

CS 360: Machine Learning

Sara Mathieson, Sorelle Friedler

Spring 2024



HVERFORD
COLLEGE

Admin

- Sorelle office hours **Thursday: 4-5pm in H110**
- **Lab 5** due Thursday (check in during lab today)
 - Goal: finish non-regularization case by *end* of lab
- Make sure you have **3 handouts!**
 - Study guide, practice problems, feedback form
- Reminder about **extra credit for handout solutions** (see Piazza)
 - Deadline: 24 hours before the midterm

Midterm Notes

- In class on **Tuesday** (week from today)
- You may bring a one page (front and back) “study sheet”
- Handwritten, created by you

Outline for Feb 27

- Correction for softmax
- Review K-nearest neighbors and KD trees
- Review gradient descent and logistic regression
- Thursday: decision trees, evaluation metrics, ML pipeline

Outline for Feb 27

- Correction for softmax
- Review K-nearest neighbors and KD trees
- Review gradient descent and logistic regression
- Thursday: decision trees, evaluation metrics, ML pipeline

Softmax

$$s_k = \vec{w}^{(k)} \cdot \vec{x}$$

↗ positive
→ 0
↘ negative

Softmax function

$$p(y=k | \vec{x}) = \frac{e^{s_k}}{\sum_{i=1}^K e^{s_i}}$$

$$K=3, P=1$$

$$\begin{aligned} \vec{w}^{(1)} &= [-2, 3]^T \\ \vec{w}^{(2)} &= [1, 0]^T \\ \vec{w}^{(3)} &= [-5, -1]^T \end{aligned}$$

result of SGD for multi-class logistic regression

$$X=2 \} \text{ test data}$$

$$\vec{x} = [1, 2]^T$$

$$\left. \begin{aligned} S_1 &= -2 + 3 \cdot 2 = 4 \\ S_2 &= 1 + 0 \cdot 2 = 1 \\ S_3 &= -5 - 1 \cdot 2 = -7 \end{aligned} \right\}$$

$$P(Y=1) = \frac{e^4}{e^4 + e^1 + e^{-7}}$$

$$\approx 0.95$$

$$\hat{Y} = 1$$

Outline for Feb 27

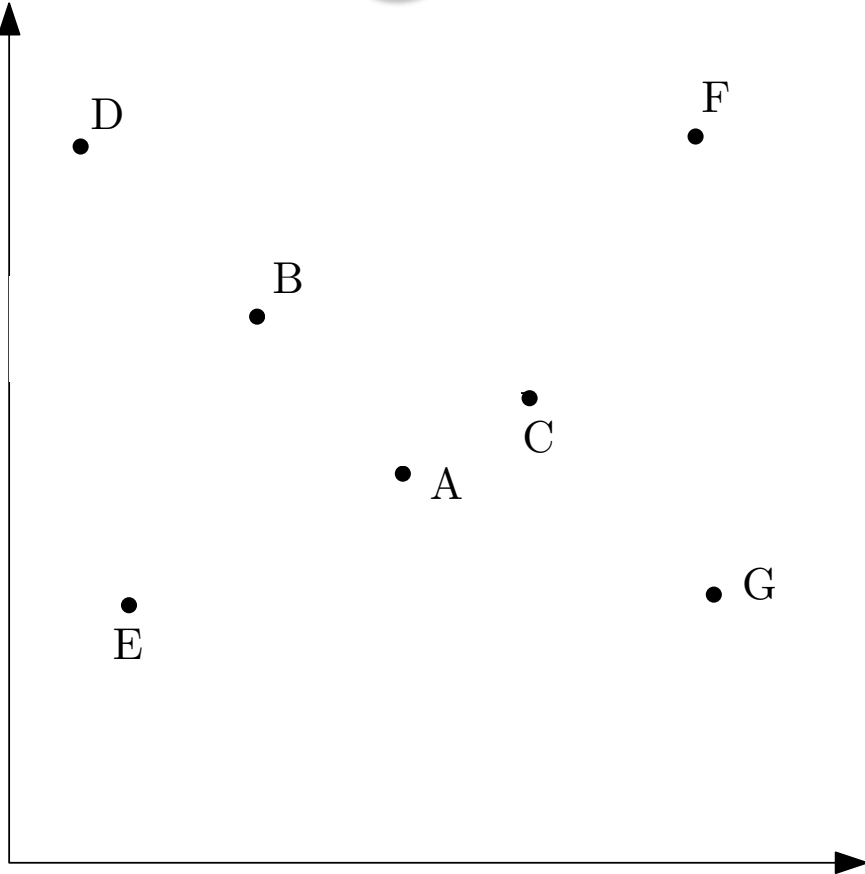
- Correction for softmax
- **Review K-nearest neighbors and KD trees**
- Review gradient descent and logistic regression
- Thursday: decision trees, evaluation metrics, ML pipeline

Handout 4, example 1

Sort by dim 0 (x-axis here) and choose median

More points *below* the median (i.e. fill in left child first)

