Principles of Artificial Intelligence(305450)

Lecture 12:

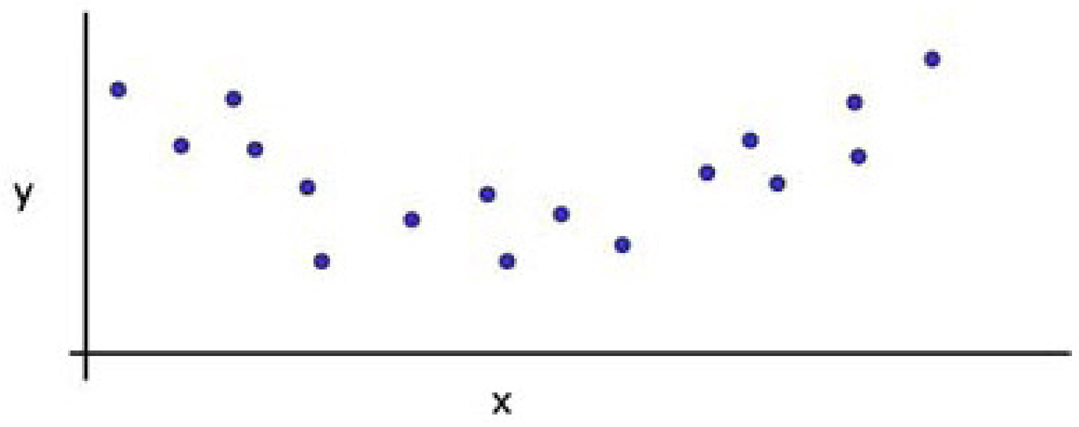
Machine Learning V

Regression

- Output is a continuous numeric value
  - Locally-weighted averaging
  - Regression trees

