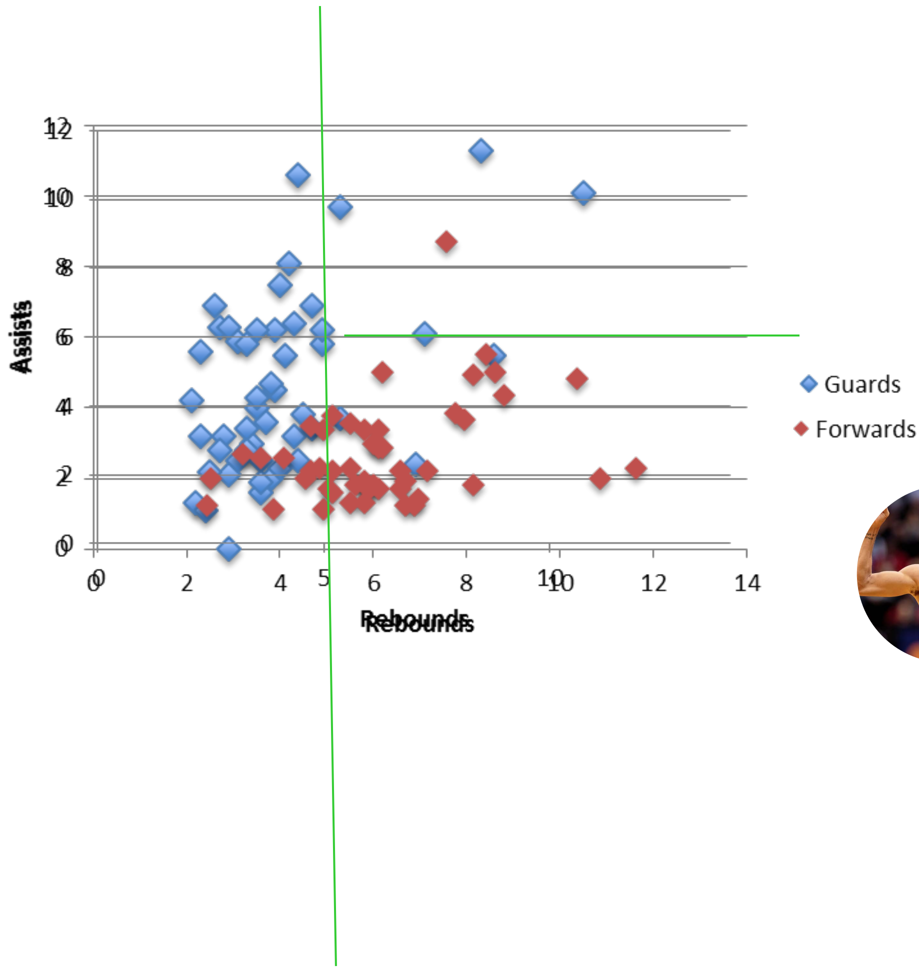
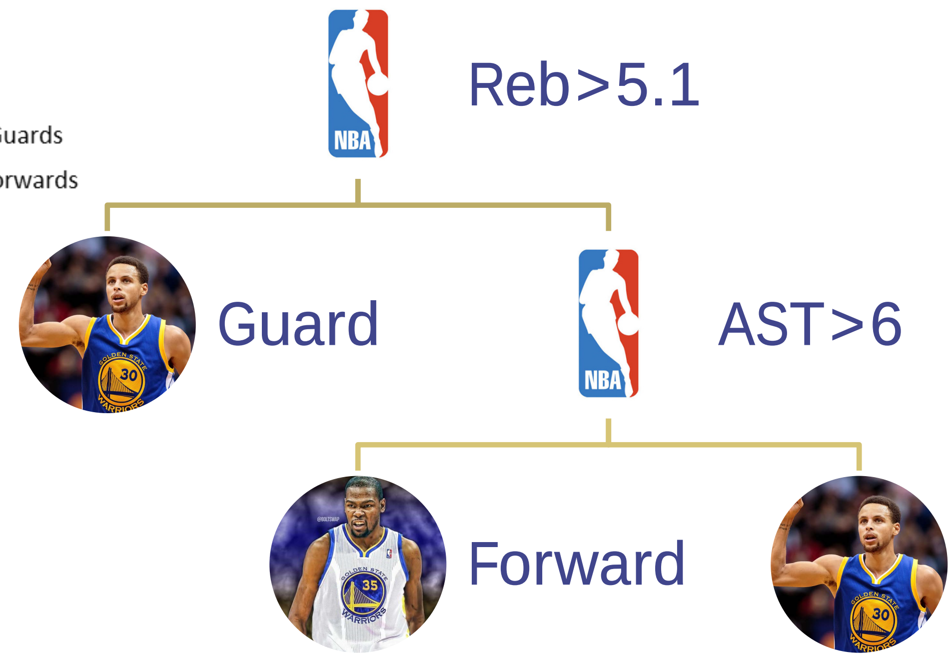
Machine Learning

