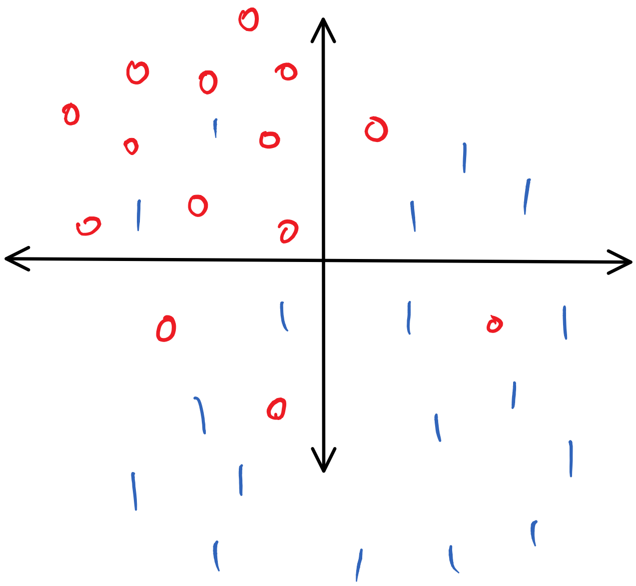
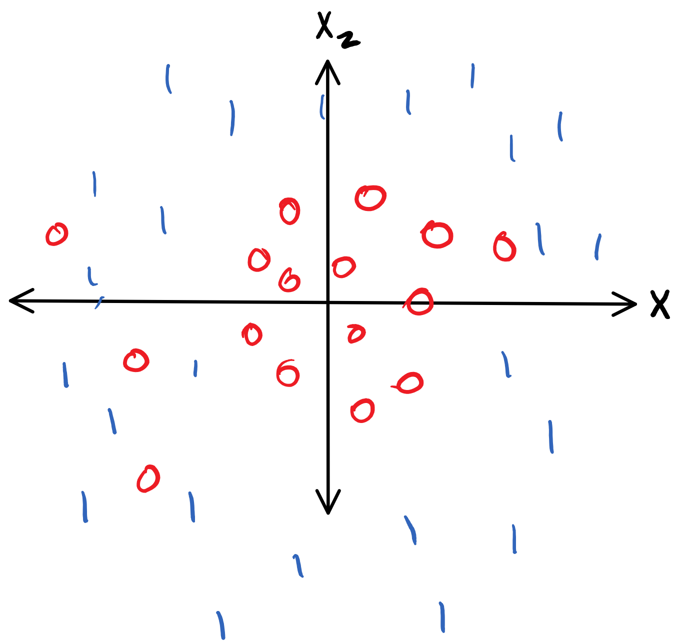
HiDi Embedding

Monday, May 8, 2017 8:11 PM



Linear decision boundary

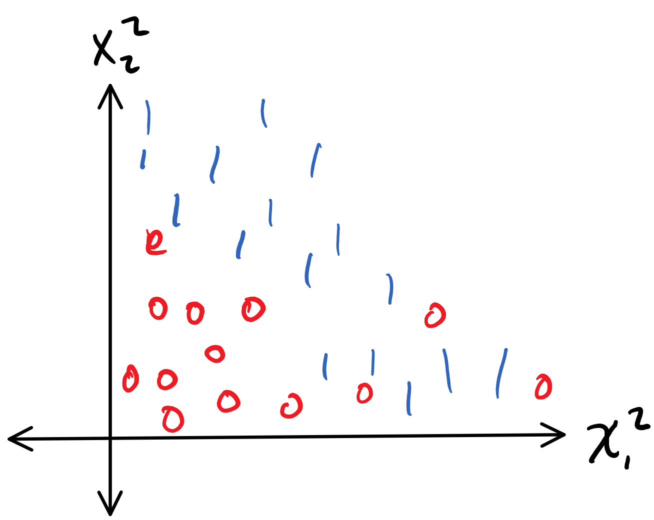


Non-linear decision boundary

<u>define</u> higher dimensional embedding $\Phi : \mathbb{R}^P \to \mathbb{R}^D$