D E B **A** C F G

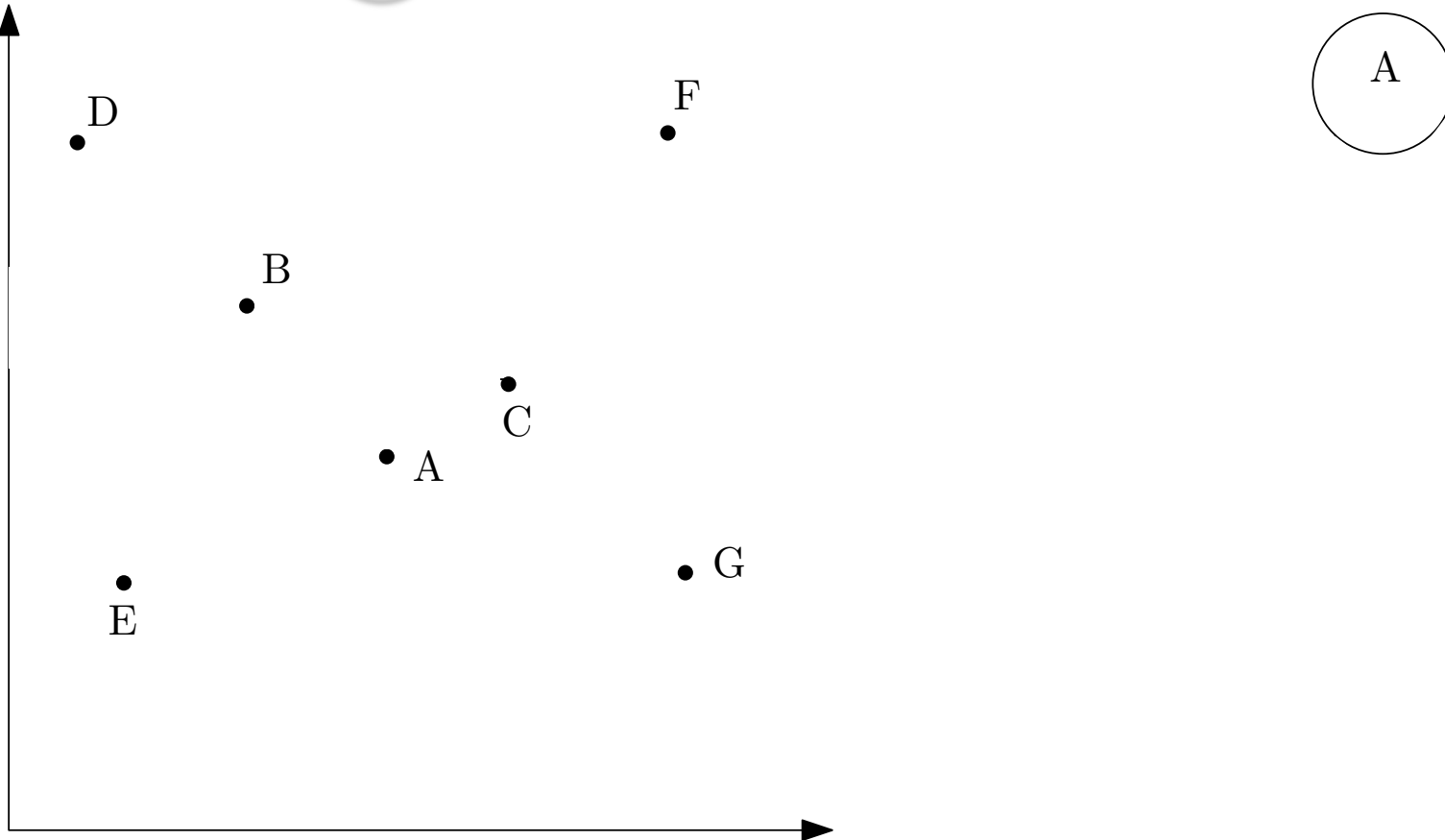


Handout 4, example 1

Sort by dim 0 (x-axis here) and choose median

More points *below* the median (i.e. fill in left child first)