4771

Instructor: Itsik Pe’er

Reminder: Decision Trees





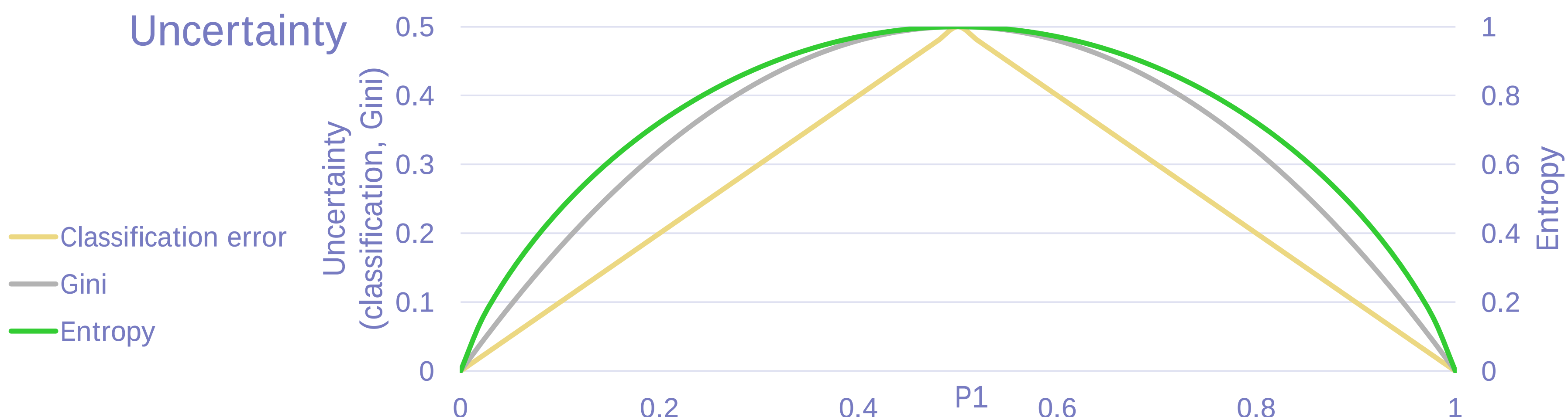
Objective: Certainty

Choose leaf and split to minimize
a measure of uncertainty $X, \Pr(X = i) = p_i$

Objective: Certainty

Choose leaf and split to minimize
a measure of uncertainty of $X, \Pr(X=i)=p_i$

- Classification error: $1-\max p_i$
- Gini index: $\Pr(x_1 \neq x_2) = 1-\sum p_i^2$
- Entropy: $E[I(X)] = \sum p_i \log_2 \frac{1}{p_i}$



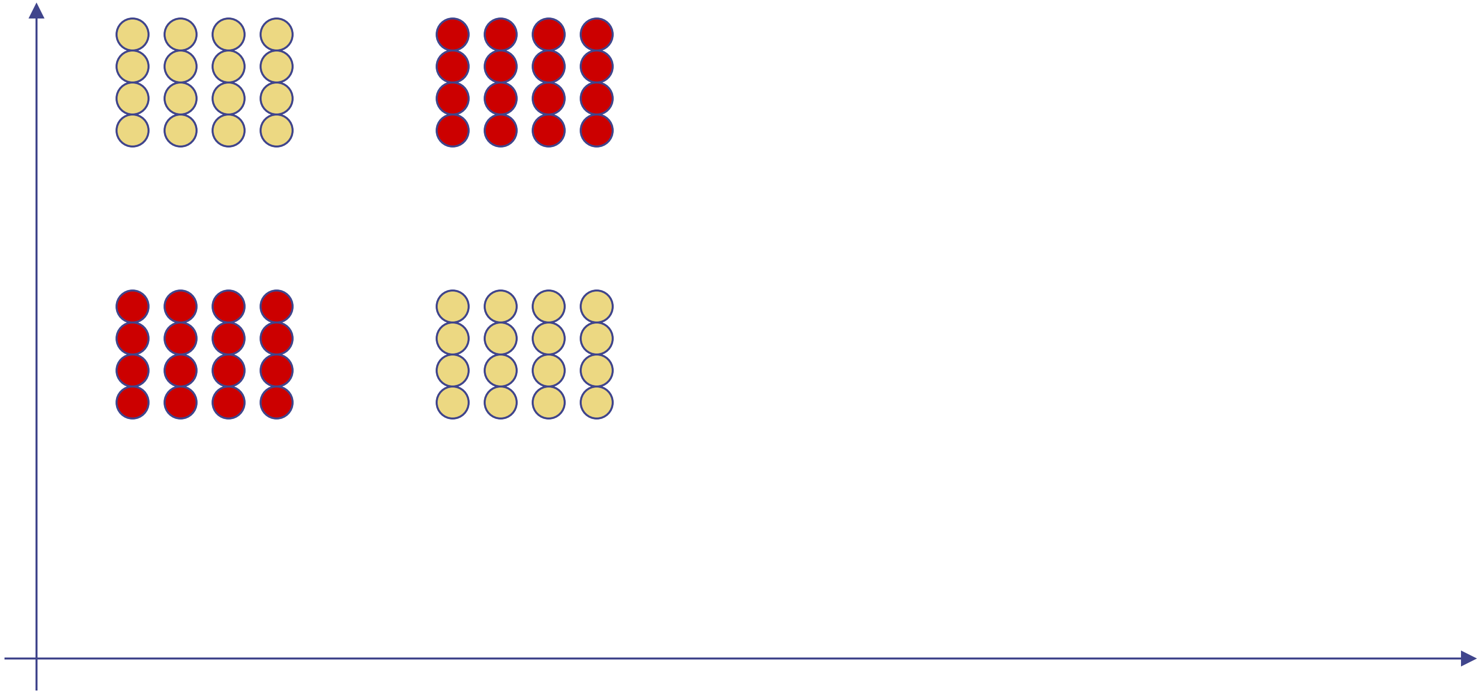
Objective: Certainty

Choose leaf and split to minimize
a measure of uncertainty of $X, \Pr(X = i) = p_i$

- Classification error: $1 - \max p_i$
- Gini index: $\Pr(x_1 \neq x_2) = 1 - \sum p_i^2$
- Entropy: $E[I(X)] = \sum p_i \log \frac{1}{p_i}$

