# Supplement B1: Data, Features, and Approximations

## Feature Vectors and Mappings

The ***feature vector*** for some input $\mathbf{x}$ is the vector

$$\phi(\mathbf{x}) = (\phi_0(\mathbf{x}), \phi_1(\mathbf{x}), \ldots, \phi_{M-1}(\mathbf{x}))^T$$

For ***basis function*** $\phi_j : \mathbb{R}^D \to \mathbb{R}$ and typically $\phi_0(\mathbf{x}) = 1$

- Our ***data points*** , $\mathbf{x}_n$ , are said to live in ***data space*** $(\subseteq \mathbb{R}^D)$
- Equally, ***feature vectors*** live in ***feature space*** $(\subseteq \mathbb{R}^M)$
- Typically $M > D$

The feature mapping, $\phi$, takes a data point and gives the corresponding feature vector, i.e.:

$$\phi : \mathbb{R}^D \to \mathbb{R}^M$$

## Linear Models and Feature Space

If you define $M$ basis functions, then your ***feature vector*** will have $M$ elements, i.e. $\phi(\mathbf{x}) \in \mathbb{R}^M$.

Predictions are from linear model:

$$y(\mathbf{x}) = \phi(\mathbf{x})^T \mathbf{w}$$

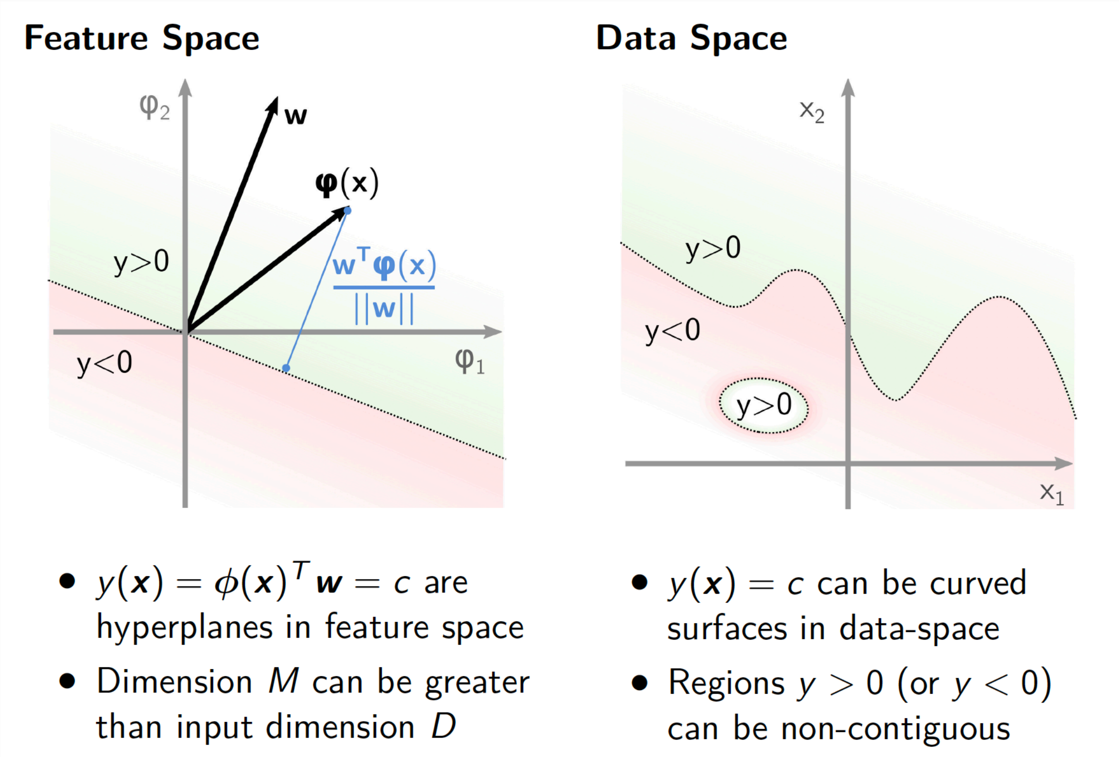
and so weights are $M$-vectors ($\mathbf{w} \in \mathbb{R}^M$)

