David R. Thompson
Jet Propulsion Laboratory, California Institute of Technology

Local Methods for Pattern Recognition

Sponsored by the Jet Propulsion Laboratory and by the Keck Institute for Space Studies, California Institute of Technology

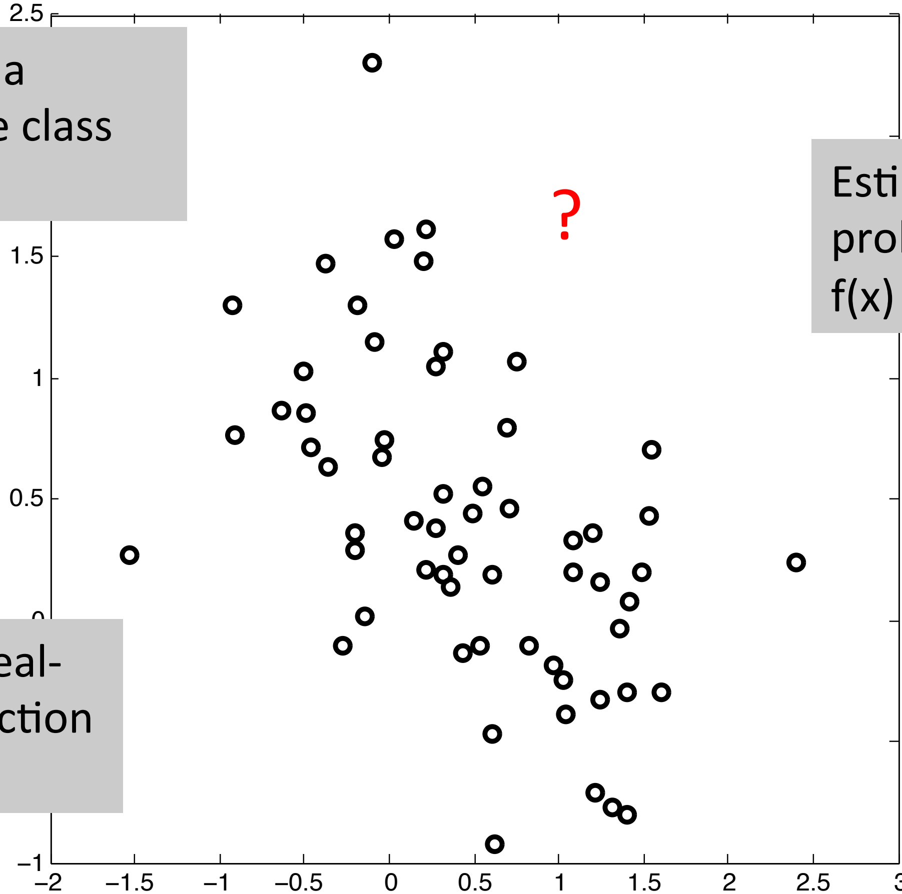
Objectives

1. Review classification and regression
2. Review nearest-neighbor methods
3. Local linear regression
4. Kernel density estimation

Kernel density

Simple pattern recognition task

Predict a discrete class label



Estimate the probability density f(x)

Predict a real-valued function y = g(x)

Simple pattern recognition task

datapoint

n features (attributes)