Stopping Criteria

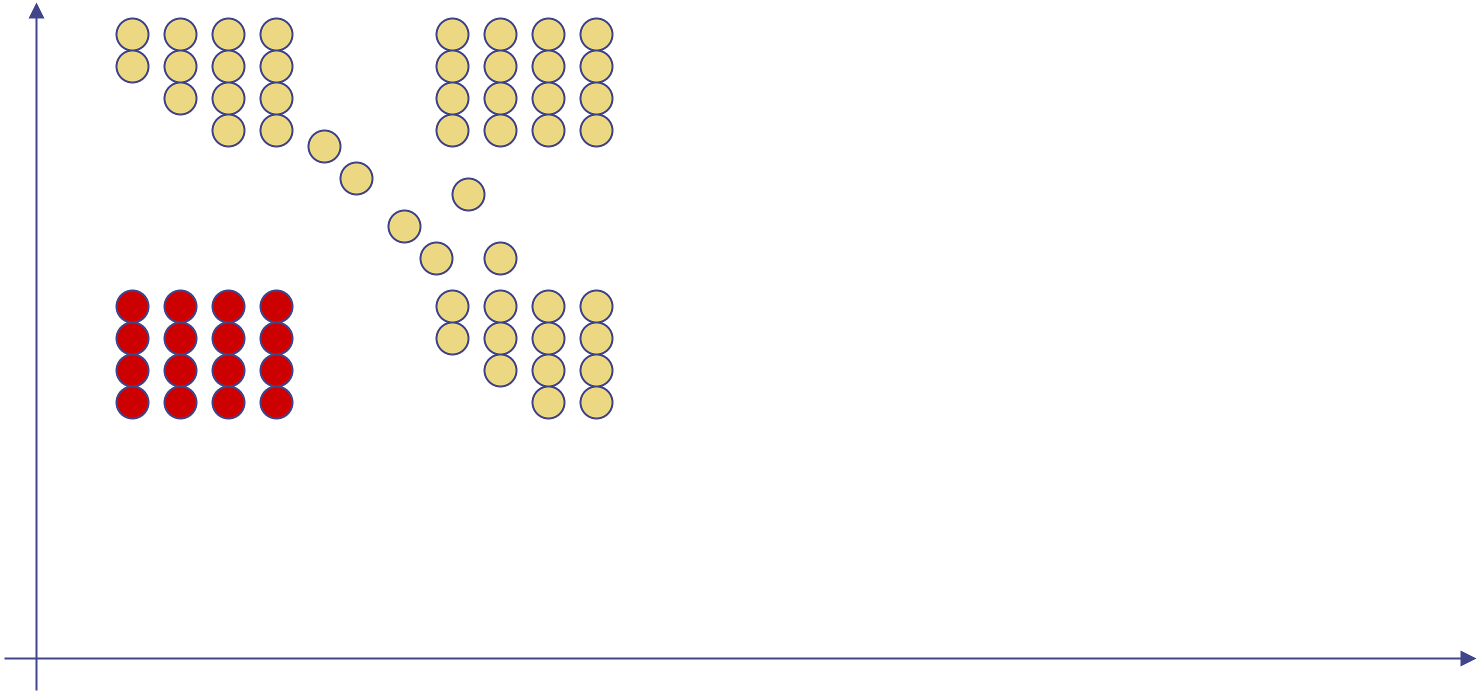
Example:
No split reduces uncertainty



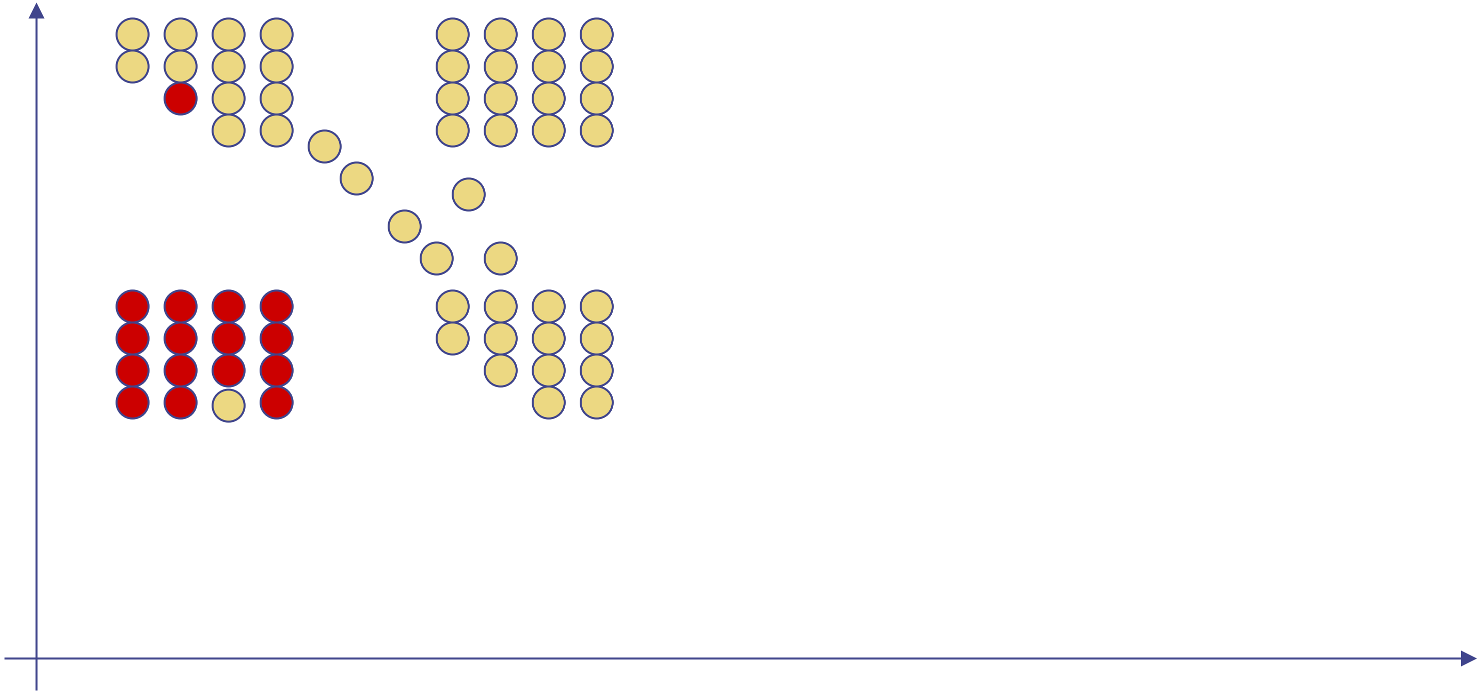
Stopping Criteria

- No improvement?
- Certain tree size
- When leaves are pure

Example:
Clear split if clean data



Example:
Overfit if noisy data



Stopping Criteria

- No improvement?
- Certain tree size
- When leaves are pure
  - Overfitting. Requires pruning
  - Address by validation set.
  - Prune the pure-training tree

