Nonparametric, Nearest Neighbors

CSci 5525: Machine Learning

Instructor: Nicholas Johnson

October 8, 2020

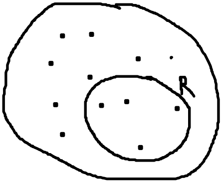
Announcements

- HW2 is posted (due Oct 15)
- Exam 1 will be posted on Oct 20 (due 48 hours later)
- Course feedback form

Parametric vs Nonparametric

- Parametric models
  - Set of parameters with fixed size
  - Independent of the number of training points
  - Example: Linear regression, Linear classification
  - Great for small amounts of data
  - May miss subtleties if huge amounts of data is available
- Nonparametric models
  - *Not* "Do not use any parameters"
  - Number of 'parameters' grow with data
  - Example: k-nearest neighbor classifier
  - Good for large datasets with 'subtleties'
  - Care is needed to avoid under- or over-fitting

Density estimation



- True density $p(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^p$, we draw N samples
- Density at a local region R: $P = \int_R p(\mathbf{x}) d\mathbf{x}$
  - Number of samples in the region $K \approx NP$
  - For small region of volume V, density is uniform, $P \approx p(\mathbf{x})V$
- Estimate of density:

$$p(\mathbf{x}) \approx \frac{K}{NV}$$

- Two approaches to estimation
  - Kernel density estimate: Keep V fixed, estimate K from data
  - Nearest neighbor estimate: Keep K fixed, estimate V from data

Kernel Density Estimation

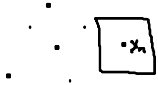
$$k(u) = \begin{cases} 1 & \text{if } |u_i| \leq 1/2, \\ 0 & \text{o/w} \end{cases}$$

- Consider a kernel $k : \mathbb{R}^p \mapsto \mathbb{R}_+$
  - Need $k(\mathbf{u}) \geq 0, \int k(\mathbf{u}) d\mathbf{u} = 1$
  - Bounded support, e.g., $k(\mathbf{u}) = 1, |\mathbf{u}_i| \leq 1/2, 0$ otherwise
- Total number of points inside cube of size h is

$$p(x) \approx \frac{K}{NV}$$

$$K = \sum_{n=1}^{N} k\left(\frac{\mathbf{x} - \mathbf{x}_n}{h}\right)$$

V:



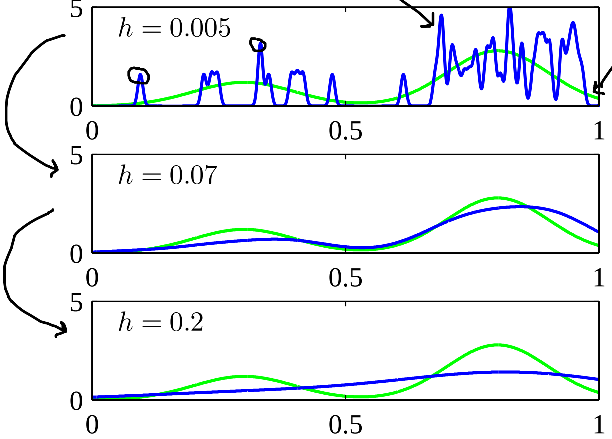
- Since volume is h^p, the density estimate is

$$p(\mathbf{x}) \approx \frac{1}{N} \sum_{n=1}^{N} \frac{1}{h^p} k\left(\frac{\mathbf{x} - \mathbf{x}_n}{h}\right)$$

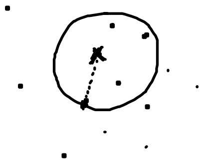
Kernel Density Estimation

$\hat{p}(x) = \frac{K}{NV} \approx p(x)$

$p(x)$

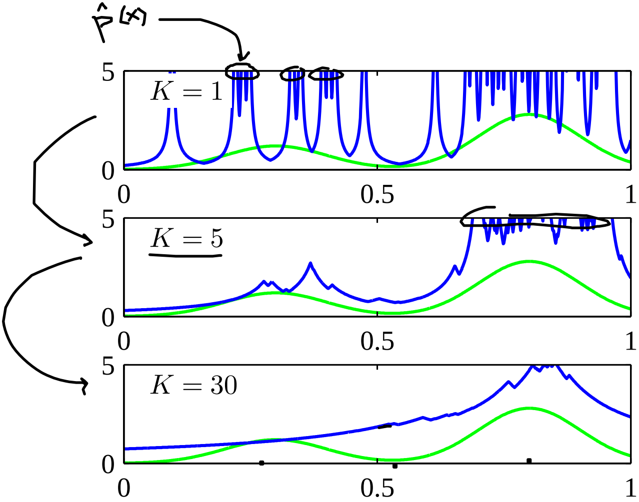


Nearest Neighbor Density Estimation



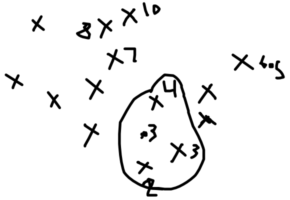
- Main idea: Fix K, estimate V from data
- Estimating density $p(\mathbf{x})$ around $\mathbf{x}$
  - Fix K, number of neighbors
  - Grow a sphere around $\mathbf{x}$, till it includes K points
  - Volume of the sphere is V
- Radius of sphere is distance to the K^{th} nearest neighbor
- Estimate density as: $p(\mathbf{x}) = \frac{K}{NV}$ 4