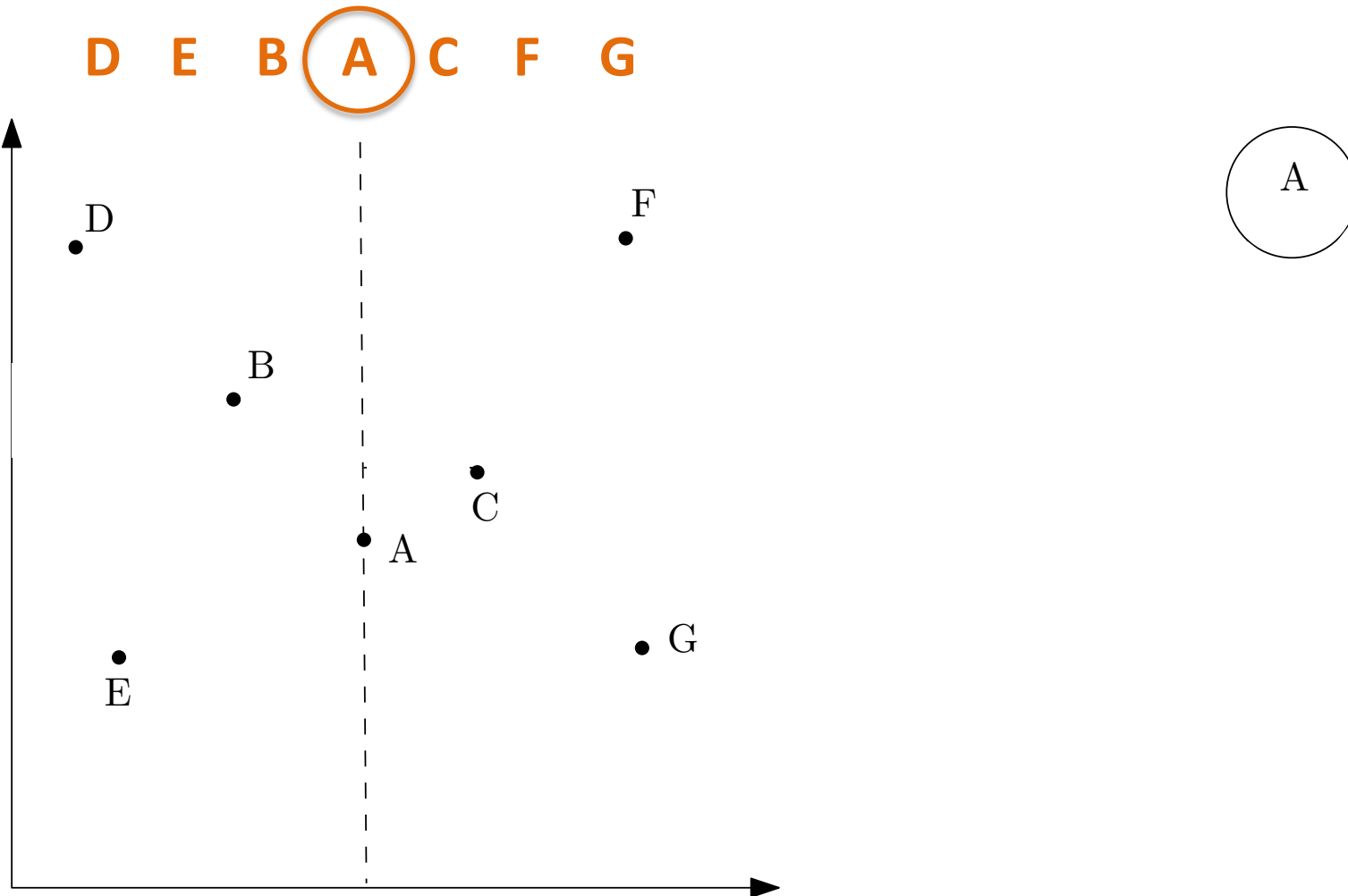
D E B **A** C F G



Handout 4, example 1

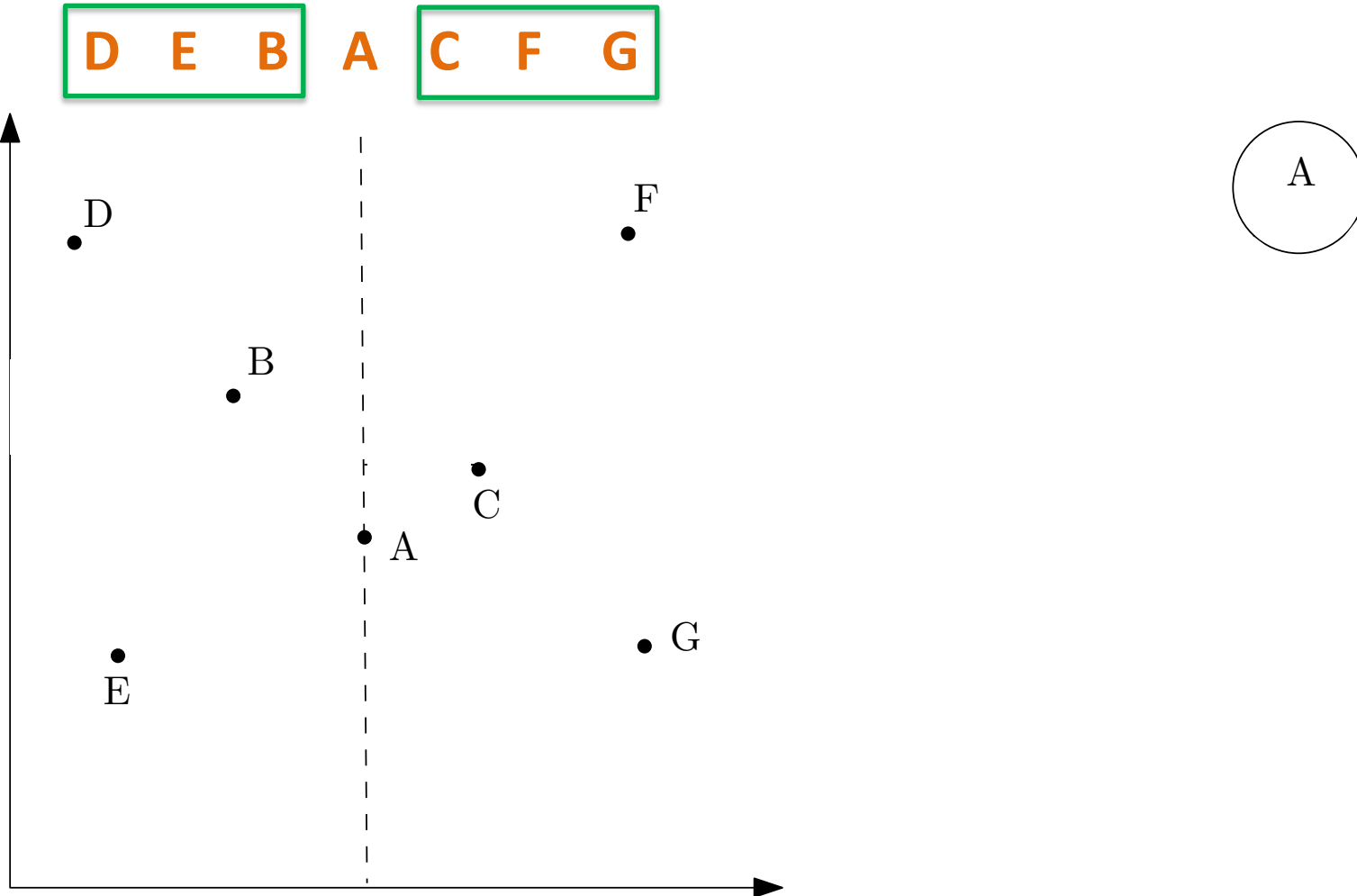
Sort by dim 0 (x-axis here) and choose median