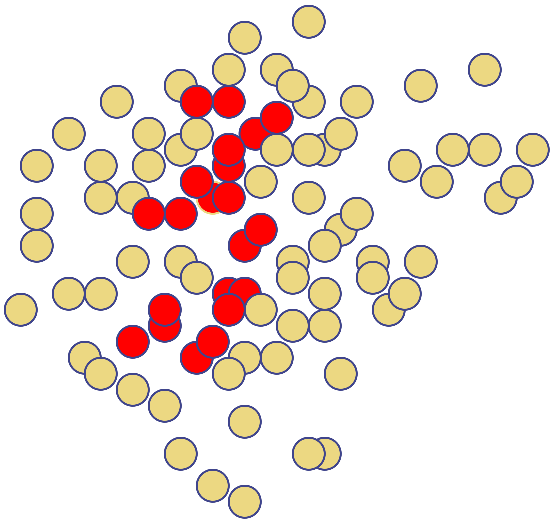
Summary – Decision Trees

- Decision trees grow greedily

- Effective when “dominant” dimensions

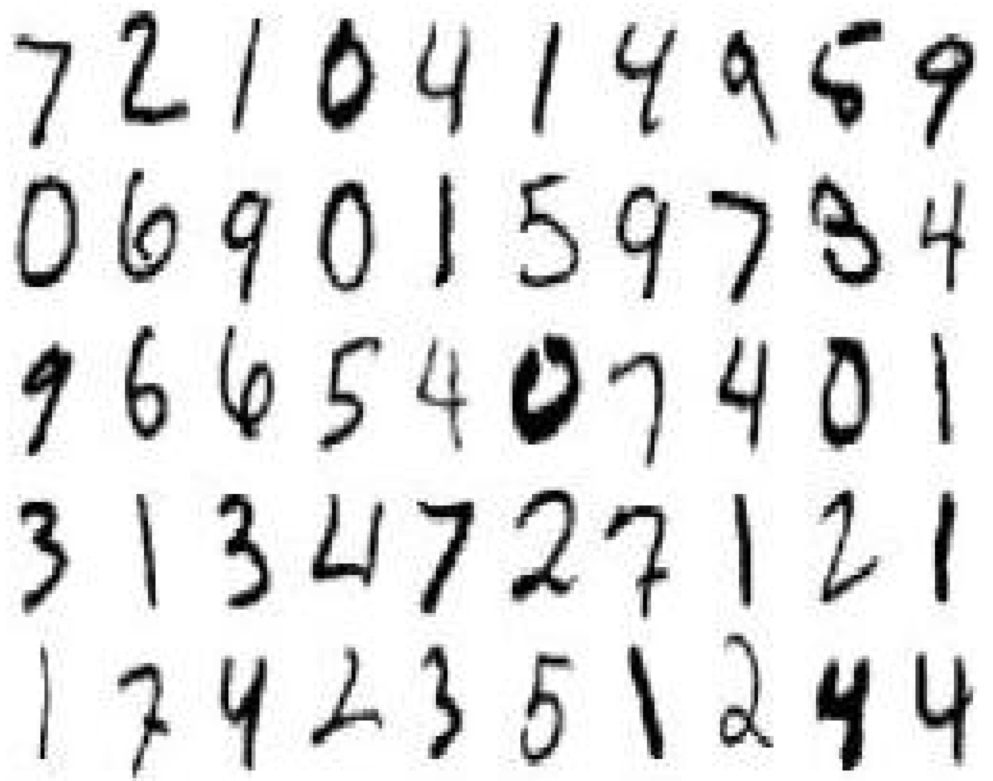
Summary – Decision Trees

- Decision trees grow greedily
- Effective when "dominant" dimensions
- Is there an RBF analog?
  Based on proximity, not axis



Nearest Neighbor

- Idea: Small $\|x - \tilde{x}\|$ implies $y = \tilde{y}$

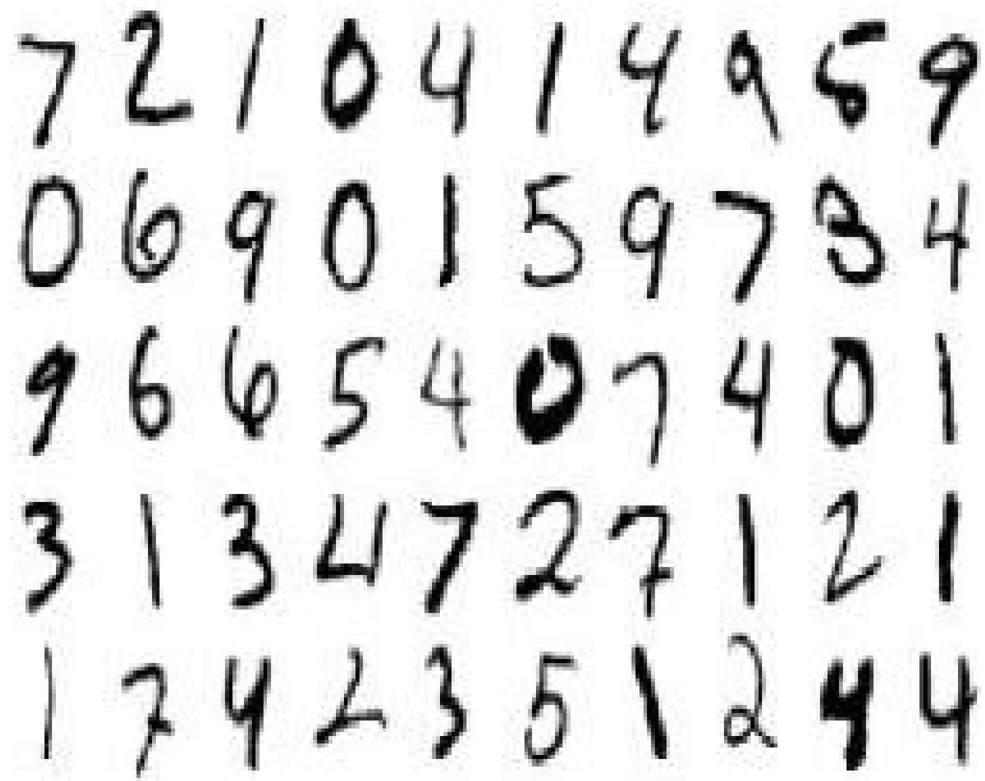


- Example: OCR

```
0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1
0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 0
0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0
0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0
0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0
0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0
0 0 1 1 1 1 1 1 0 0 0 0 0 0 0 0
0 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0
1 1 1 1 1 0 0 0 1 1 1 0 0 0 0 0
0 1 1 1 0 0 0 0 0 0 1 1 0 0 0 0
1 1 1 0 0 0 0 0 0 1 1 0 0 0 0 0
1 1 0 0 0 0 0 0 0 1 1 0 0 0 0 0
1 1 0 0 0 0 0 0 1 1 0 0 0 0 0 0
1 1 0 0 0 0 0 1 1 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0
0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0
```

