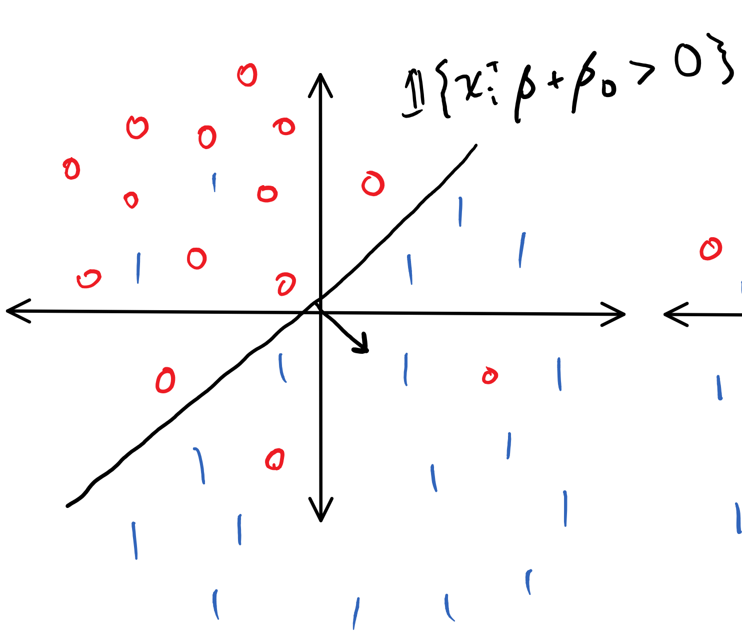
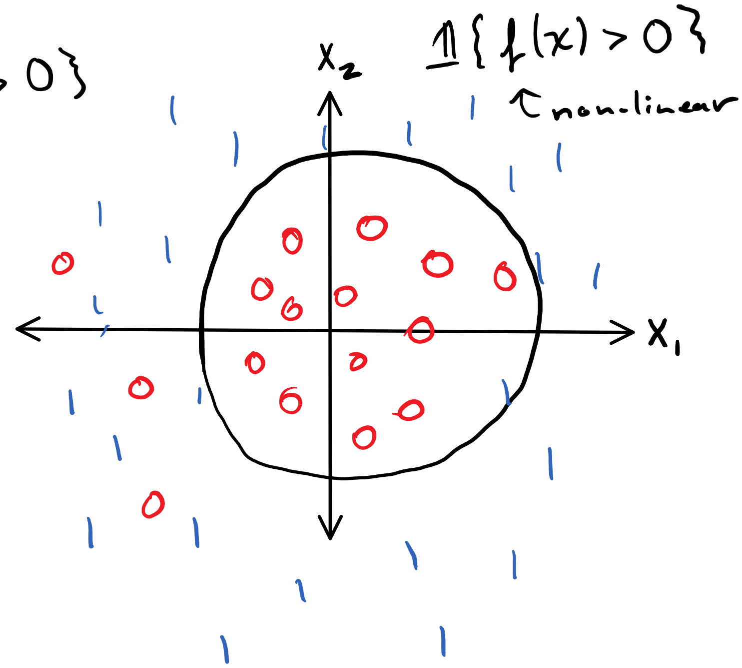
HiDi Embedding