More points *below* the median (i.e. fill in left child first)



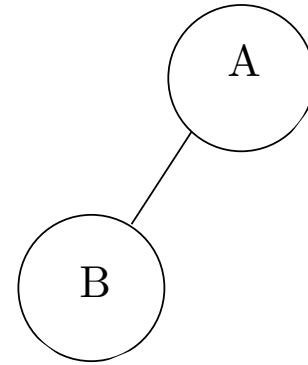
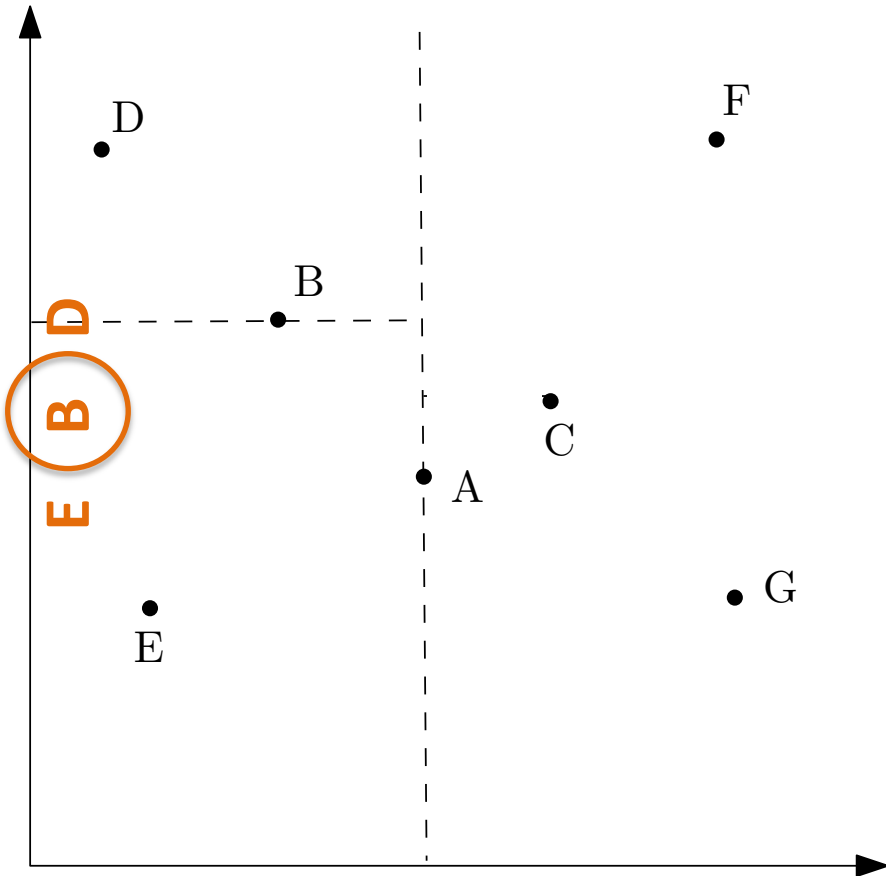
Handout 4, example 1