$\Phi(x) \in \mathbb{R}^D$

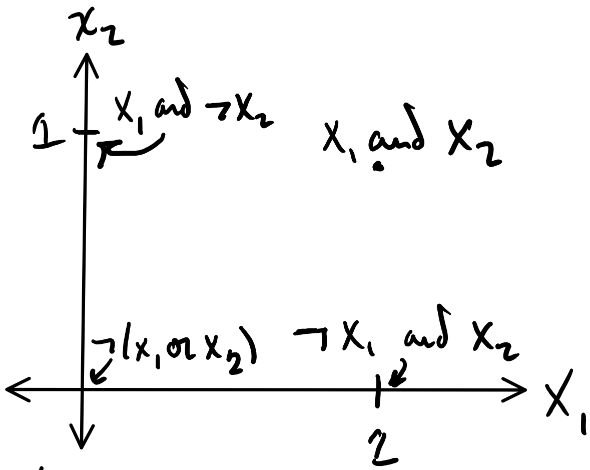
<u>ex</u> $\Phi(x_1, x_2) = (1, x_1, x_2, x_1^2, x_2^2)$

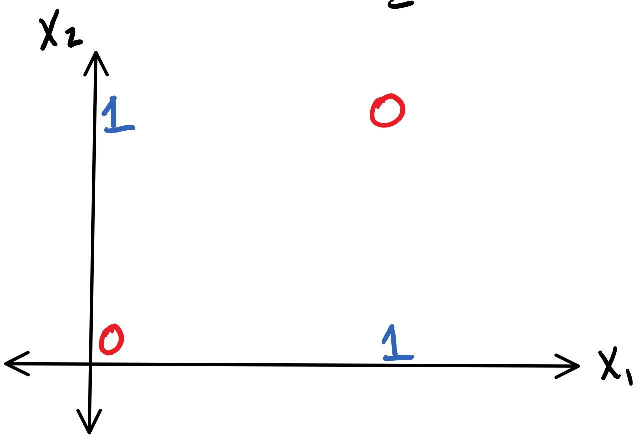


"Hi-di embedding can make linear methods non-linear"

— me, just now

ex Logic: $X_1, \dots, X_p$ are propositions encoded as $\{0,1\}$.





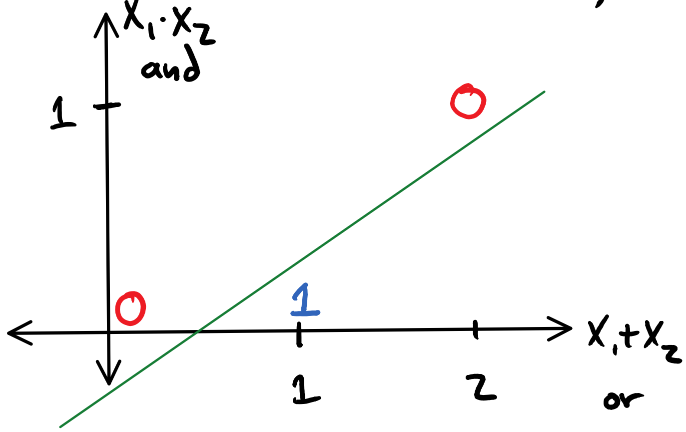
X_1 xor X_2 is

$(X_1 \text{ and } \neg X_2)$ or

$(X_2 \text{ and } \neg X_1)$

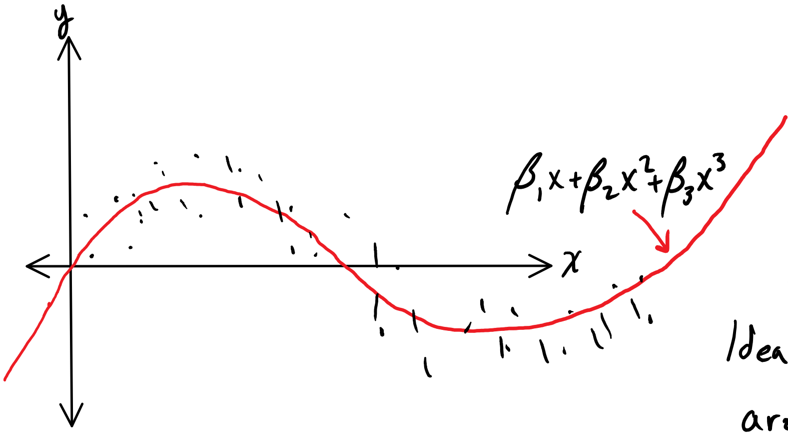
$\equiv (X_1 \text{ or } X_2) \text{ and } \neg(X_1 \text{ and } X_2)$