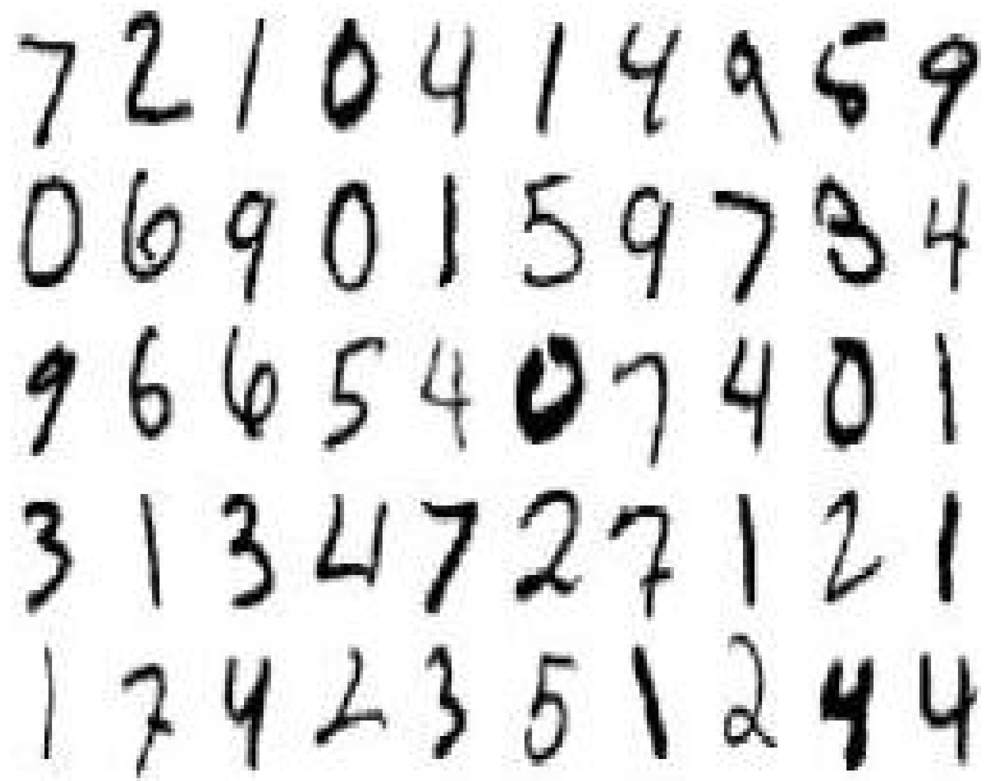
Nearest Neighbor

- Idea: Small $||x - \tilde{x}||$ implies $y = \tilde{y}$
- Example: OCR
- $x \in \boldsymbol{R}^{28\times 28}$
  $y \in \{0, \dots, 9\}$



Nearest Neighbor

- Idea: Small $||x - \tilde{x}||$ implies $y = \tilde{y}$



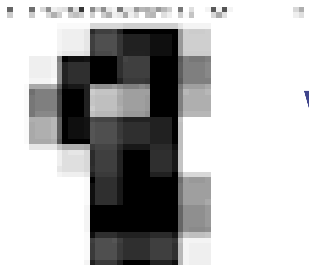
- Example: OCR
- Training $\{(x_i, y_i)\}$
- Classify(x):

$j \leftarrow argmin_i ||x - x_i||$

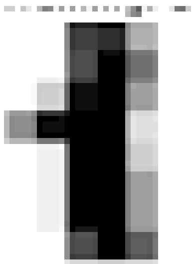
return y_j

Nearest Neighbor(s)

- Problem: sensitivity to class outliers
  this 9



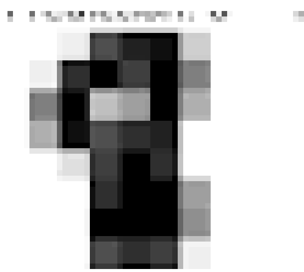
will cause this



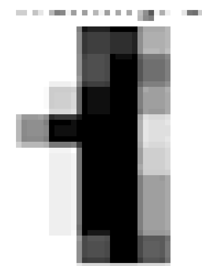
to be labeled 9

Nearest Neighbor(s)

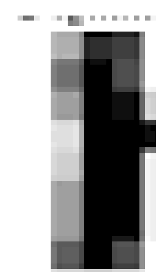
- Problem: sensitivity to class outliers this 9

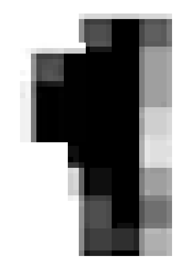


will cause this



to be labeled 9

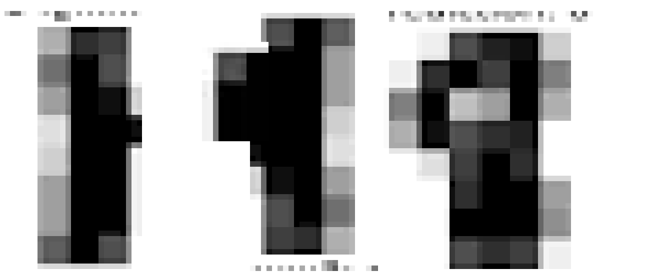




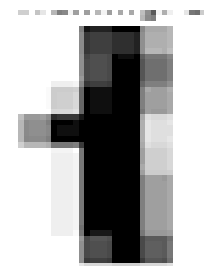
Nearest Neighbor(s)

- Problem: sensitivity to class outliers

  this 9



will cause this



to be labeled 9

- Solution: rely on k>1 neighbors

  Idea: 1NN of 0.1 error worse than Maj(9 NN) each 0.3 error

k Nearest Neighbors

- Classify(x):

  while $|J| < k$

  $$J \leftarrow J \cup \{argmin_{i \notin J} \|x - x_i\|\}$$

  return $Plurality(\{y_j | j \in J\})$

k Nearest Neighbors

- Classify(x):

  while $|J| < k$

  $$J \leftarrow J \cup \{argmin_{i \notin J} ||x - x_i||\}$$

  return $Plurality(\{y_j | j \in J\})$

- What is k ?
- Distance?

k Nearest Neighbors

- Classify(x):

  while $|J| < k$

  $$J \leftarrow J \cup \{argmin_{i \notin J} ||x - x_i||\}$$

  return $Plurality(\{y_j | j \in J\})$

- What is k ? Determine by validation set
- Distance? Domain dependent

Distance Functions

- Euclidean
- Images, audio:
  How much warping need to turn x to $\tilde{x}$?
- Strings: edit distance

Known issues: Bad Features

Known issues: Bad Features

Need distance to prioritize “good” features

Complexity

- How complex is FindClosest(X,x) ?