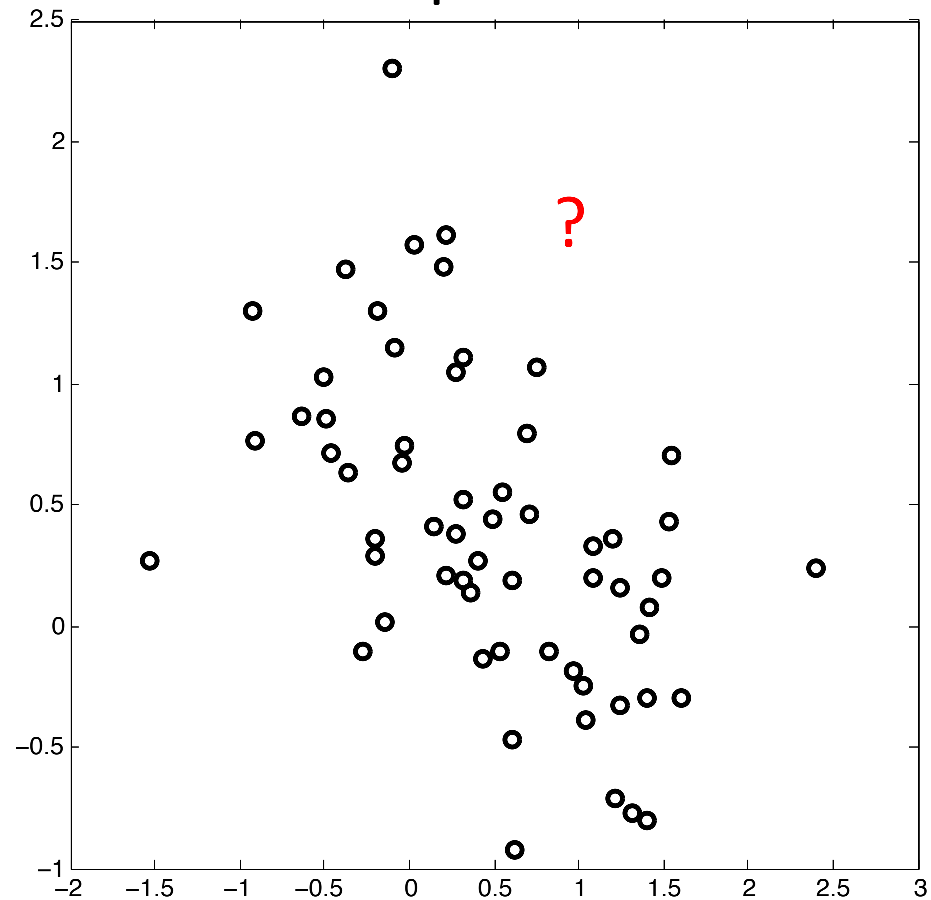
$$\mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \dots \\ x_{in} \end{bmatrix}$$

