Generalized Linear Models (GLM) and Radial Basis Functions (RBF)

Lyle Ungar

Computer and information Science

Learning Objectives

Extend linear regression with link functions,
basis functions

Know RBF algorithm and its uses

Generalized linear models (GLM)

- **Linear Model:** $\hat{y}(\boldsymbol{x}) = \sum_{j=1}^{p} w_j x_j$
- **GLM with link fn $f()$:** $\hat{y}(\boldsymbol{x}) = f(\sum_{j=1}^{p} w_j x_j)$
- **Basis transformation:** $\hat{y}(\boldsymbol{x}) = \sum_{j=1}^{d} w_j \phi_j(\boldsymbol{x})$

Based on slide by Geoff Hinton

Link functions

- **Link function** $\hat{y} = f(\boldsymbol{w}^T\boldsymbol{x})$
  - $f(x) = e^x$
  - $f(x) = log(x)$
- **Equivalent to** $f^{-1}(\hat{y}(\boldsymbol{x})) = \boldsymbol{w}^T\boldsymbol{x}$

Linear Basis Function Models

- Generally,

$$\hat{y}(\boldsymbol{x}) = \sum_{j=1}^{d} w_j \underbrace{\phi_j(\boldsymbol{x})}_{\text{basis function}}$$

- Typically, $\phi_0(\boldsymbol{x}) = 1$ so that θ_0 w_0 acts as a bias
- In the simplest case, we use linear basis functions

$$\phi_j(\boldsymbol{x}) = x_j$$

- Could use polynomials or Gaussians

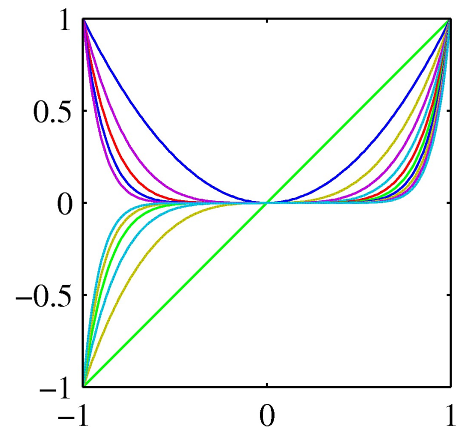
Based on slide by Christopher Bishop (PRML)

Linear Basis Function Models

- **Polynomial basis functions**

$$\phi_j(x) = x^j$$

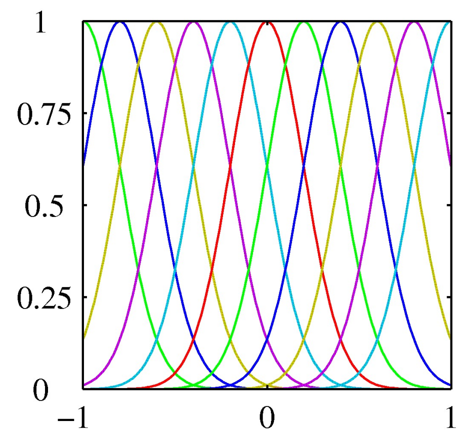
  - **Global – mostly crappy**



- **Gaussian basis functions:**

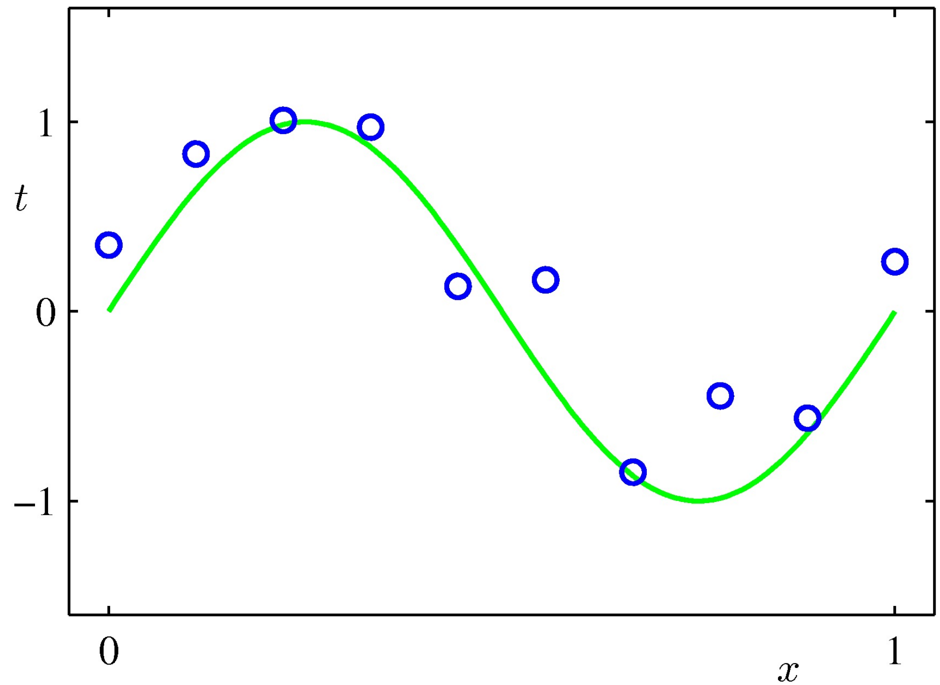
$$\phi_j(x) = \exp\left\{-\frac{(x-\mu_j)^2}{2s^2}\right\}$$