Complexity

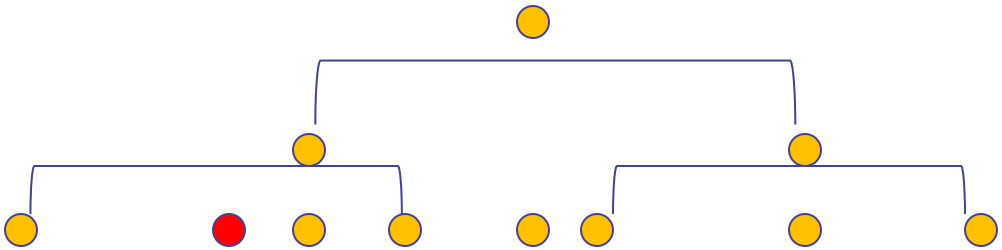
- How complex is FindClosest(X,x) ?
- Naively, O(N) distance evaluations
- 1D:

  Preprocessing: O(logN)

Complexity

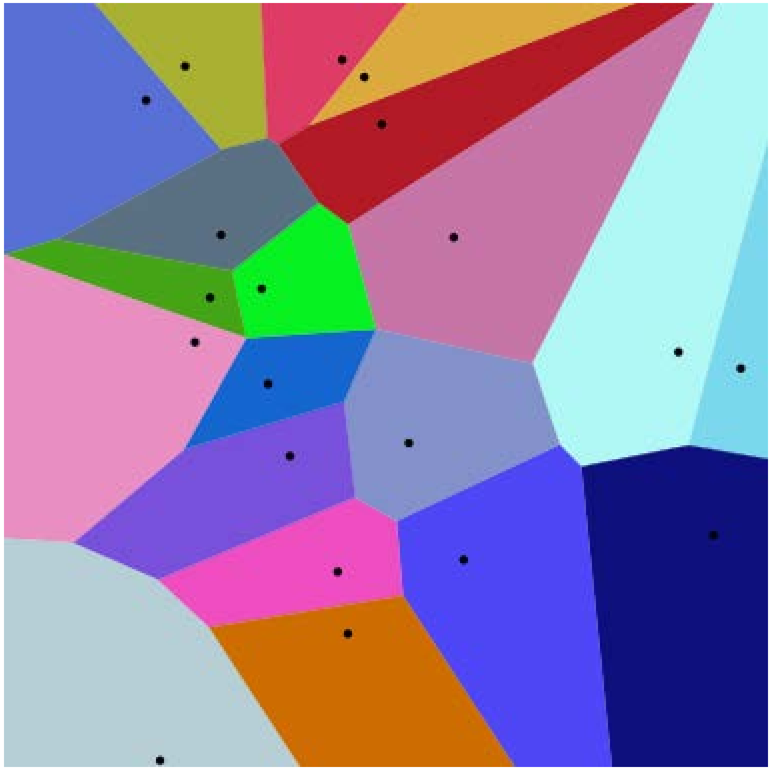
- How complex is FindClosest(X,x) ?
- Naively, O(N) distance evaluations
- 1D:

  Preprocessing: O(logN)