Nearest Neighbor Density Estimation



Nearest Neighbors: Classification, Regression

$k=3$



- Find $NN_K(\mathbf{x})$, the k nearest neighbors of $\mathbf{x}$
- Classification: Majority class in $NN_K(\mathbf{x})$
- Regression: Solve linear regression using $NN_K(\mathbf{x})$
- Choice of distance metric
  - Suitable L_p norm
  - z-scored metrics, Mahalanobis distance
  - Metrics defined by nonlinear 'kernels'
  - Application domain dependent metrics

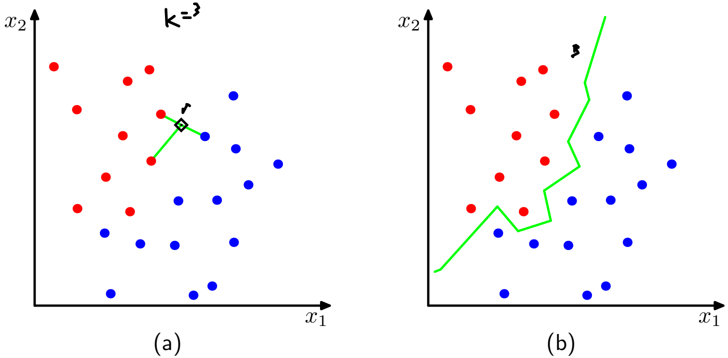
$$\hat{y}(x) = \frac{2+3+4}{3} = \frac{9}{3} = 3$$

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

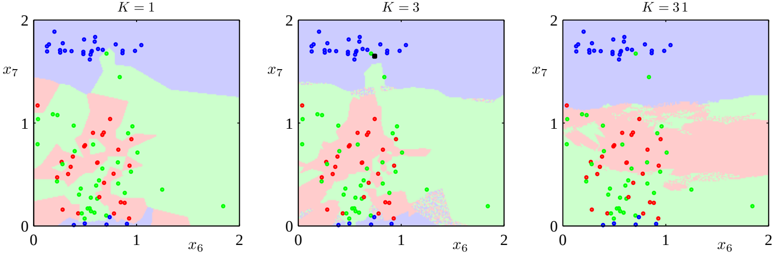
$$\underset{w}{\operatorname{argmax}} \frac{1}{n} \sum_{i \in n_K} (y_i - w^T x_i)^2$$

n_K = KNN to point x

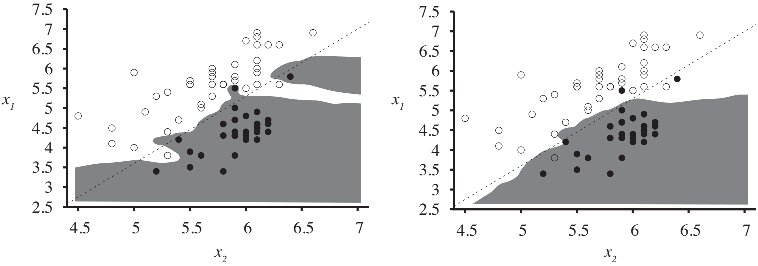
Nearest Neighbor Classification



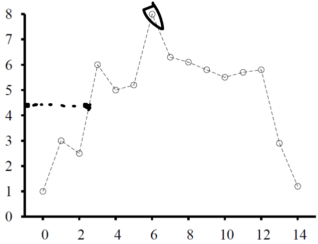
Nearest Neighbor Classification

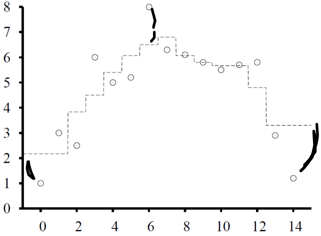


Over-fitting, Under-fitting



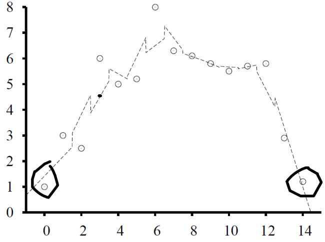
Nonparametric Regression



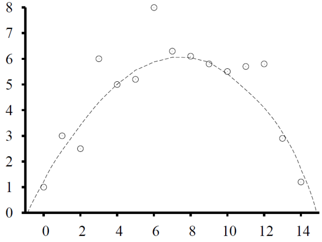


- ‘Connect-the-dots’ regression can be ‘spiky’
- K-nearest neighbor regression: mean of K points

Nonparametric Regression (Contd.)

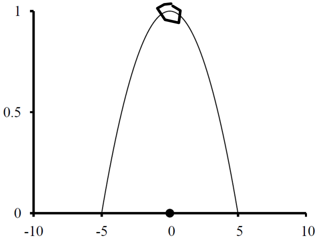


(j)



(k)

- k-nearest neighbor linear regression: best line
- Locally weighted regression: Weighting is done using a kernel



- Kernel $k(\cdot)$ is symmetric around 0 and maximum at 0
- The area under the kernel should remain bounded
- Bandwidth of the kernel: Underfitting vs Overfitting
- Locally weighted regression:

$$\mathbf{w}^*(\mathbf{x}) = \operatorname{argmin}_{\mathbf{w}} \sum_{n=1}^{N} k\left(\frac{\mathbf{x}-\mathbf{x}_n}{h}\right)(y_n - \mathbf{w}^\top \mathbf{x}_n)^2$$

$$\Rightarrow \qquad y(\mathbf{x}) = (\mathbf{w}^*(\mathbf{x}))^\top \mathbf{x}$$