  - **Local – good!**



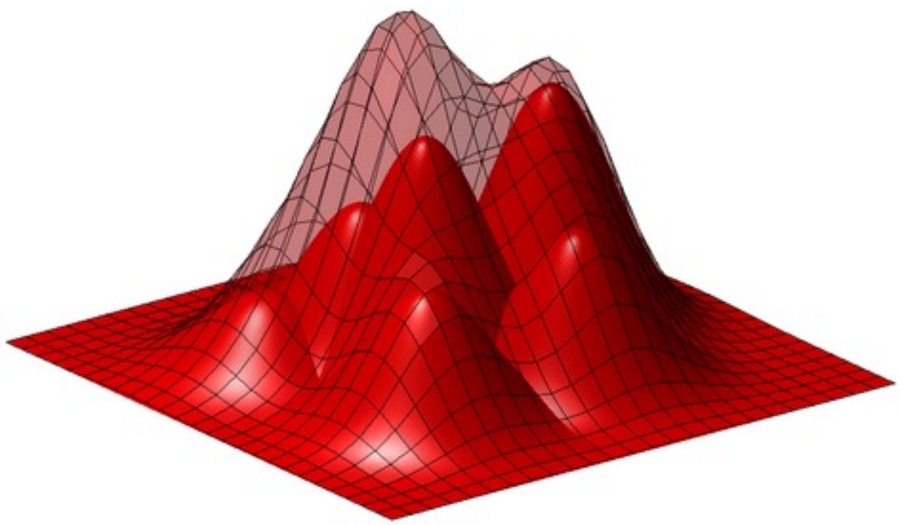
Based on slide by Christopher Bishop (PRML)

Fitting a Polynomial Curve with a Linear Model



$$y = \theta_0 + \theta_1 x + \theta_2 x^2 + \ldots + \theta_p x^p = \sum_{j=0}^{p} \theta_j x^j$$