Recurse on the points below and above the median



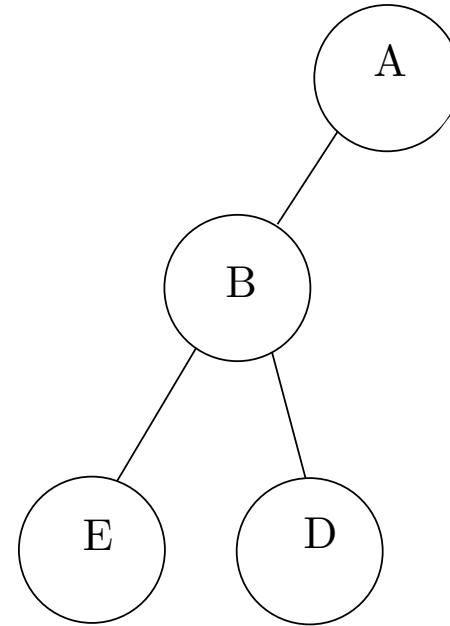
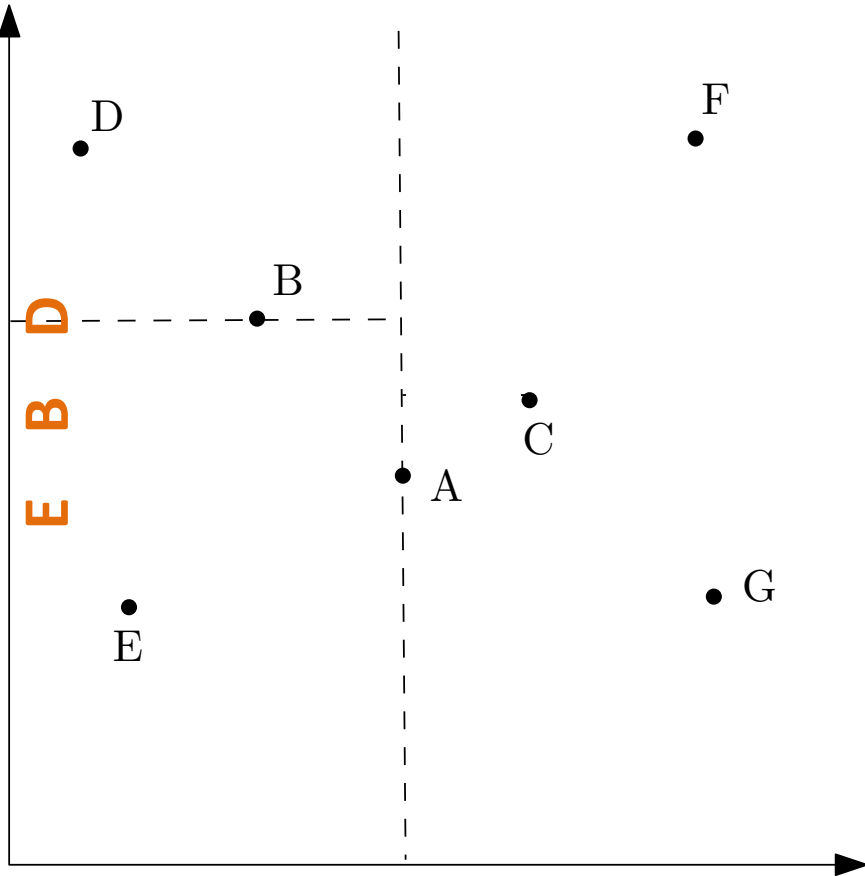
Handout 4, example 1

Recurse on the points below and above the median



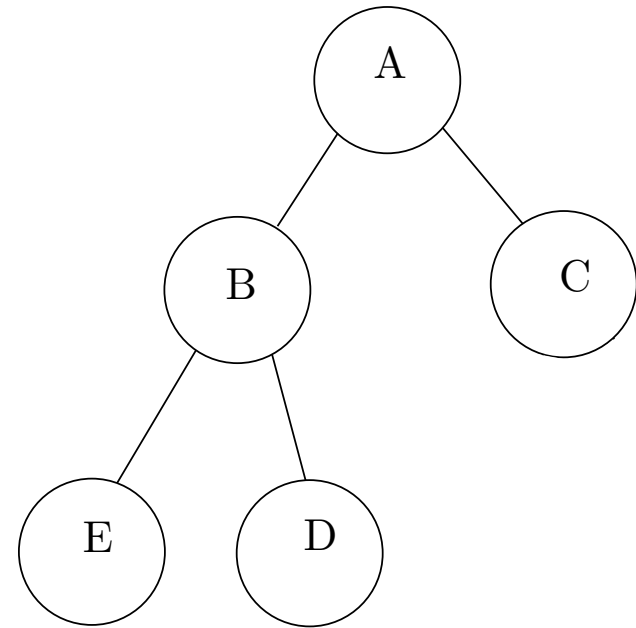
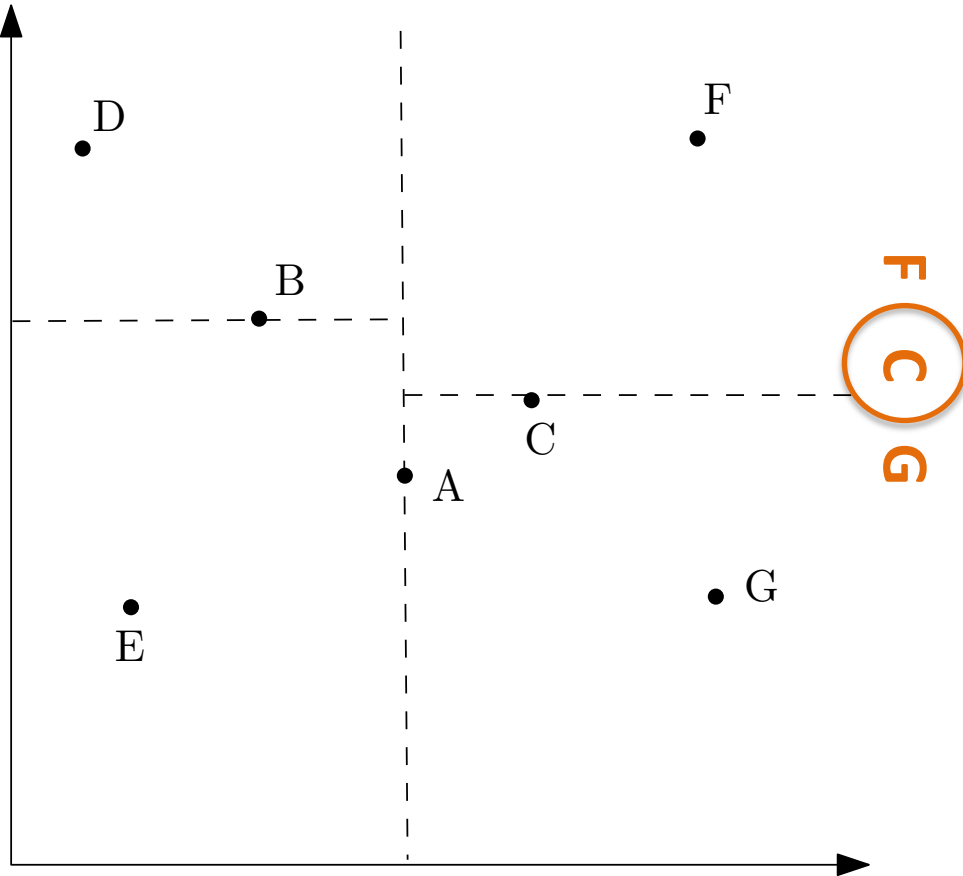
Handout 4, example 1