Wednesday, May 8, 2019 1:20 PM



Linear Decision Boundary

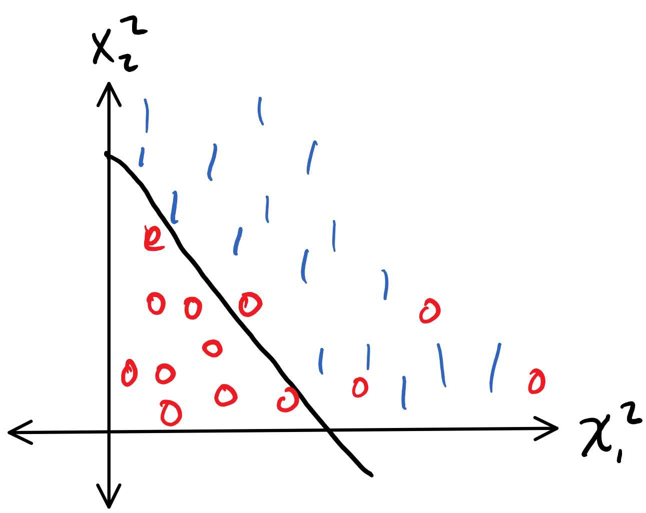


Non-linear decision boundary

def higher-dimensional embedding

$$\Phi : \mathbb{R}^p \to \mathbb{R}^D \qquad \Phi(x) \in \mathbb{R}^D$$

ex $\Phi(x_1, x_2) = (1, x_1, x_2, x_1^2, x_2^2)$



Φ makes linear methods into non-linear ones

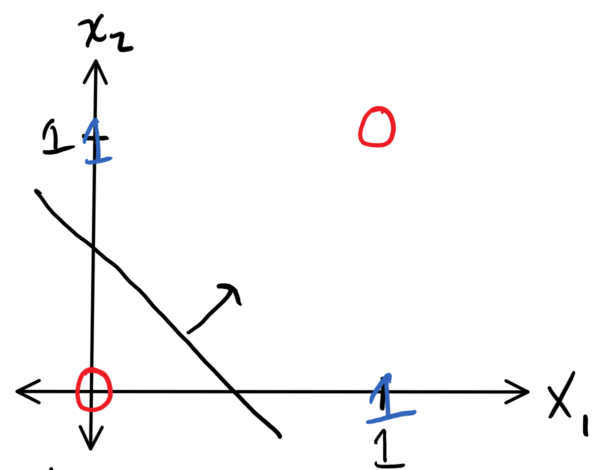
ex $x_1, \dots, x_p$ are proposition and we want to answer ~~logi~~ complex props such as $(x_1 \text{ or } x_2)$ and $(x_3 \text{ or } x_4)$