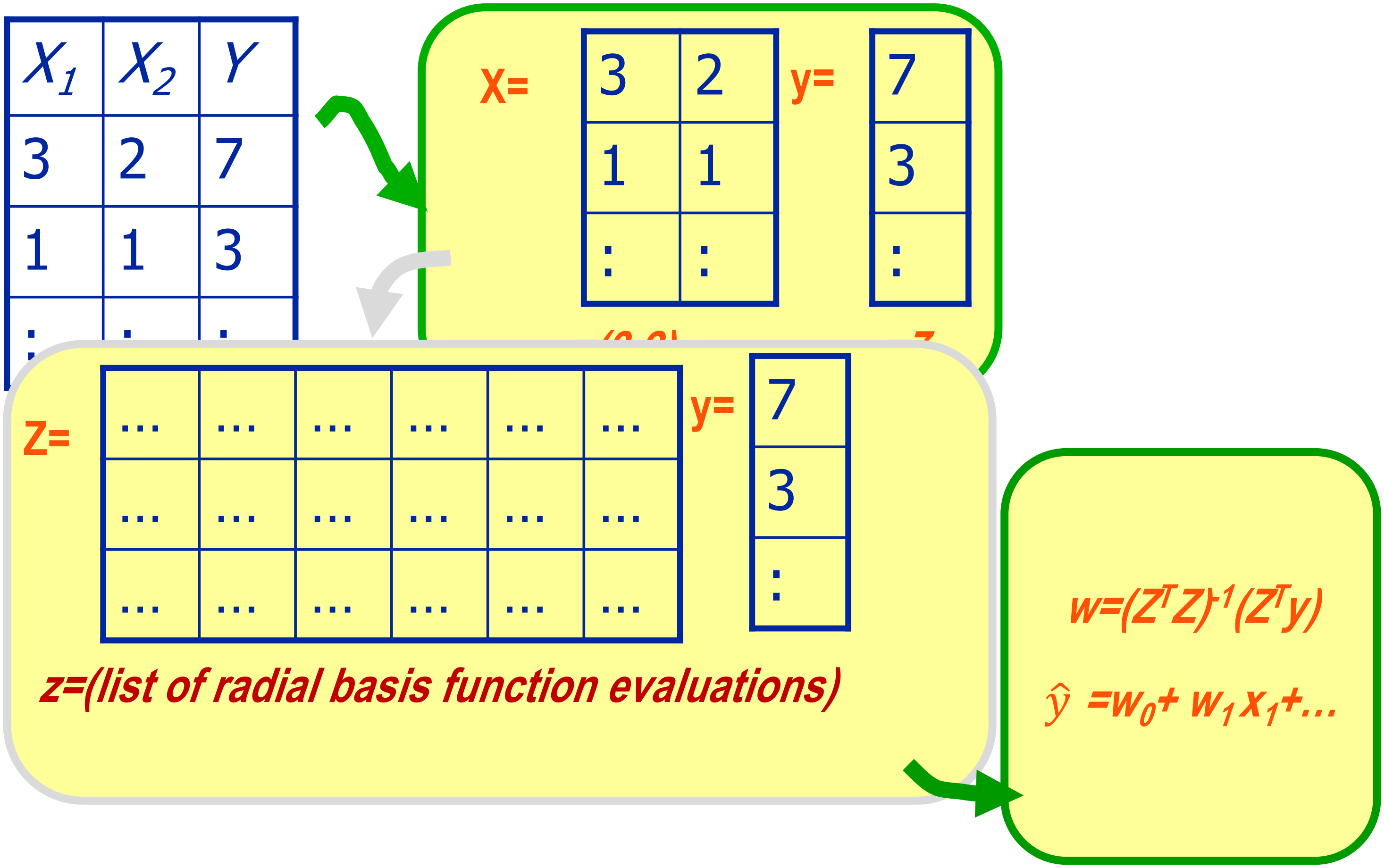
Radial Basis Functions



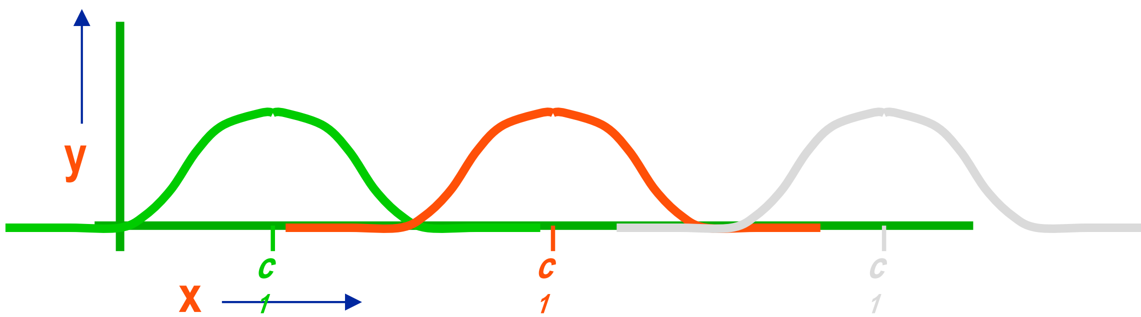
http://www.it.uu.se/research/project/rbf/rbf.png

Originally by Andrew Moore;
now heavily edited by Lyle Ungar

Radial Basis Functions (RBFs)