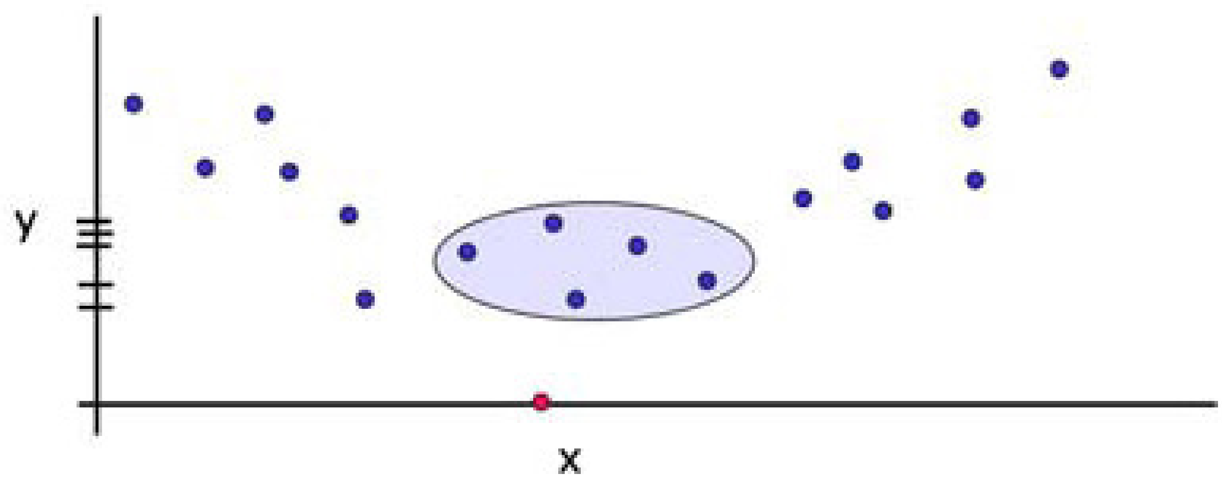
Local Averaging

- Remember all your data



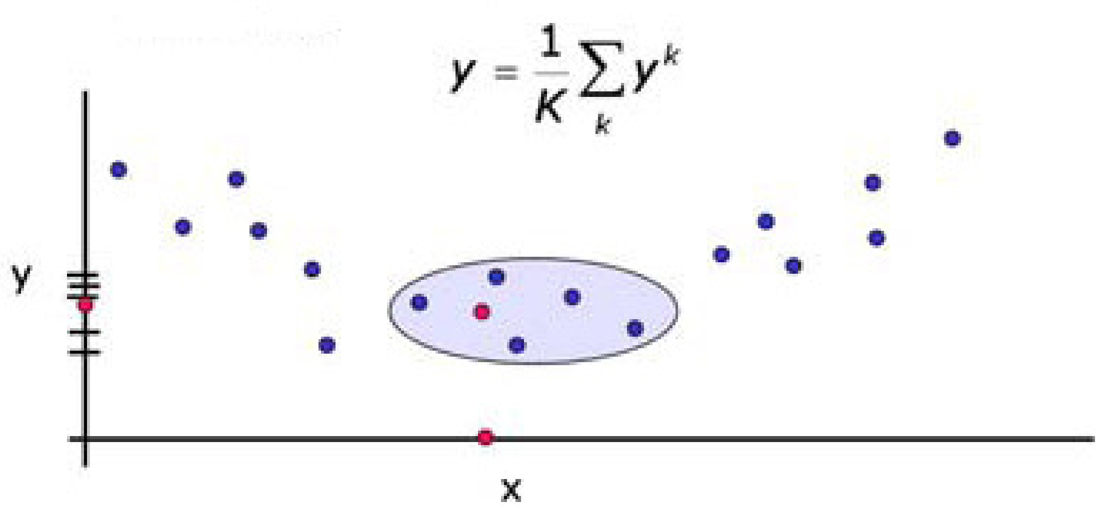
Local Averaging

- Remember all your data
- When someone asks a question,
  - Find the K nearest old data points



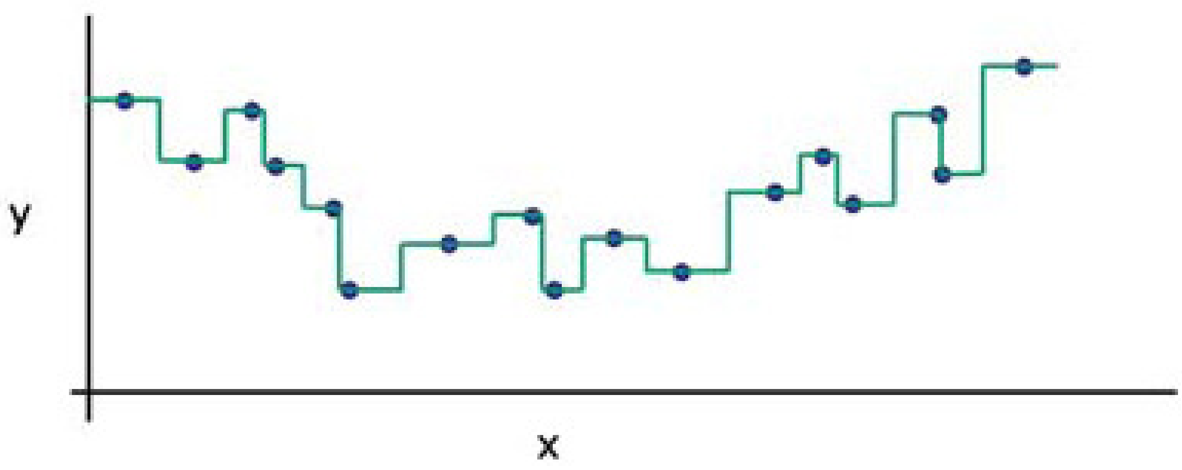
Local Averaging

- Remember all your data
- When someone asks a question,
  - Find the K nearest old data points
  - Return the average of the answers associated with them