x_1 xor x_2 : $(x_1 \text{ and not } x_2)$ or $(x_2 \text{ and not } x_1)$

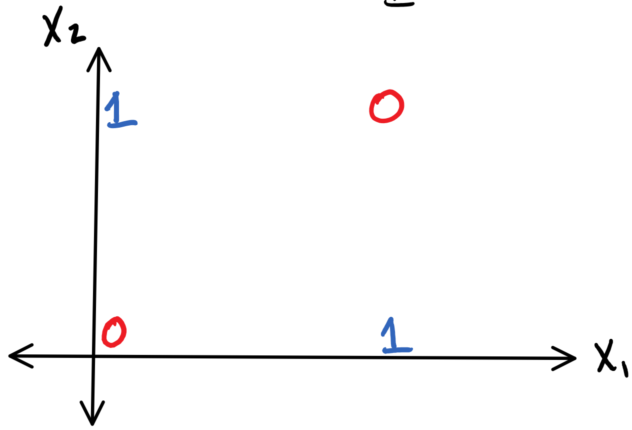
$x_1, x_2 \in \{0, 1\}$ False True

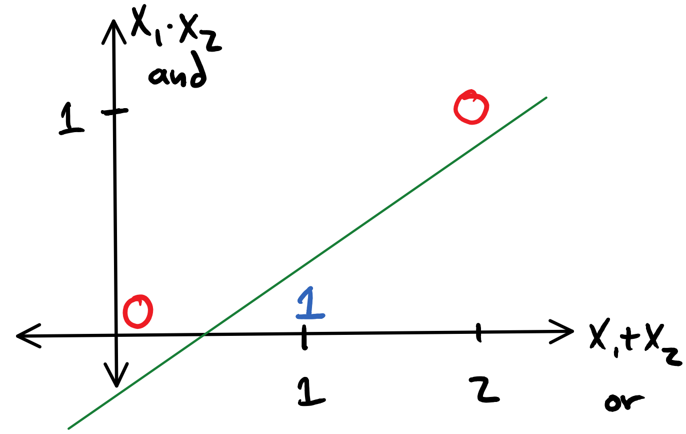
$y = x_1 \text{ xor } x_2$

y	x_1	x_2
0	0	0
1	0	1
1	1	0
0	1	1



$$\Phi(x_1, x_2) = (x_1, x_2, x_1 \cdot x_2)$$





Kernel Trick

Wednesday, May 8, 2019 1:23 PM

Let $Z_{ik} = \Phi_k(x_i)$ $\quad x_i \in \mathbb{R}^p \quad k = 1, \dots, D$

linear - SVM for $y_i \in \{-1, 1\}$

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \big(1 - \overset{(I)}{y_i \underbrace{z_i^T}_{\text{row}} \beta}\big)_+ + \overset{(II)}{\lambda \|\beta\|^2}$$

<u>claim</u> $\hat{\beta}$ solves SVM can be written $Z^T\alpha$, $\alpha \in \mathbb{R}^n$

i.e. $\hat{\beta}_j = \underbrace{Z_j^T}_{\text{column}} \alpha = \sum_i \alpha_i z_{ij} = \left(\sum_i \alpha_i z_i\right)_j$

$\beta = \underbrace{\sum_i \alpha_i z_i}_{Z^T\alpha} + \beta^{\perp} \qquad z_i^T \beta^{\perp} = 0 \quad \forall i$

$z_i^T \beta = \sum_j \alpha_j z_i^T z_j + \cancel{z_i^T \beta^{\perp}} \to 0$

$= z_i^T Z^T \alpha \quad \Rightarrow \beta^{\perp}$ does not impact (I)

$\|\beta\|^2 = \|Z^T\alpha + \beta^{\perp}\|^2 = \|Z^T\alpha\|^2 + \underbrace{2\beta^{\perp T} Z^T\alpha}_{= \alpha^T (\cancel{Z\beta^{\perp}}) \to 0} + \|\beta^{\perp}\|^2$

$= \|Z^T\alpha\|^2 + \|\beta^{\perp}\|^2$

$\Rightarrow \beta^{\perp}$ at the min is 0

Rewrite SVM:

$$\min_{\alpha \in \mathbb{R}^n} \frac{1}{n} \sum_i \left(1 - y_i z_i^T Z^T \alpha\right)_+ + \lambda \|Z^T\alpha\|^2$$

$$\hat{\beta} = Z^T \hat{\alpha} = \sum_i \hat{\alpha}_i x_i$$

<u>General</u>

$\min \quad R(y, Z\beta) + \lambda \|\beta\|^2$

$$\min_{\beta \in \mathbb{R}^D} R_n(y, Z\beta) + \lambda \|\beta\|^2$$

$$\parallel\!\!\mid$$

$$\min_{\alpha \in \mathbb{R}^n} R_n(y, K\alpha) + \lambda \alpha^T K \alpha$$

w/ $K = ZZ^T \quad (z_i^T Z^T \alpha) = (K\alpha)_i$

↳ kernel matrix

$$K_{ij} = z_i^T z_j = \Phi(x_i)^T \Phi(x_j) \quad (1)$$

$$\hat{\beta} = Z^T \hat{\alpha}$$

Method 1: define Φ, apply $\to Z$, make K

Method 2: define <u>kernel function</u>

$k(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$ (closed form)

compute K from (1)

predict $\Phi(x)^T \hat{\beta} = \Phi(x)^T \left(\sum_i \hat{\alpha}_i z_i\right)$

$$= \sum_i \hat{\alpha}_i \Phi(x)^T \Phi(x_i)$$

$$= \sum_i \hat{\alpha}_i k(x, x_i)$$

<u>ex</u> d^{th} degree poly. $k(x, x') = (1 + x^T x')^d$

$d=2$: $(1 + x_1 x_1' + x_2 x_2')^2 = 1 + 2x_1 x_1' + 2x_2 x_2' +$
$x_1^2 x_1'^2 + x_2^2 x_2'^2 + 2x_1 x_1' x_2 x_2'$

$$= \underbrace{(1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, x_1^2, x_2^2)^T}_{} (\ldots x' \ldots)$$