$$\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, \dots \mathbf{x_d}]$$

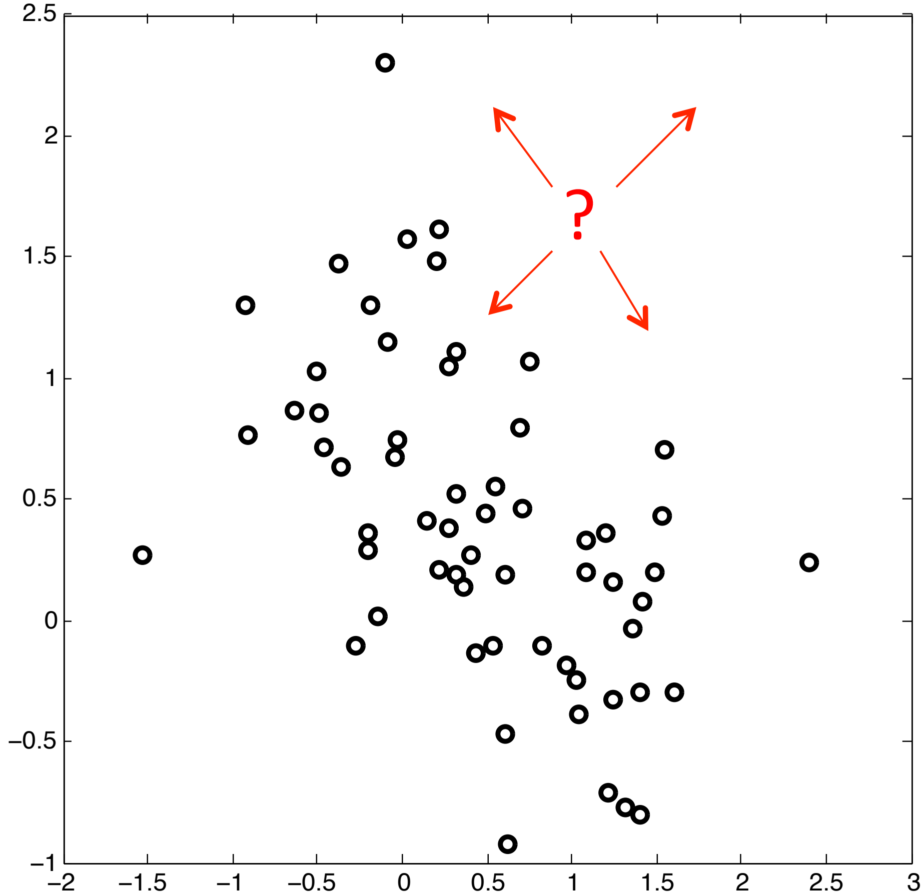
data set

n x d matrix

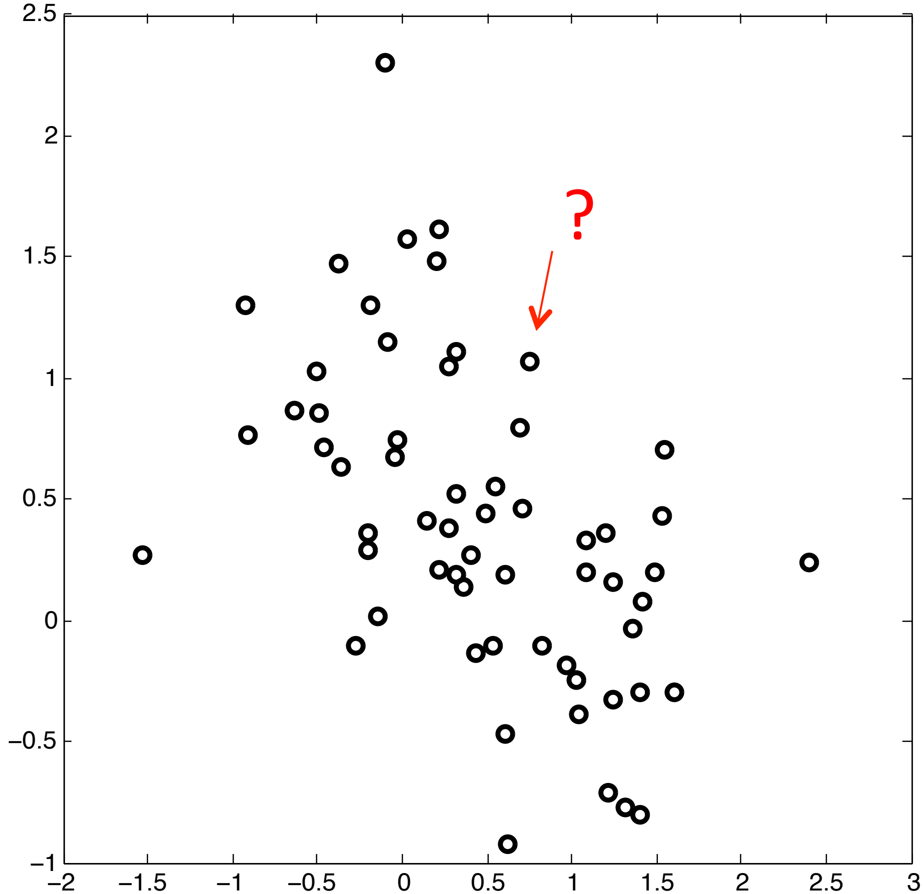
Example for n=2



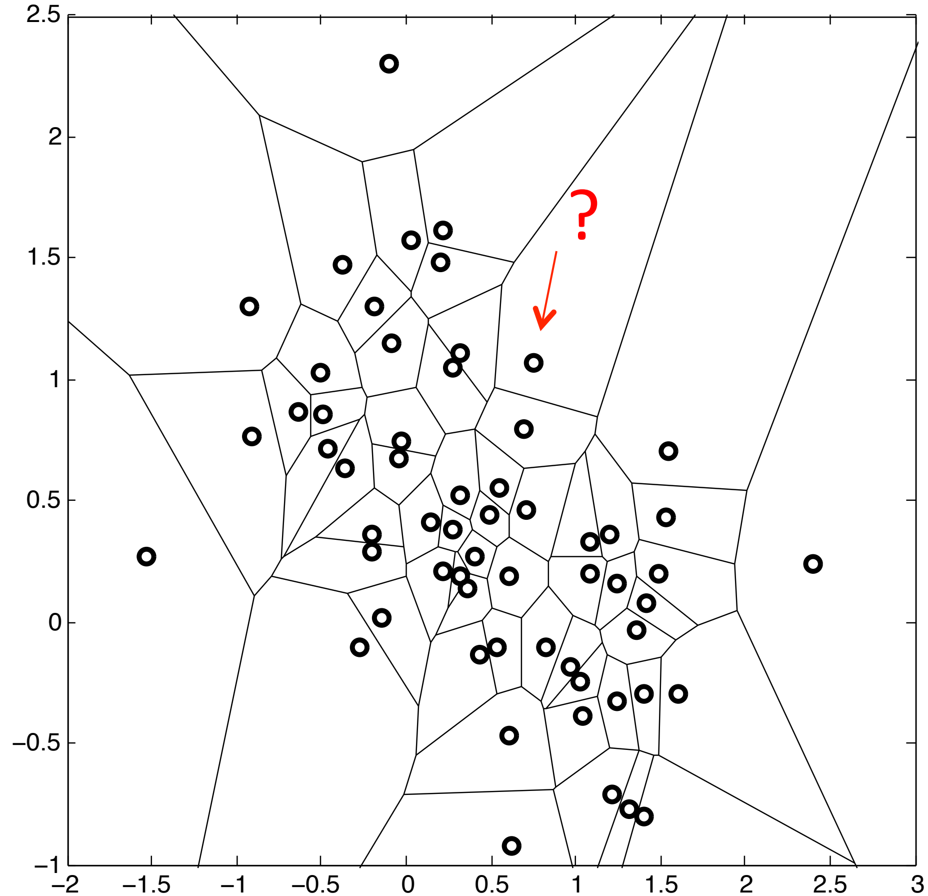