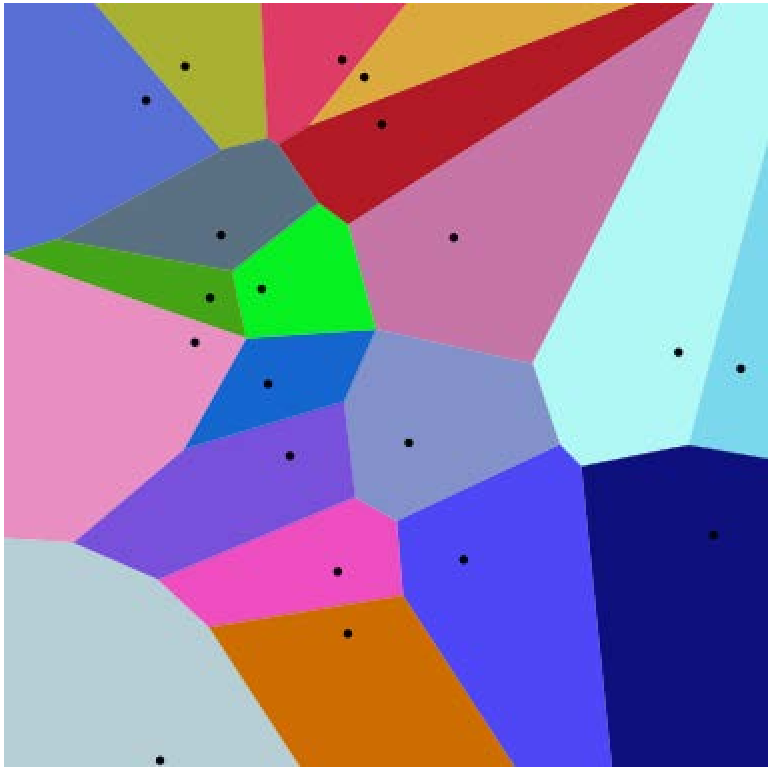
Preprocessing Higher D

- Voronoi Diagrams

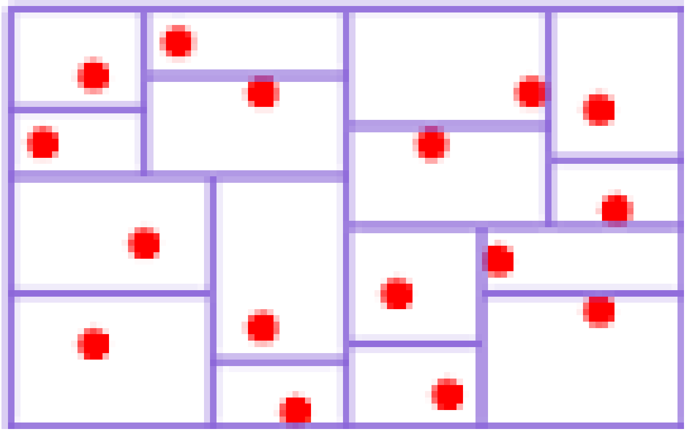


Preprocessing Higher D

- Voronoi Diagrams

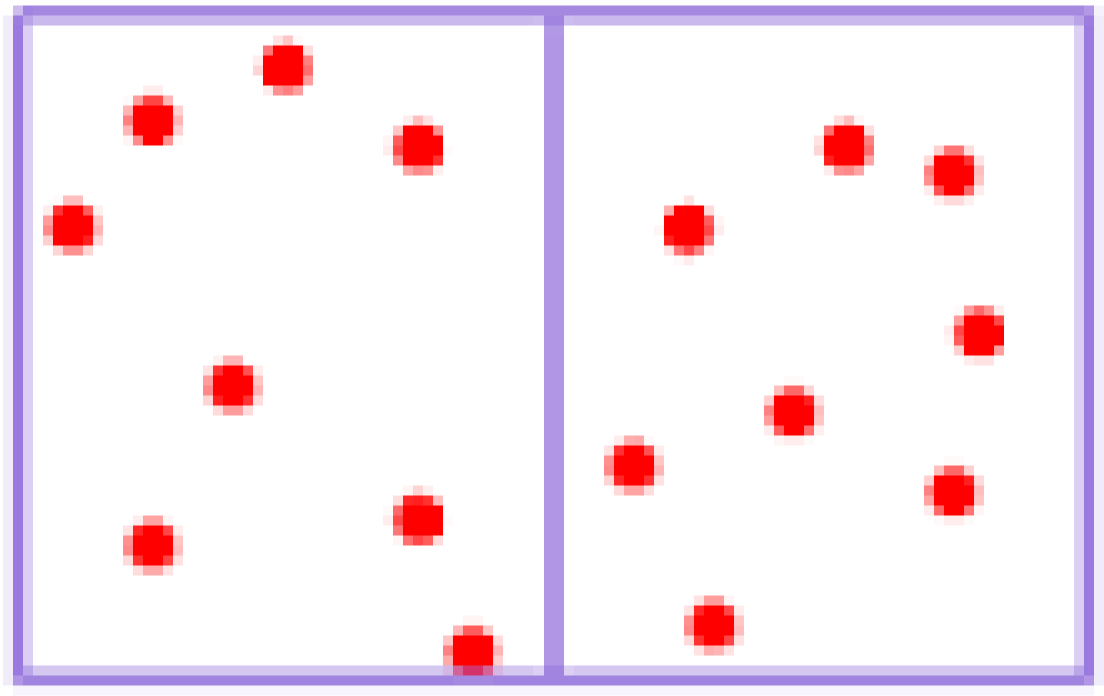


- k-d tree



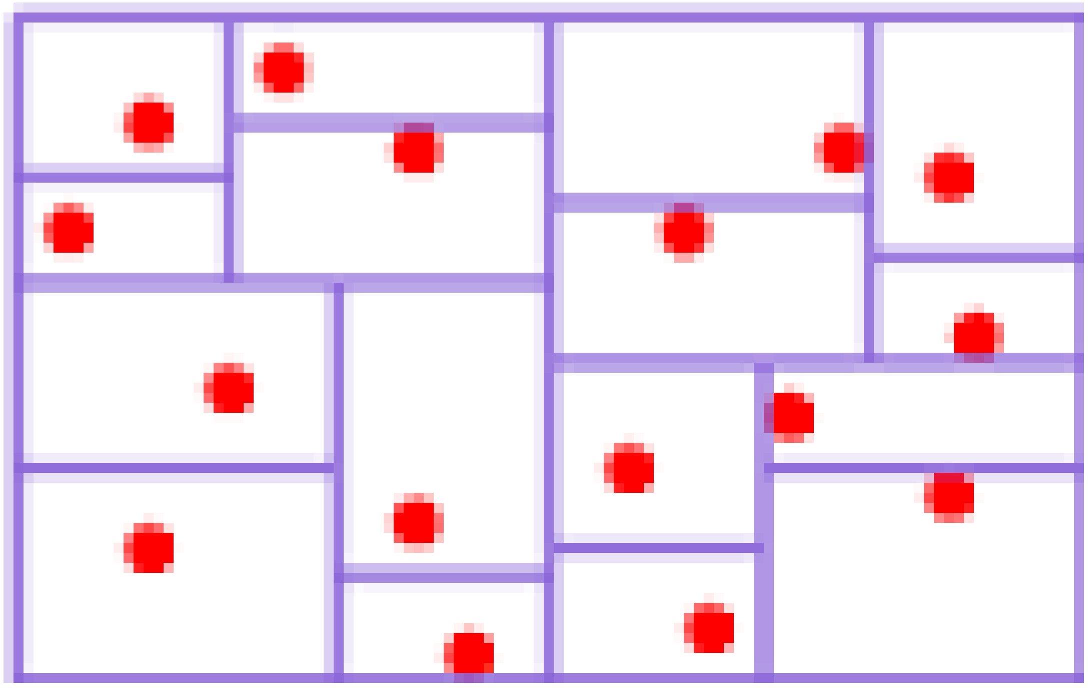
k-d Tree Construction

- Repeat
  - Pick dimension
  - Split by median



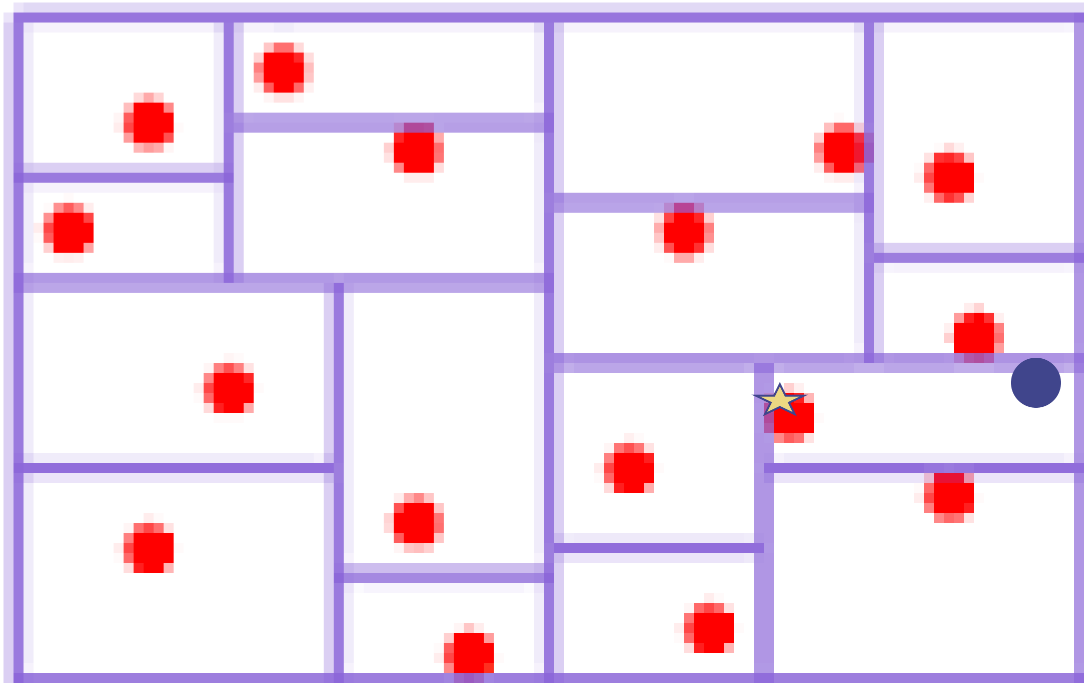
k-d Tree Search

- Repeat
  - Compare to median
  - Choose side
  - Recurse



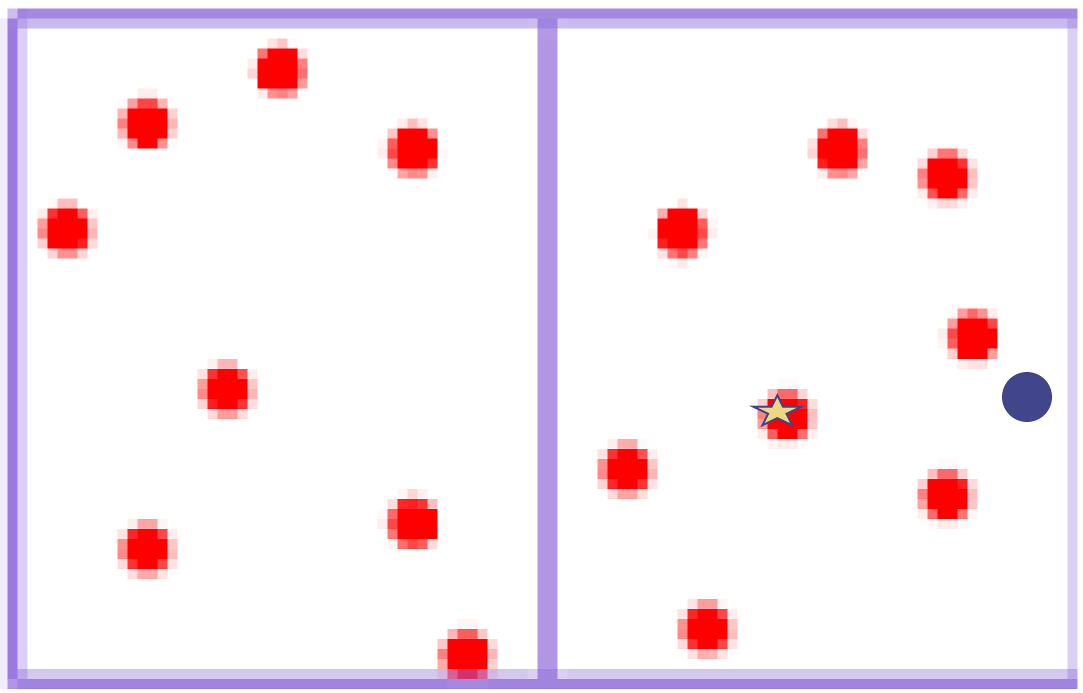
k-d Tree Search (training x)

- Repeat