Terminate when we don't have any more points



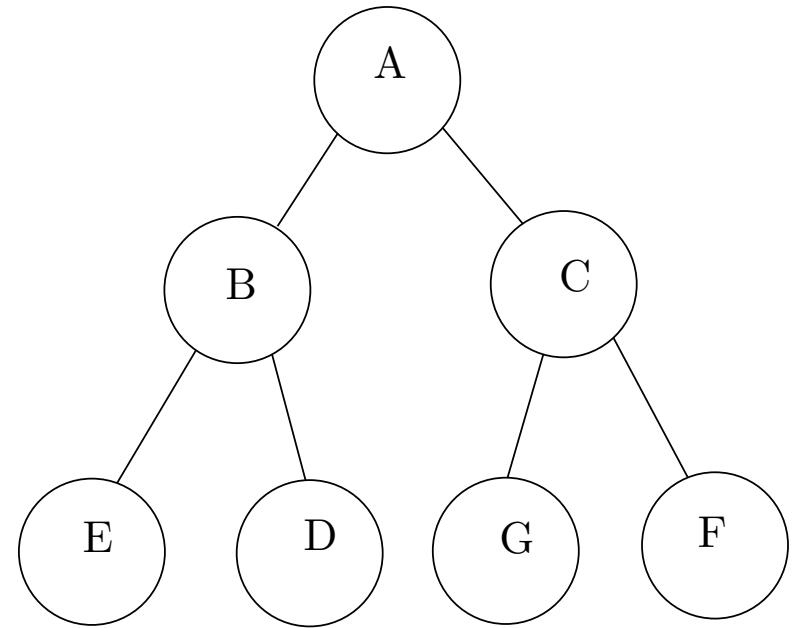
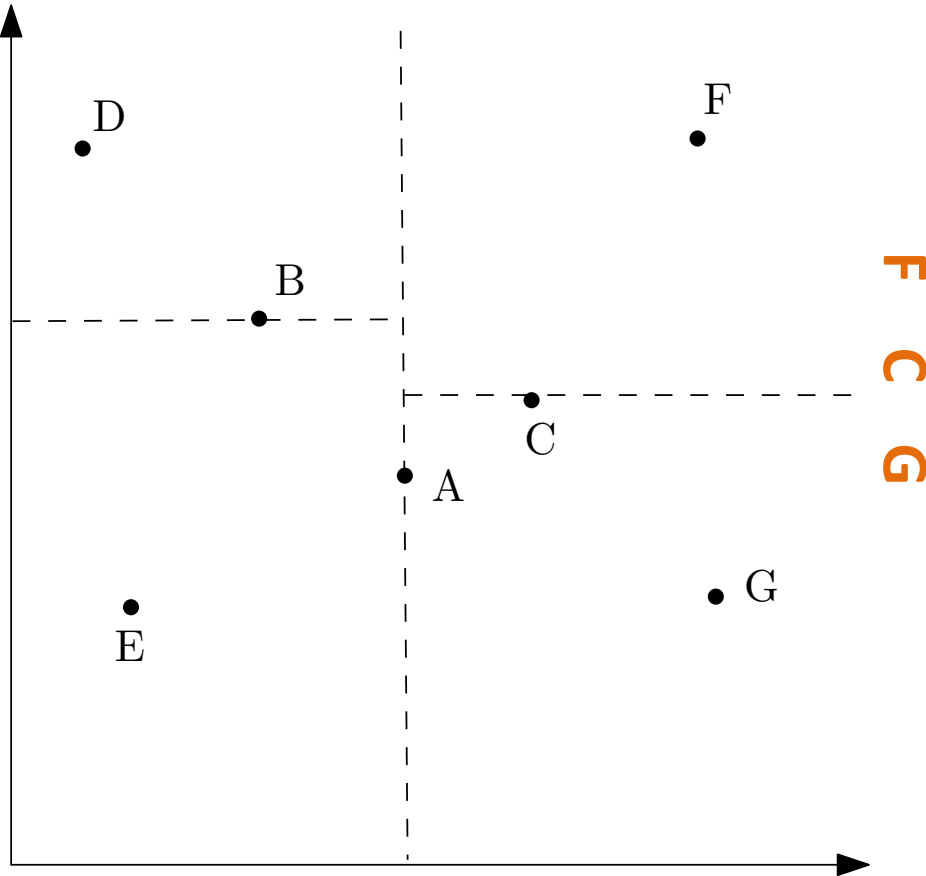
Handout 4, example 1