$$\Phi(X_1, X_2) = \Phi(X_1, X_2, X_1 \cdot X_2)$$



Basis Expansion

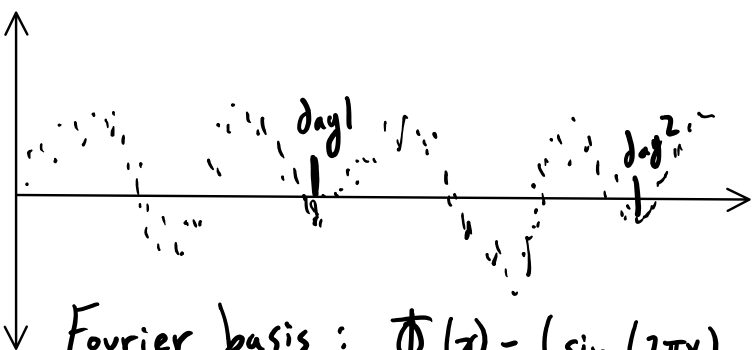
Monday, May 8, 2017 8:42 PM



Polynomial Basis

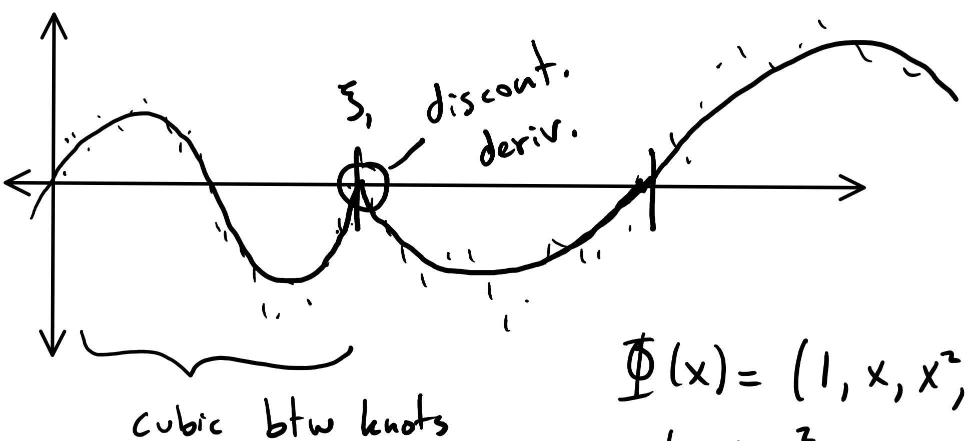
$\Phi(x) = (1, x, x^2, \ldots, x^d)$

Idea: Taylor expansion around $x = 0$.



Periodic (1 day period)

Fourier basis: $\Phi(x) = (\sin(2\pi x), \cos(2\pi x), \sin(4\pi x), \cos(4\pi x), \ldots)$



Piecewise polynomial

$\Phi(x) = (1, x, x^2, x^3, \underline{(x-\xi_1)_+}, \underline{(x-\xi_1)_+^2}, (x-\xi_1)_+^3, \ldots)$