$\Phi(x)$

Kernel SVM

Wednesday, May 8, 2019 1:22 PM

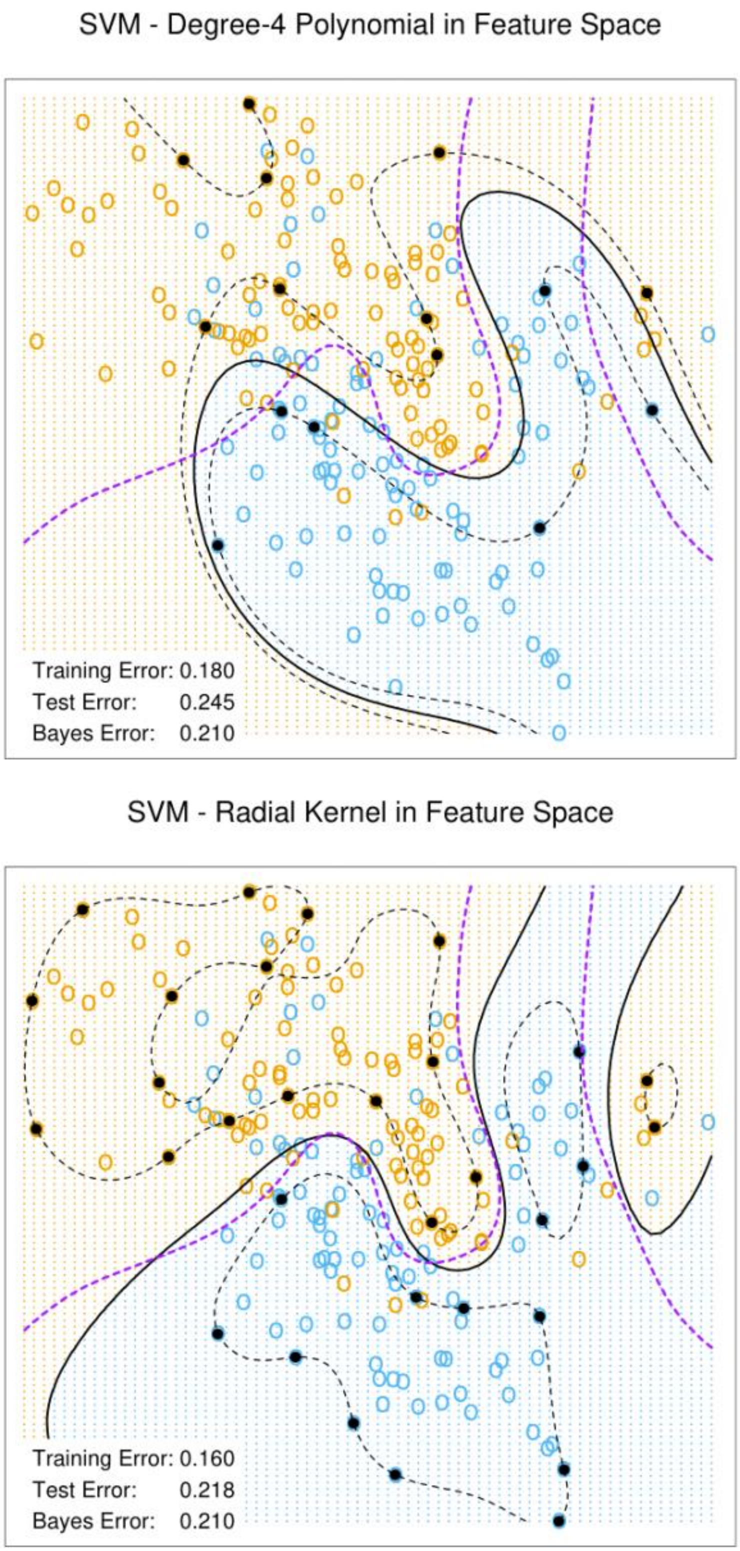


FIGURE 12.3. *Two nonlinear SVMs for the mixture data. The upper plot uses a 4th degree polynomial kernel, the lower a radial basis kernel (with $\gamma = 1$). In each case C was tuned to approximately achieve the best test error performance, and $C = 1$ worked well in both cases. The radial basis kernel performs the best (close to Bayes optimal), as might be expected given the data arise from mixtures of Gaussians. The broken purple curve in the background is the Bayes decision boundary.*

def Mercer kernel is a function

$k : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}_+$ that is PSD

(for any $\{x_i\} \subseteq \mathbb{R}^p$ $(k(x_i, x_j))_{ij}$ is PSD)

thm Every Mercer kernel has a Hi-di embedding

$\underline{\Phi}$ s.t. $k(x, x') = \underline{\Phi}(x)^T \underline{\Phi}(x')$

($\underline{\Phi}$ may be ∞ - dimensional)

