) = e^{-a\|\underline{x}_i - \underline{x}\|_2^2}$ = "Gaussian kernel"

$\boxed{Ex}$ polynomial $d=1$

The decision boundary is

$$0 = f(\underline{x}) = \beta_0 + \sum_{i=1}^{n} \alpha_i y_i k(\underline{x}_i, \underline{x})$$

$$= \beta_0 + \sum_{i=1}^{n} \alpha_i y_i [1 + \underline{x}_i^T \underline{x}] \leftarrow d=1$$

$$= \left[\beta_0 + \sum_{i=1}^{n} \alpha_i y_i\right] + \sum_{i=1}^{n} \alpha_i y_i \underline{x}_i^T \underline{x}$$

$=$ linear in $\underline{x}_j$, same as our previous SVC.

$[d=2]$ $f(\underline{x}) = \beta_0 + \sum_{i=1}^{n} \alpha_i y_i \, k(\underline{x}_i, \underline{x})$

$= \beta_0 + \sum_{i=1}^{n} \alpha_i y_i \underbrace{(1 + \underline{x}_i^T \underline{x})^2}$

suppose $\underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ $p=2$ features

$(1 + \underline{x}_i^T \underline{x})^2 = (1 + x_{i1} x_1 + x_{i2} x_2)^2$

$= \underbrace{1} + 2 x_{i1} \underbrace{x_1} + 2 x_{i2} \underbrace{x_2} + x_{i1}^2 \underbrace{x_1^2} +$

$x_{i2}^2 \underbrace{x_2^2} + 2 x_{i1} x_{i2} \underbrace{x_1 x_2}$

$=$ quadratic function in (x_1, x_2)

$\boxed{\text{Aside}}$ Two ways to get quadratic boundary:

append original vector with quadratic terms & use SVC, or SVM with poly kernel $d=2$.

$\boxed{\text{Ex}}$ Radial kernel

Suppose we have $\underline{x}_1, \underline{x}_2, \underline{x}_3$ w/ radial kernel

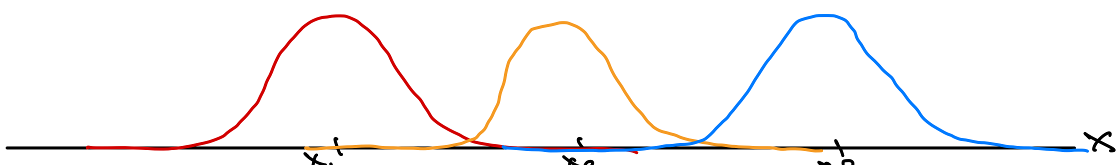
$$f(\underline{x}) = \beta_0 + \alpha_1 y_1 K(\underline{x}_1, \underline{x}) + \alpha_2 y_2 K(\underline{x}_2, \underline{x}) + \alpha_3 y_3 K(\underline{x}_3, \underline{x})$$

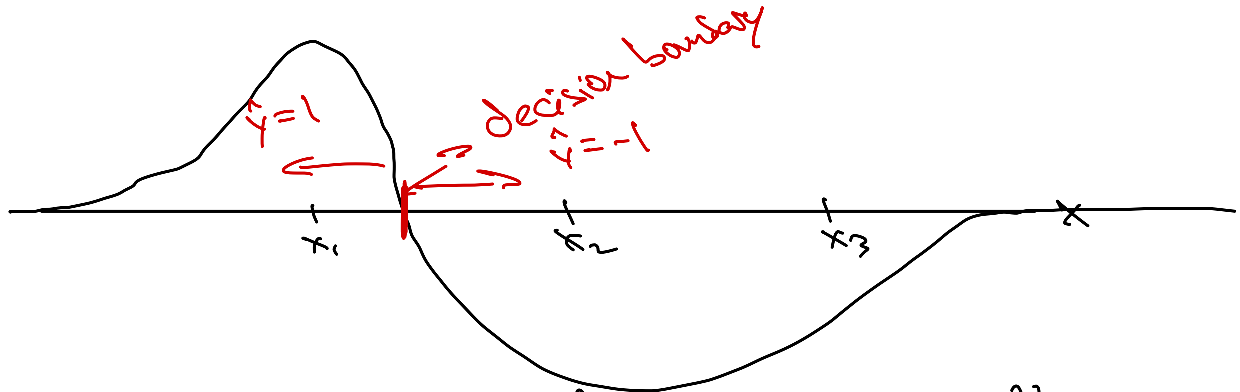
$$= \beta_0 + \alpha_1 y_1 e^{-a\|\underline{x}_1 - \underline{x}\|^2} + \alpha_2 y_2 e^{-a\|\underline{x}_2 - \underline{x}\|^2}$$

$$+ \alpha_3 y_3 e^{-a\|\underline{x}_3 - \underline{x}\|^2}$$

$\boxed{p=1}$

$e^{-a(x_i - x)^2}$





$$F(x) = \beta_0 + \underbrace{\alpha_1 y_1}_{\text{big positive coef}} e^{-a(x_1-x)^2} + \underbrace{\alpha_2 y_2}_{\text{big neg. coef}} e^{-a(x_2-x)^2} + \underbrace{\alpha_3 y_3}_{\text{also negative but smaller}} e^{-a(x_3-x)^2}$$