For constants $\mathbf{b}, \mathbf{c} \in \mathbb{R}^D$, a $D-1$ dimensional plane (***hyper-plane***) is defined by:

$$\mathbf{b}^T \mathbf{z} = \mathbf{c}$$

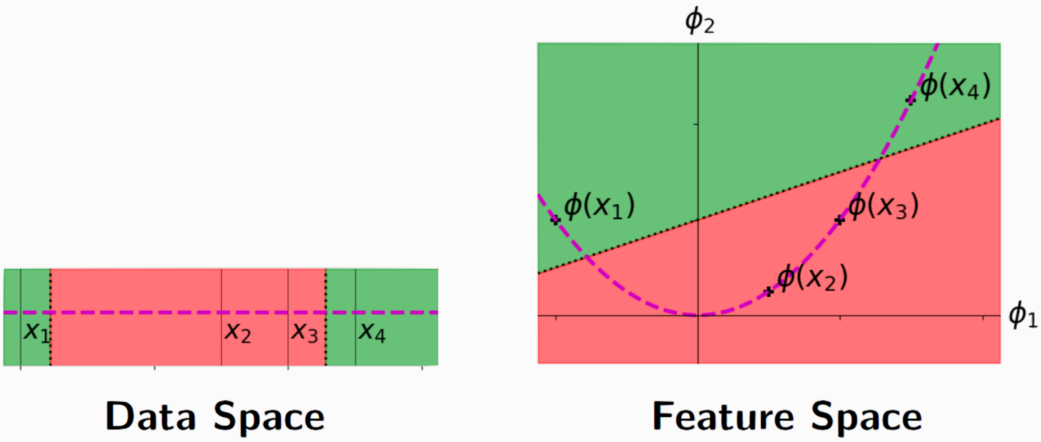
for some $D$-dimensional vector $\mathbf{z}$

## Linear Models: Geometric Intuition




### Example 1: Data vs. Feature Space

A ***scalar data space*** ($x \in \mathbb{R}$) and a *2-dimensional feature space* ($\phi(x) \in \mathbb{R}^2$) with $\phi_1(x) = x$ and $\phi_2(x) = x^2$

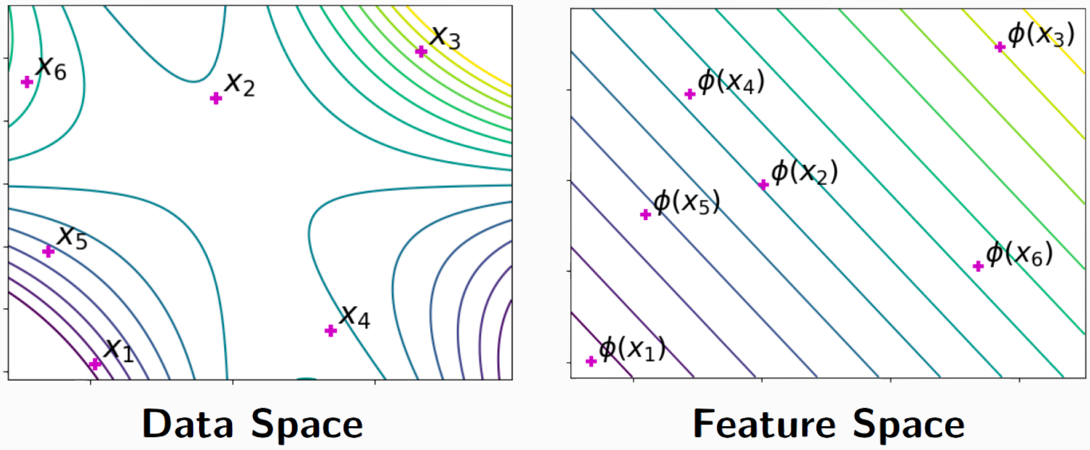



> $\phi_1$ ***axis*** in Feature Space is the ***real axis*** in Data Space, $\phi_2$ ***axis*** in Feature Space follows a $x^2$ ***relationship***. The *purple* real line is curved

A linear function in features defines positive region in green and negative regions in red. In data space, these regions are ***non-contiguous***

### Example 2: Data vs. Feature Space

A 2-dimensional data space ($\mathbf{x} \in \mathbb{R}^2$) and a 2-dimensional feature space ($\phi(x) \in \mathbb{R}^2$), with $\phi_1((a, b)) = a^2 b$ and $\phi_2((a, b)) = ab^2$




The contours of a linear function over features are shown. In feature space, these are linear and equally spaced. In data space, the function is clearly non-linear.