ex $k(x, x') = e^{-\|x - x'\|^2 / \sigma^2}$

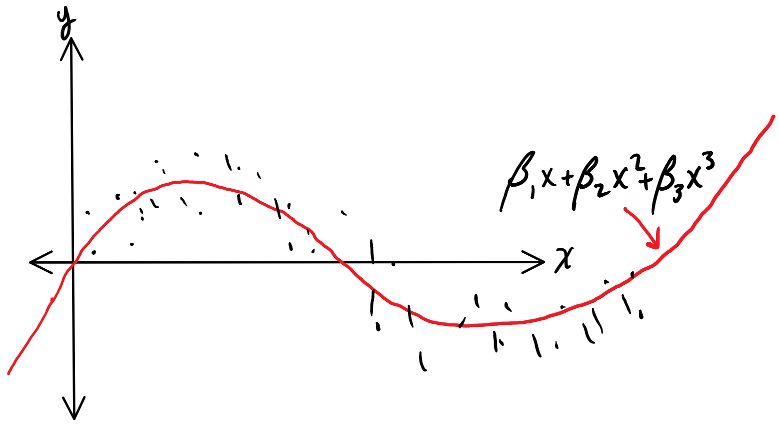
$\uparrow$ bandwidth parm

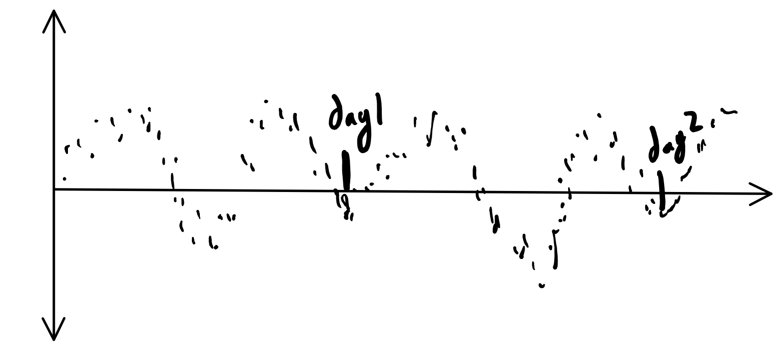
RBF in 1-D has

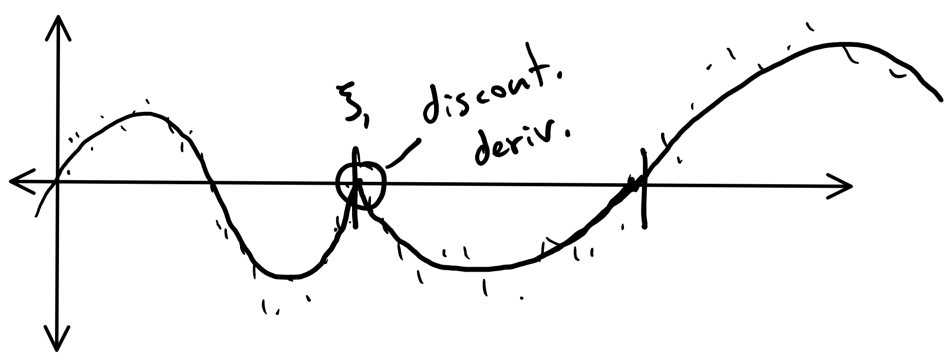
$$\underline{\Phi}(x) = e^{-x^2/2\sigma^2} \left[1, \sqrt{\frac{1}{1!\,\sigma^2}}\, x, \sqrt{\frac{1}{2!\,\sigma^4}}\, x^2, \dots \right]$$