Look at local behavior



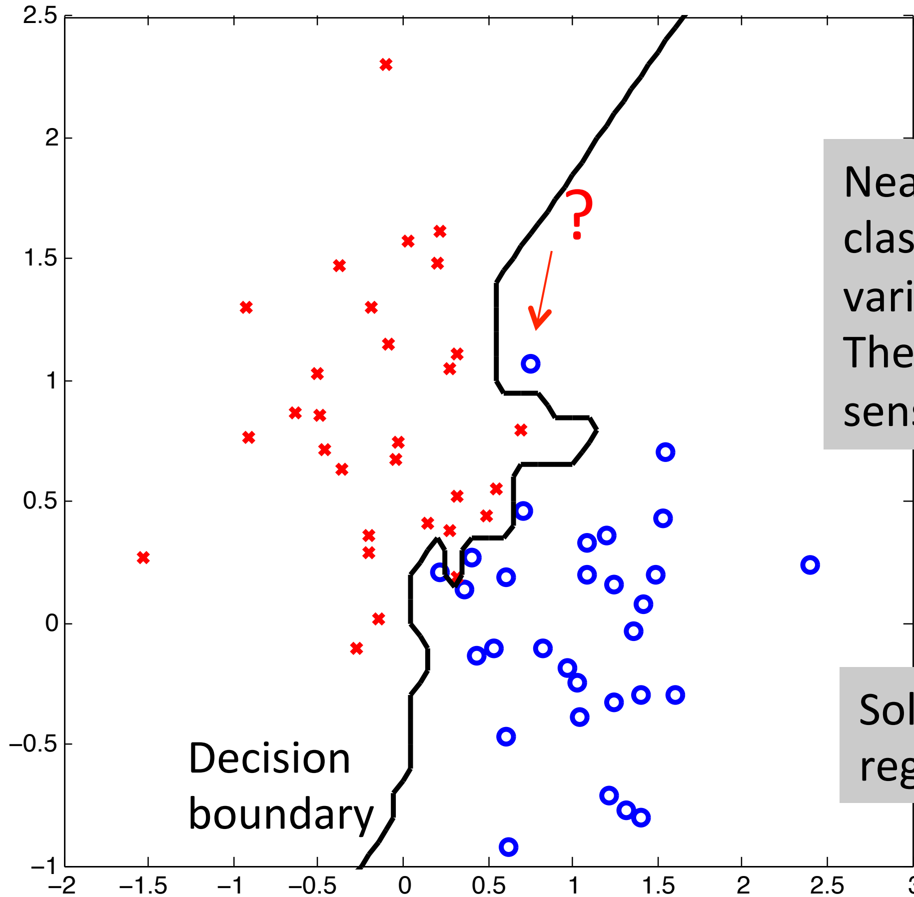
Nearest neighbor classification



Equates to a Voronoi partitioning



Binary classification

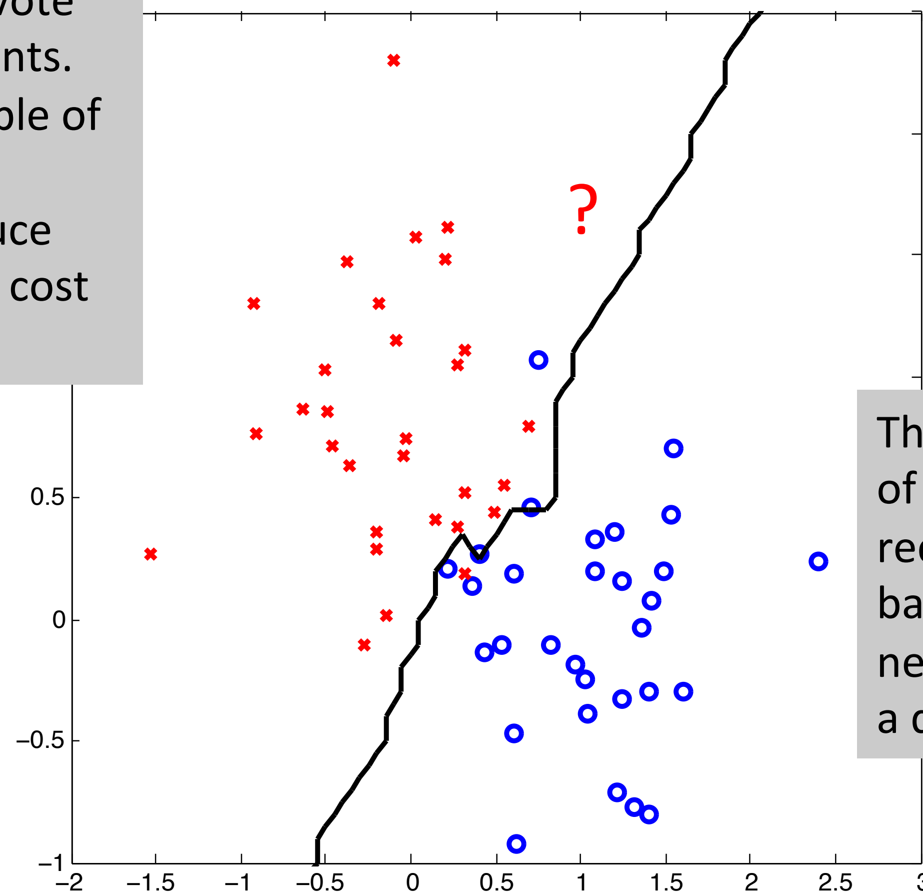


Nearest neighbor classification has high variance…
The solution is sensitive to noise.

Solution: regularizaton!

K-nearest neighbor, k=5

Use a majority vote of nearest 5 points. This is an example of a *regularization* strategy to reduce variance, at the cost of bias.