<u>derivatives and constraints</u>

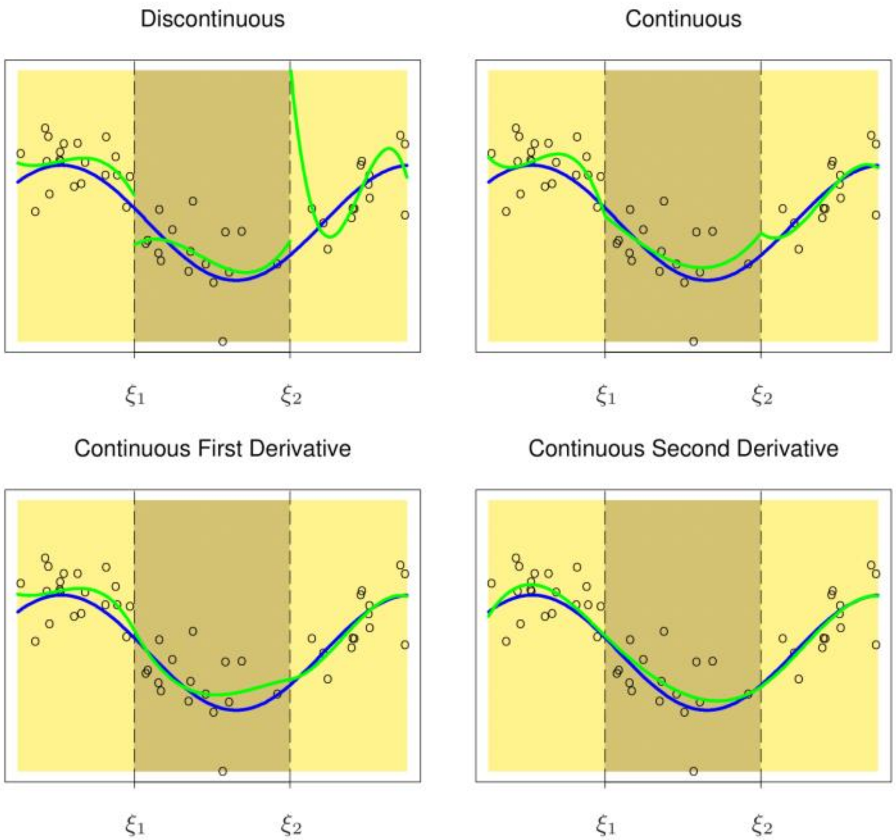


FIGURE 5.2. *A series of piecewise-cubic polynomials, with increasing orders of continuity.*

ESL 5.2

$\overline{III}$ / Change β changes fit everywhere (globally supported)

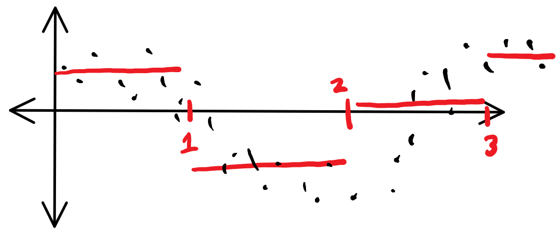
Localized Bases

Monday, May 8, 2017 9:55 PM

B-spline basis (cardinal)

knots at 1, 2, 3, ...

0th order $\varphi_i(x) = \mathbb{1}\{i-1 \le x < 1\}$



<u>translation</u> : $\varphi_{i+1}(x) = \varphi_1(x-i)$

<u>convolution</u> : $(g * h)(x) = \int_{-\infty}^{\infty} g(y)\, h(x-y)\, dy$

$(\varphi_1 * \varphi_1)(x) = \int \mathbb{1}\{0 \le y < 1\}\, \mathbb{1}\{0 \le x-y < 1\}\, dy$

<u>derive</u>

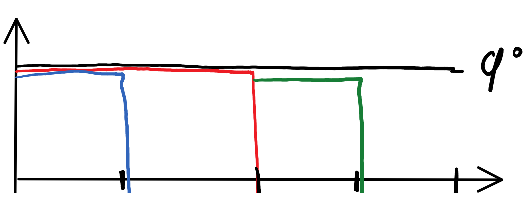
<u>def</u> $\varphi_1^{(k)} = \varphi_1^{(k-1)} * \varphi_1$ and $\varphi_{i+1}^{(k)} = \varphi_1(x-i)$

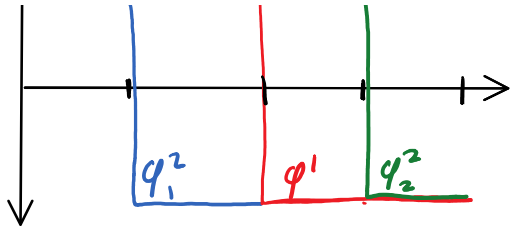
is the <u>k-th order (cardinal) B-spline basis</u>.