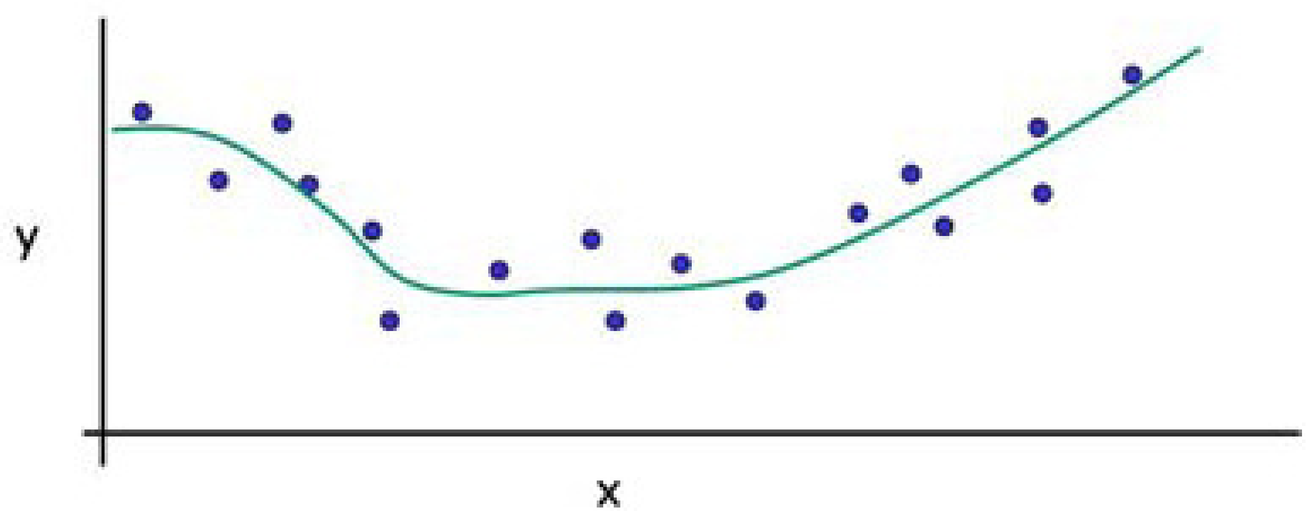
K=1

- Tracks data very closely
- Prone to overfitting



Bigger K

- Smoothes out variations in data
- May introduce too much bias



Locally Weighted Averaging

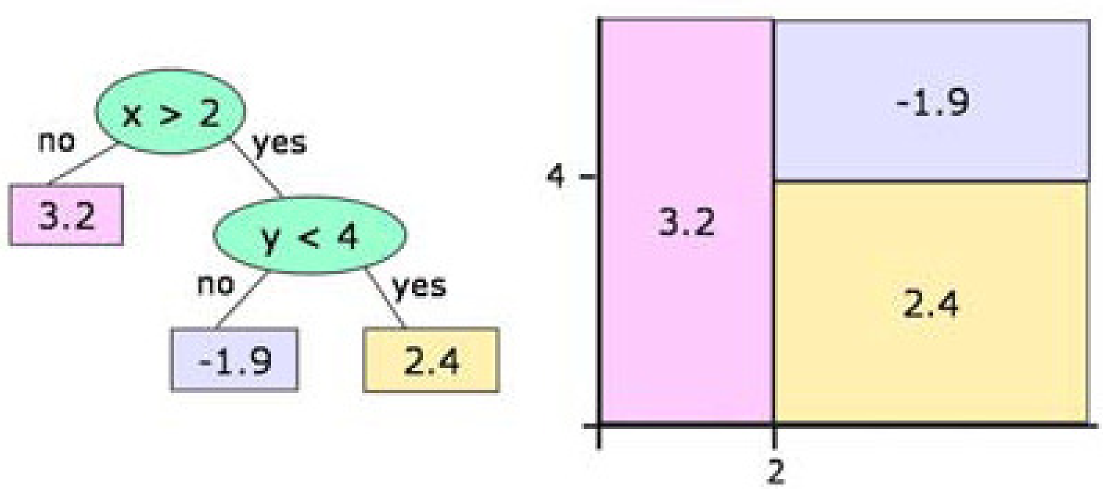
- Find all points within distance λ from target point
- Average the outputs, weighted according to how far away they are from the target point
- Given a target x, with k ranging over neighbors,

$$y = \frac{\sum_k K(x, x^k) y^k}{\sum_k K(x, x^k)}$$

weighting "kernel"

Regression Trees

- Like decision trees, but with real-valued constant outputs at the leaves



Leaf Values

- Assign a leaf node the average of the y values of the data points that fall there
- We'd like to have groups of points in a leaf that have similar y values (because then the average is a good representative)