Recurse on the points below and above the median



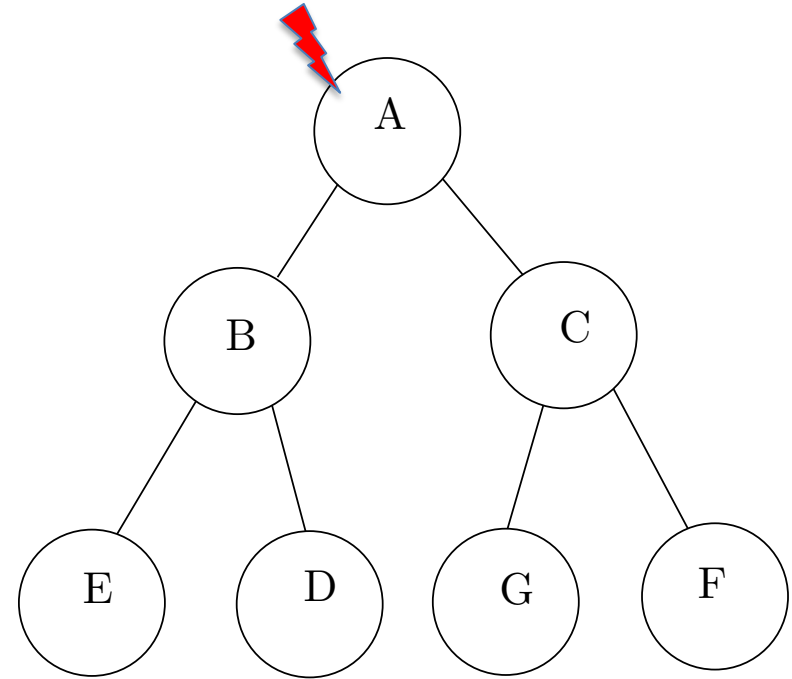
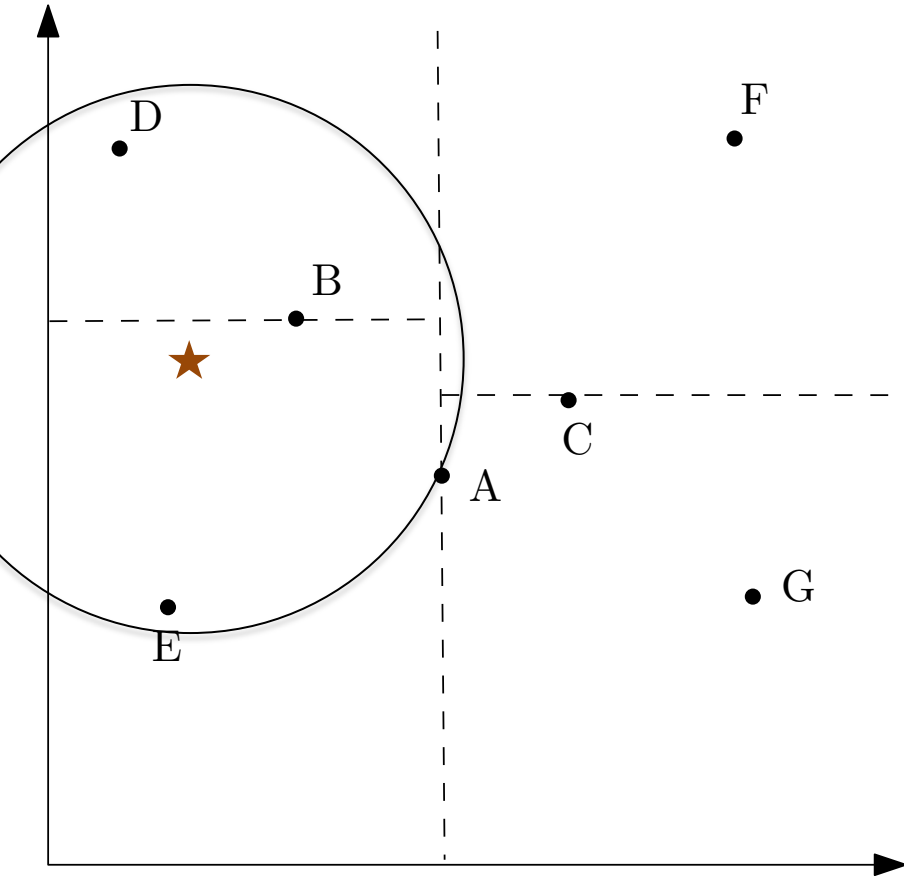
Handout 4, example 1