1-D RBFs



$\hat{y} = w_1 \phi_1(x) + w_2 \phi_2(x) + w_3 \phi_3(x)$

where

$\phi_i(x) = k(\|x - \mu_j\| / C)$

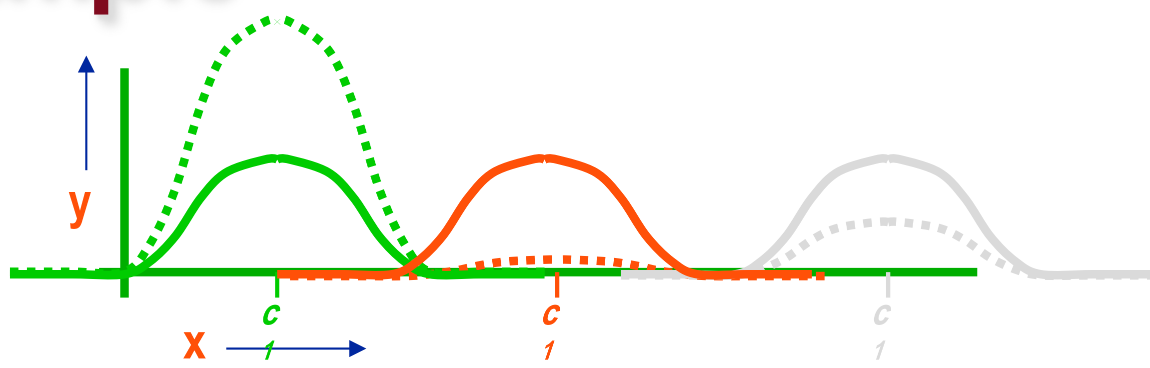
For RBF:

$k(\| x - \mu_j\| / C) = \exp\{-\|x - \mu_j\|_2^2 / C\}$

C = "Kernel Width"