- localized, not orthogonal. $\left(\int \varphi_{i+1}^{(k)}(x)\, \varphi_i^{(k)}(x)\, dx \ne 0\right)$

<u>Haar Wavelets</u>

orthogonal!





- Wavelets are orthogonal
- Take "mother wavelet" φ^1 and translate and dilate $\varphi^{(k)}_{i+1}(x) = \varphi^1(2^k x - i)$
- Many other wavelets : Daubechies, Coiflets, etc.

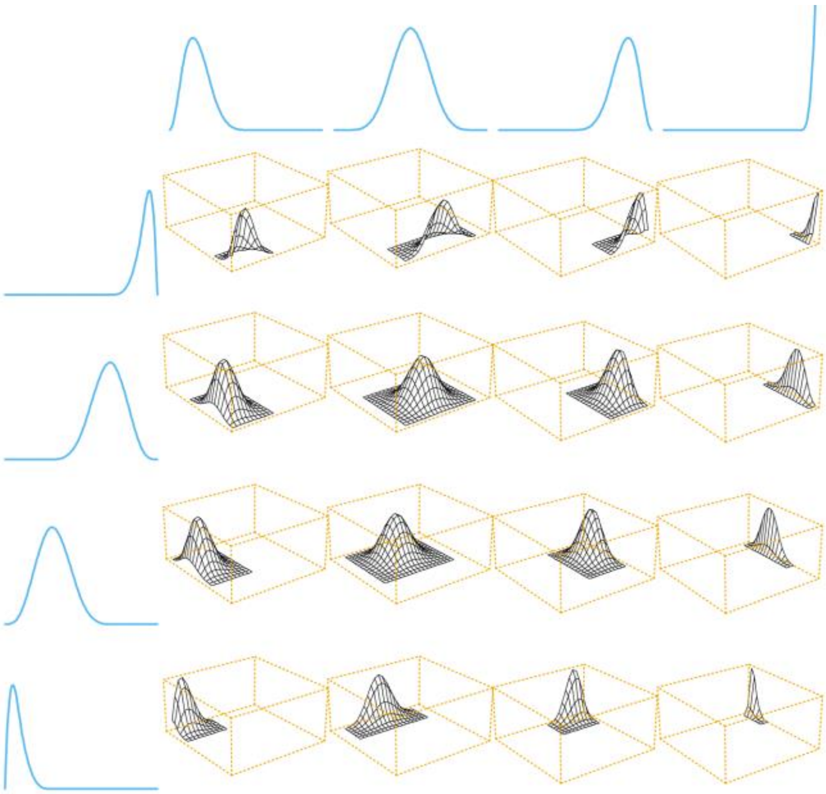
Define $Z_{j\ell} = \varphi_\ell(j)$ for ℓ'th basis element,

lasso: $\min_\beta \|y - Z\beta\|_2^2 + \lambda \|\beta\|_1$

Z orthogonal $\longrightarrow$ soft-thresholding

Multidimensional Bases $x \in \mathbb{R}^2$ then

$\varphi_{ij}(x) = \varphi_i(x_1) \cdot \varphi_j(x_2)$ is Tensor product basis.



ESL 5.7

Kernel Trick

Monday, May 8, 2017 10:32 PM

Let $Z_{jl} = \varphi_l(x_j)$ for any basis then

SVM's for $y_i \in \{-1, 1\}$

$$\min_{\beta} \underbrace{\frac{1}{n}\sum_{i=1}^{n}(1 - y_i Z_i^T \beta)}_{\text{function of } Z\beta} + \lambda \|\beta\|_2^2$$

Derive kernel trick

$$\beta = \sum_i \alpha_i Z_i + \beta^{\perp} \quad w/ \quad Z_i^T \beta^{\perp} = 0$$

<u>def</u> Mercer kernel is function

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

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

<u>ex</u> Radial basis function

$$k(x, x') = e^{-\frac{\|x - x'\|_2^2}{\sigma^2}} \quad \sigma \text{ is bandwidth}$$

<u>thm</u> Every mercer kernel has a Hidi embedding

Φ (perhaps ∞-dimensional) s.t.

$$k(x, x') = \Phi(x)^T \Phi(x').$$