## Euclidean Distance

***Radial basis function*** are defined in terms of the ***Euclidean distance*** between 2 points. Let's call these points $\mathbf{x}, \mathbf{y} \in \mathbb{R}^D$. The Euclidean distance between $\mathbf{x}$ and $\mathbf{y}$ is then:

$$||\mathbf{x} - \mathbf{y}|| = \sqrt{\sum_{d=1}^{D} (x_d - y_d)^2}$$

We can write the square of this in ***vector form*** as:

$$||\mathbf{x} - \mathbf{y}||^2 = (\mathbf{x} - \mathbf{y})^T (\mathbf{x} - \mathbf{y})$$

## Radial Basis Functions

For vector inputs, $\mathbf{x}_n \in \mathbb{R}^D$, a ***Radial Basis Function***, $\phi_j$, needs a centre $\mu_j \in \mathbb{R}^D$ and s scale $s \in \mathbb{R}$, giving:

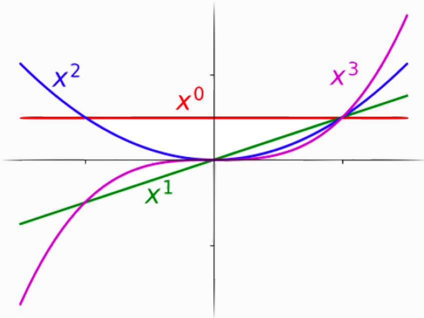
$$\phi_j(\mathbf{x}) = \exp(-\frac{||\mathbf{x} - \mu_j||^2}{2s^2})$$

Commonly, the same scale, $s$, is shared across all basis functions. This has similarities with the ***Multivariate Normal Distribution*** with ***isotropic covariance***.

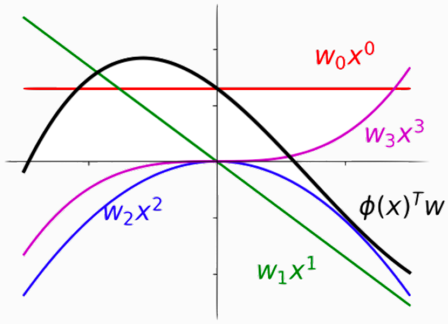
> $\phi_j(\mathbf{x})$ is closer to $1$ as the data points are closer to centre and is closer to 0 as the data points are furthur away from the centre.

## Monomial Basis Function

A ***linear model*** with *monomial basis functions* is ***a weighted mixture of monomial components.*** Below on the left are the raw basis functions, and on the right are ***weighted components*** (coloured) and the ***composite function*** (black).



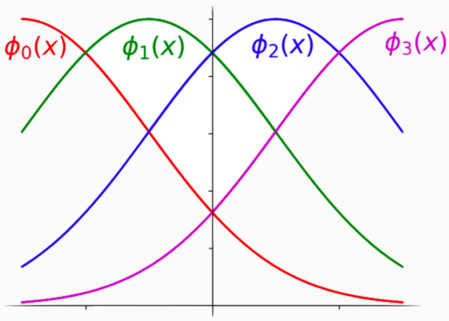

Basis functions



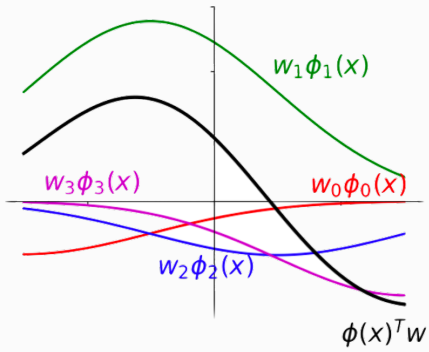

Weighted components and composite

## RBF Basis Functions

A linear model with ***radial basis functions*** (RBFs) is a ***weighted mixture of RBF components***. Below on the left are the raw basis functions, and on the right are ***weighted components*** (coloured) and the ***composite function*** (black).




Basis functions




Weighted components and composite

> ***Note that:***
>
> We can consider all of the composite functions (black) that we can construct from our weighted components as a ***function space***, which is a set of functions available to us that we can fit to our data.