This is characteristic of many local pattern recognition methods based on the neighborhood around a query point.

Other local methods: kernel smoothing for regression

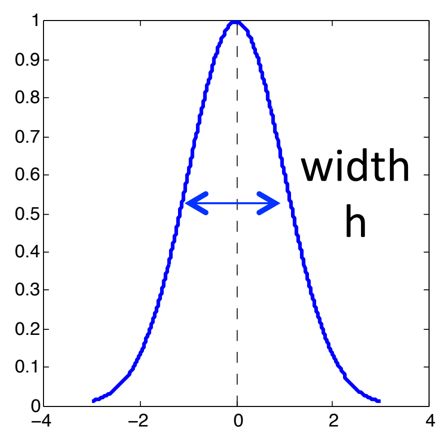
kernel

training data

$$\hat{y}(x_0) = \frac{\sum_i k(x_0, x_i) y(x_i)}{\sum_i k(x_0, x_i)}$$

predicted

A local average, weighted by some kernel function k()

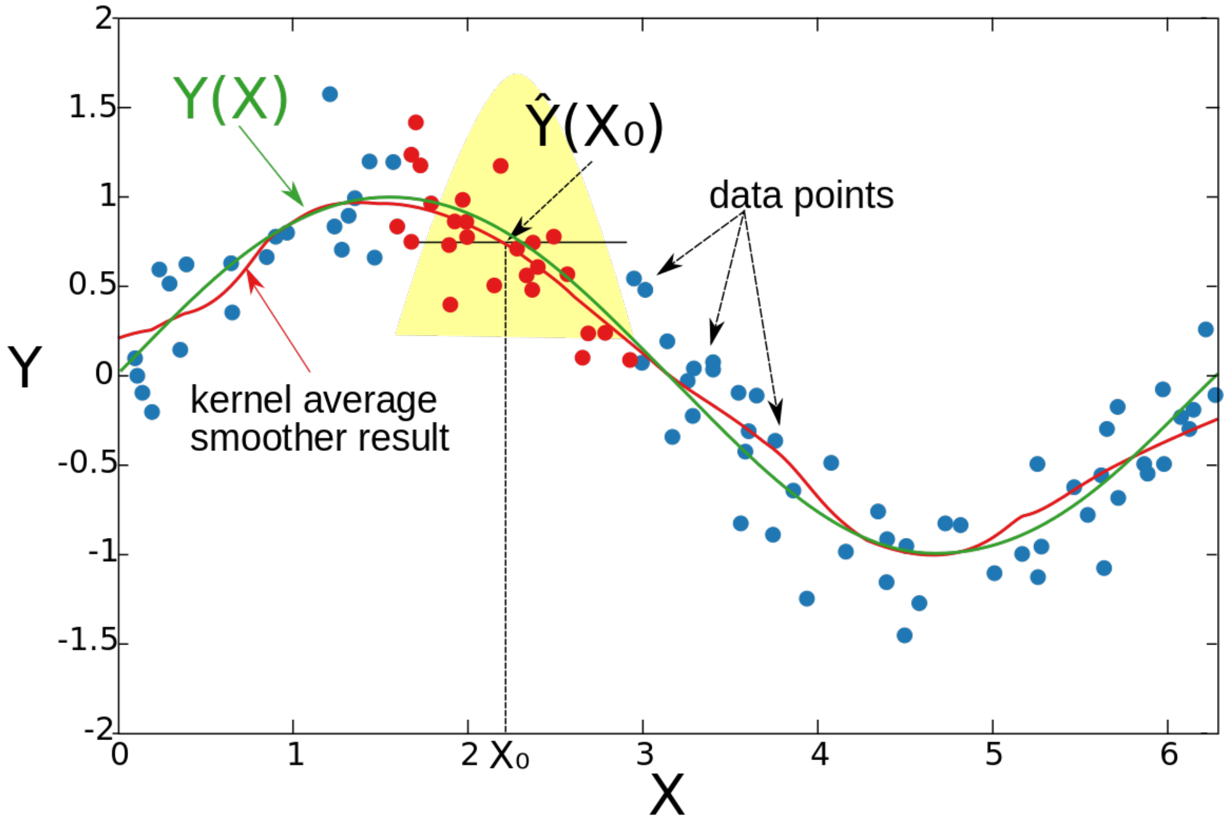


Example: a Gaussian kernel function. Can set the width using cross validation

$$k(x_0, x_i) = \frac{1}{z} e^{-\frac{h(x_0 - x_i)^2}{h}}$$

david.r.thompson@jpl.nasa.gov

Other applications: kernel smoothing for regression



Source: Wikipedia

david.r.thompson@jpl.nasa.gov

Local linear regression

$$\hat{Y}(\mathbf{X}_0) = (1, \mathbf{X}_0)(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{y}$$

Standard linear regression – project onto the column space of the design matrix B

$$B^T = \begin{pmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_N \end{pmatrix}$$

$$y = (Y(X_1), \dots, Y(X_N))^T$$

Local linear regression

$$\hat{Y}(\mathbf{X}_0) = (1, \mathbf{X}_0)(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{y}$$

$$B^T = \begin{pmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_N \end{pmatrix}$$

Standard linear regression – project onto the column space of the design matrix B