  - Compare to median
  - Choose side
  - Recurse
- Does not find closest!



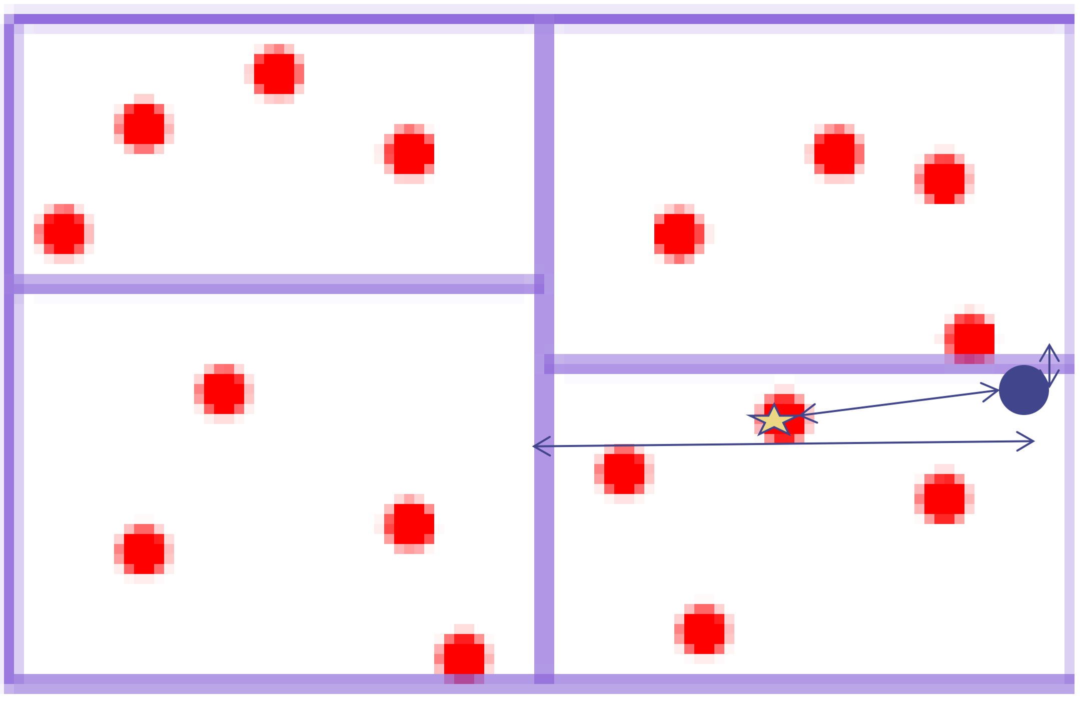
k-d Tree Search (any x)

- Repeat
  - Compare to median
  - Choose side
  - Recurse(side)
  - If needed
    - Recurse(other)



k-d Tree Search (any x)

- Repeat
  - Compare to median
  - Choose side
  - Recurse(side)
  - If needed
    - Recurse(other)
- No O(log n) guarantee



Summary+notes

- kNN : distance-based effective classification
  - Non-parametric
- Data structures for preprocessing
- Consistency guarantee