Variance

- Measure of how much a set of numbers is spread out
- Mean of m values, z_1 through z_m:

$$\mu = \frac{1}{m}\sum_{k=1}^{m} z_k$$

- Variance: average squared difference between z's and the mean

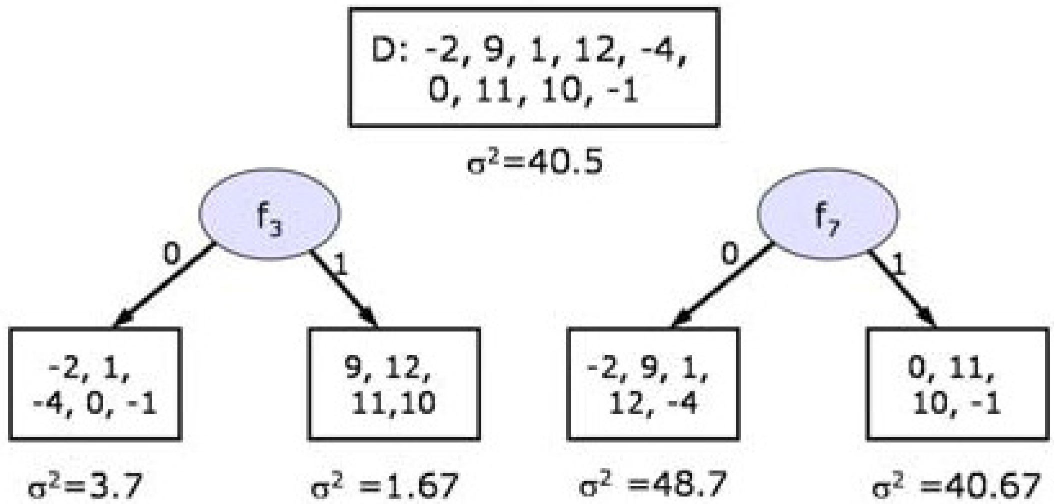
$$\sigma^2 = \frac{1}{m-1}\sum_{k=1}^{m}(z_k - \mu)^2$$