Basis Expansion

Monday, May 8, 2017 8:42 PM





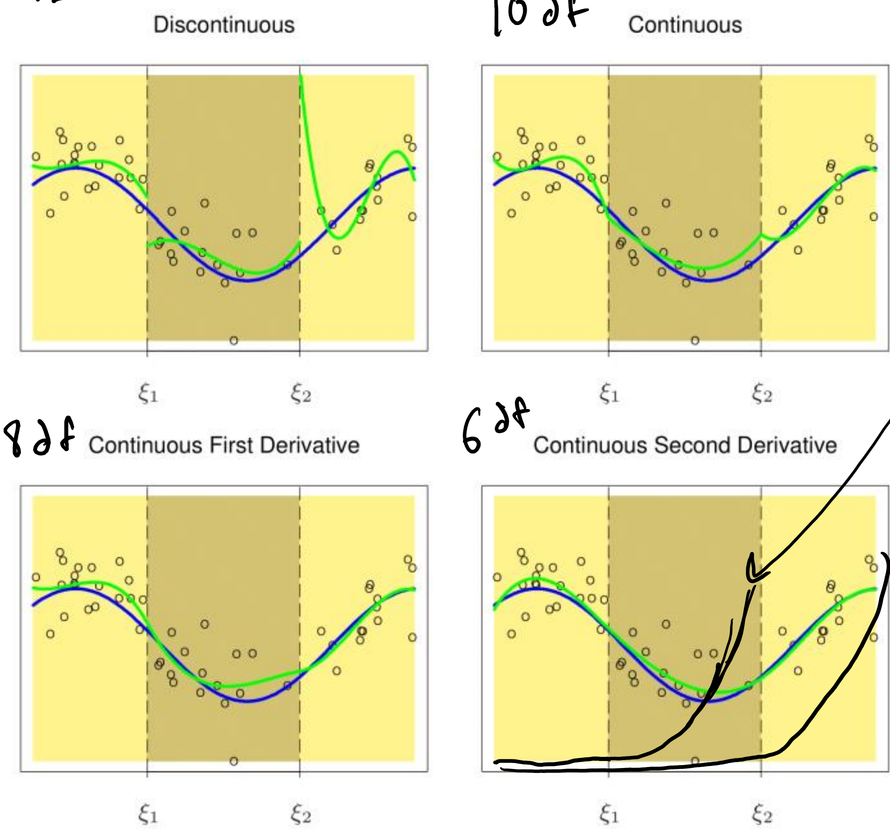


12 df

10 df

8 df

6 df



$\underline{1}, \underline{x}, \underline{x^2}, \underline{x^3},$

$\underline{(x-\xi_1)^3_+}, \underline{(x-\xi_2)^3_+}$

FIGURE 5.2. *A series of piecewise-cubic polynomials, with increasing orders of continuity.*

ESL 5.2