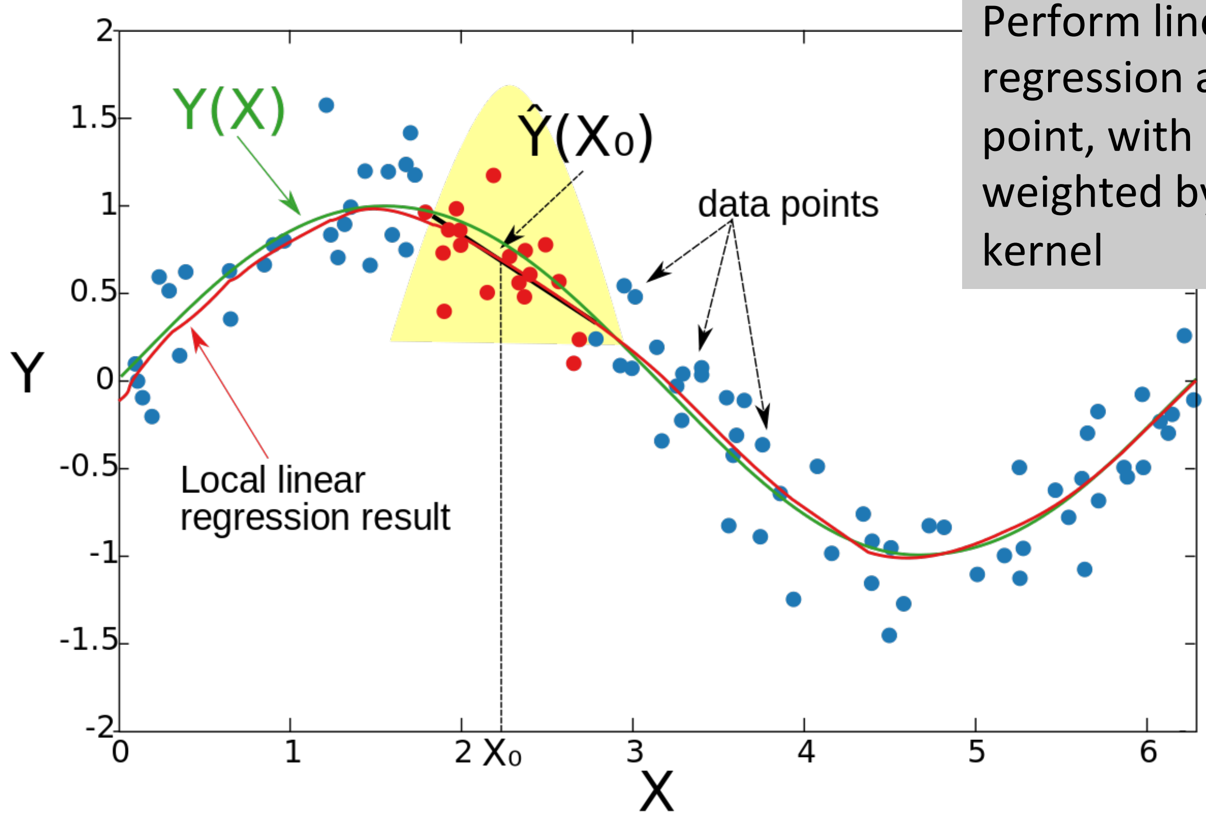
$$y = (Y(X_1), \dots, Y(X_N))^T$$

$$\hat{Y}(\mathbf{X}_0) = (1, \mathbf{X}_0)(\mathbf{B}^T W(\mathbf{X}_0)\mathbf{B})^{-1}\mathbf{B}^T W(\mathbf{X}_0)\mathbf{y}$$

Local linear regression - data weighted by the local kernel

$$W(X_0) = \mathrm{diag}\left(K_{h_\lambda}(X_0, X_i)\right)_{N \times N}$$

Local linear regression



Perform linear regression at each point, with data weighted by the local kernel

Source: Wikipedia

david.r.thompson@jpl.nasa.gov

Kernel density estimation

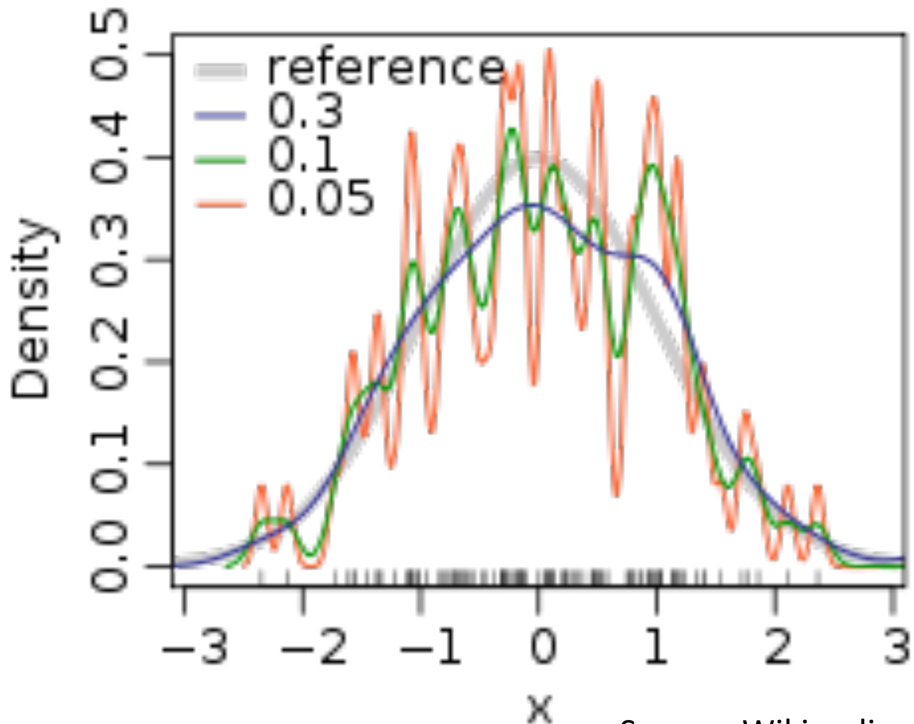
density

kernel

$$\hat{f}(x_0) = \frac{1}{nz} \sum_i k(x_0, x_i)$$

Number of datapoints

Normalizing factor (so each kernel has volume of 1)



Source: Wikipedia

Summary

Local nonparametric methods are powerful, flexible tools for varied pattern recognition problems

Classification: K-nearest neighbor

Regression: kernel smoothing or local linear regression

Density estimation: kernel density estimation

Set regularization parameters (# nearest neighbors or kernel width) using cross-validation

Kernel density

david.r.thompson@jpl.nasa.gov