Let’s split

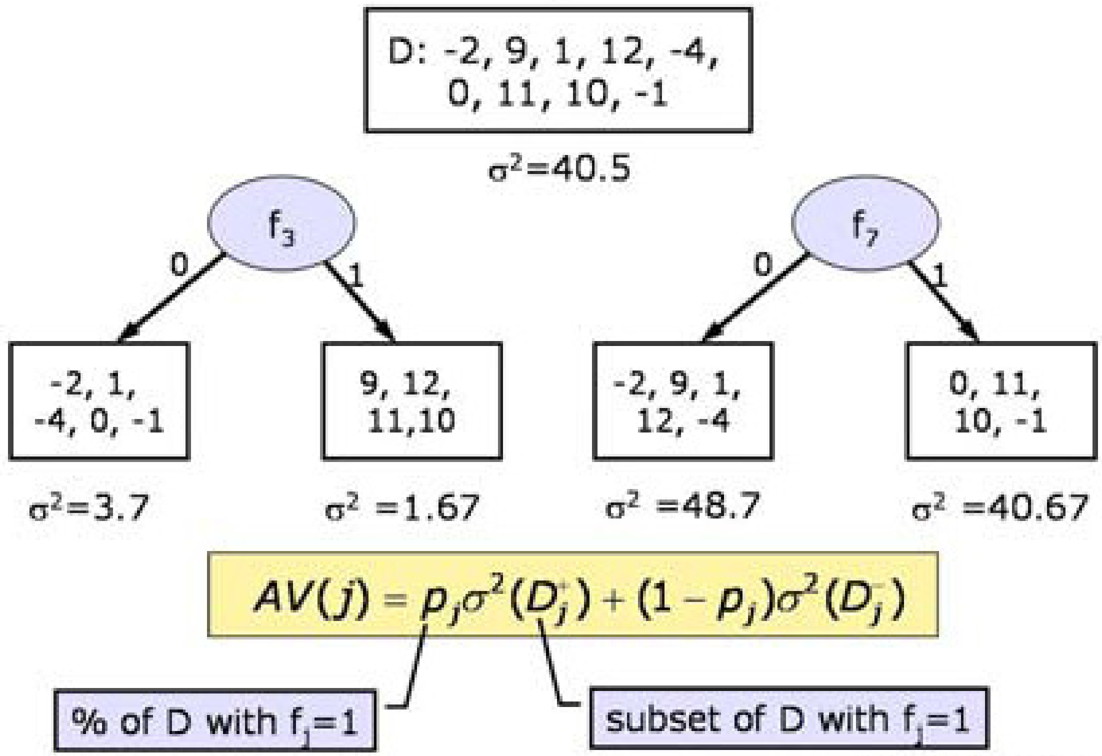
D: -2, 9, 1, 12, -4,
0, 11, 10, -1

σ^2=40.5

Let’s split

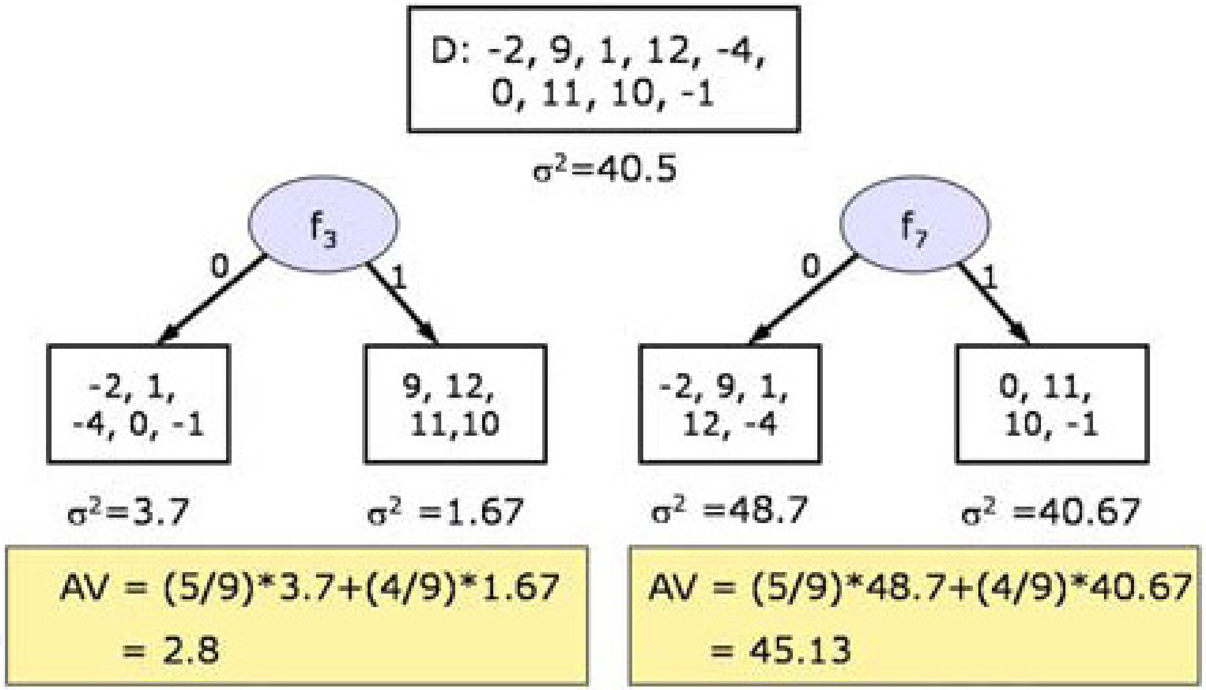

D: -2, 9, 1, 12, -4, 0, 11, 10, -1
σ²=40.5
f3
0
1
-2, 1, -4, 0, -1
σ²=3.7
9, 12, 11,10
σ² =1.67
f7
0
1
-2, 9, 1, 12, -4
σ² =48.7
0, 11, 10, -1
σ² =40.67

Let’s split



D: -2, 9, 1, 12, -4, 0, 11, 10, -1
σ²=40.5
f3
0
1
-2, 1, -4, 0, -1
9, 12, 11,10
f7
0
1
-2, 9, 1, 12, -4
0, 11, 10, -1
σ²=3.7
σ² =1.67
σ² =48.7
σ² =40.67
$AV(j) = p_j\sigma^2(D_j^+) + (1-p_j)\sigma^2(D_j^-)$
% of D with fj=1
subset of D with fj=1

Let's split



D: -2, 9, 1, 12, -4, 0, 11, 10, -1
σ²=40.5
f3
0
1
-2, 1, -4, 0, -1
9, 12, 11,10
σ²=3.7
σ² =1.67
f7
0
1
-2, 9, 1, 12, -4
0, 11, 10, -1
σ² =48.7
σ² =40.67
AV = (5/9)*3.7+(4/9)*1.67
= 2.8
AV = (5/9)*48.7+(4/9)*40.67
= 45.13

Stopping

- Stop when variance at a leaf is small enough
- Or when you have fewer than min-leaf elements at a leaf
- Set y at a leaf to be the mean of the y values of the elements