k = kernel function

Example



$\hat{y} = 2\phi_1(x) + 0.05\phi_2(x) + 0.5\phi_3(x)$

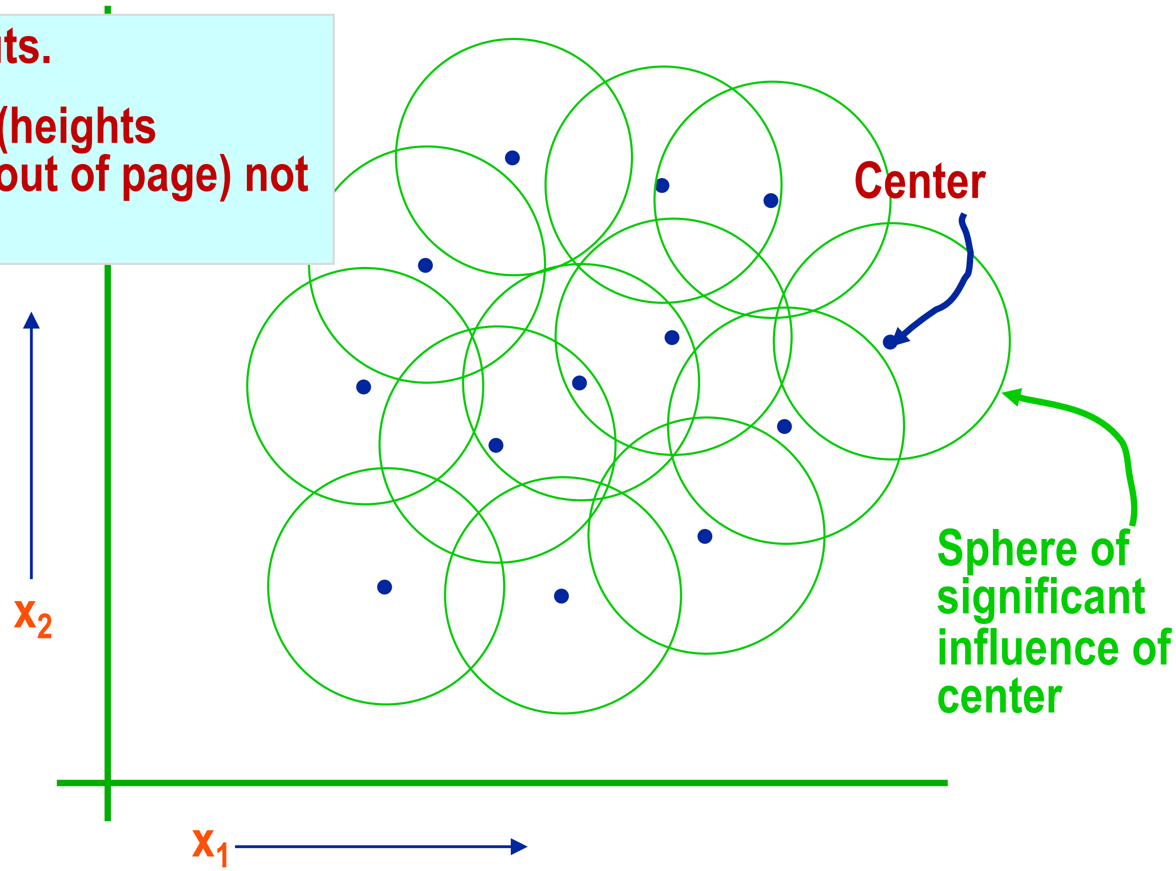
where

$\phi_j(x) = k(|| x - \mu_j|| / C)$

Radial Basis Functions in 2-d

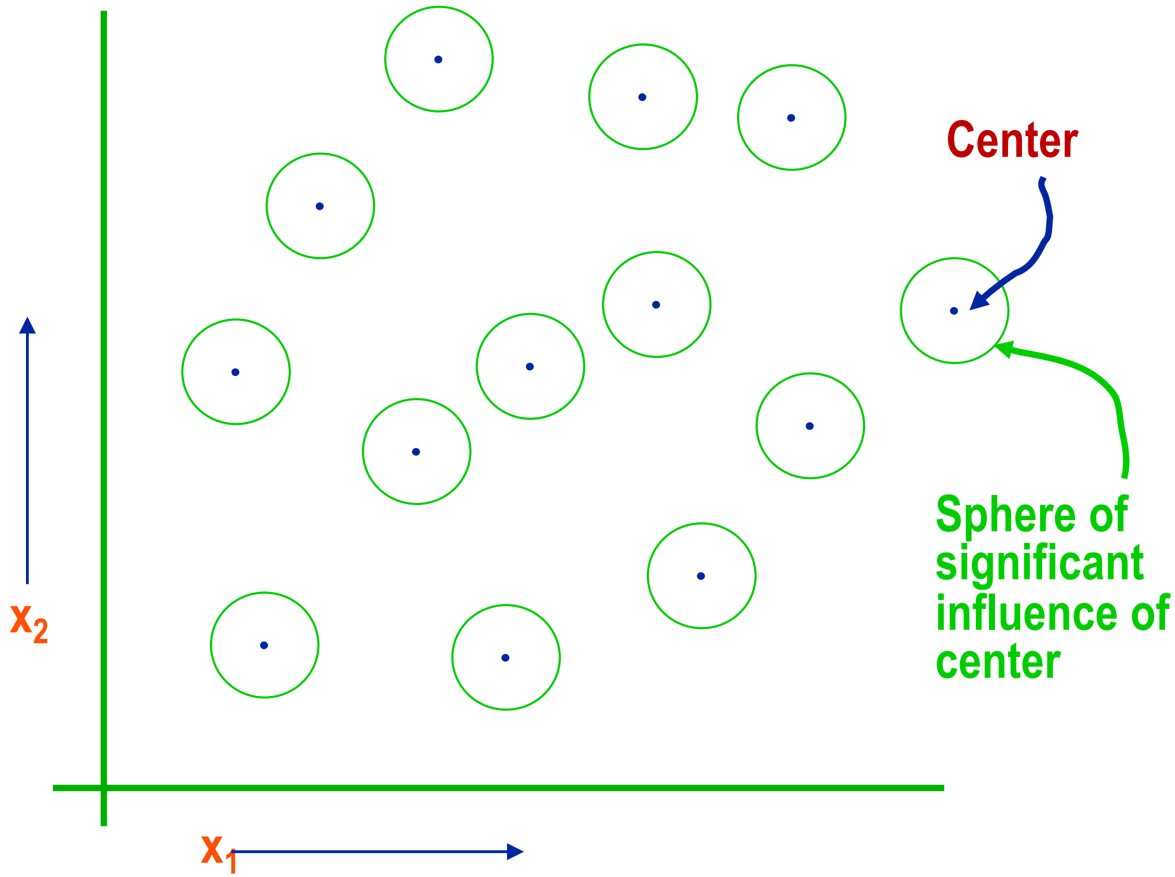
Two inputs.

Outputs (heights sticking out of page) not shown.

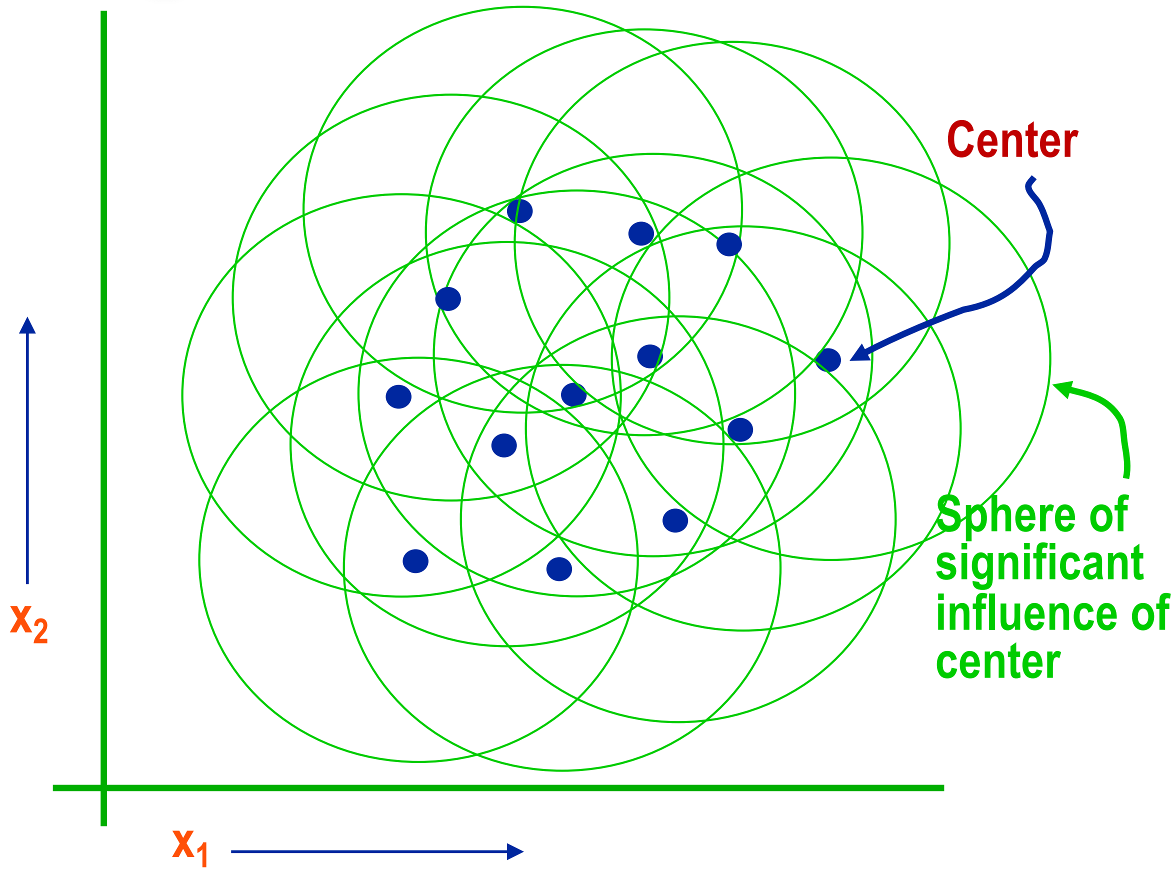


Too small!

Even before seeing the data, you should understand that this is a disaster!



Too big!!!



So what do we do?

Search to find the optimal size "width" for the Gaussians (on a test set, of course!)

RBFs can do …

- **Use $d < p$ basis vectors**
  - Dimensionality reduction
  - Good for high dimensional feature spaces
- **Use $d > p$ basis vectors**
  - Increases the dimensionality
  - Can make a formerly nonlinear problem linear
- **Use $d = n$ basis vectors**
  - We can use this to switch to a *dual* representation

How to find the kernel centers?

- **Pick random points**
  - Generally a bad idea
- **Standard RBF: do k-means clustering and use the centers of the clusters**
  - Works great!
- **Use all n of the training data points as kernel centers**
  - Requires regularization
- **Estimate them: nonlinear regression**
  - A good initialization helps

What you should know

- **Link functions give a nonlinear regression**
- **Basis functions allow one to fit a nonlinear function using linear regression**
- **RBF**
  - Cluster points
  - Put a Gaussian basic function at each cluster center
    - Pick the Gaussian width
  - Fit a linear regression

How is my speed?