Recurse on the points below and above the median

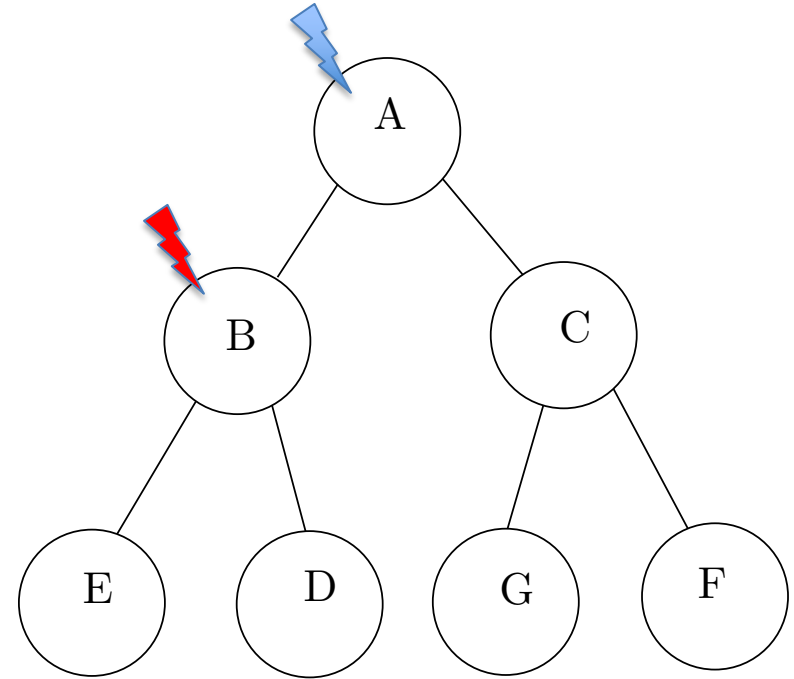
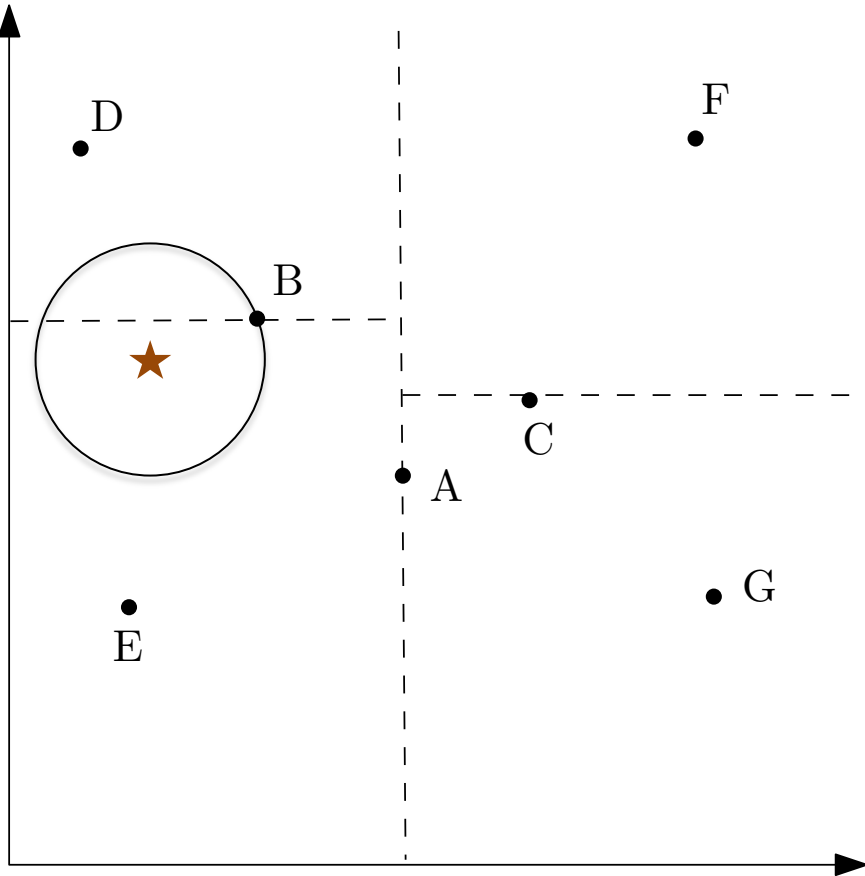


Handout 4, example 1

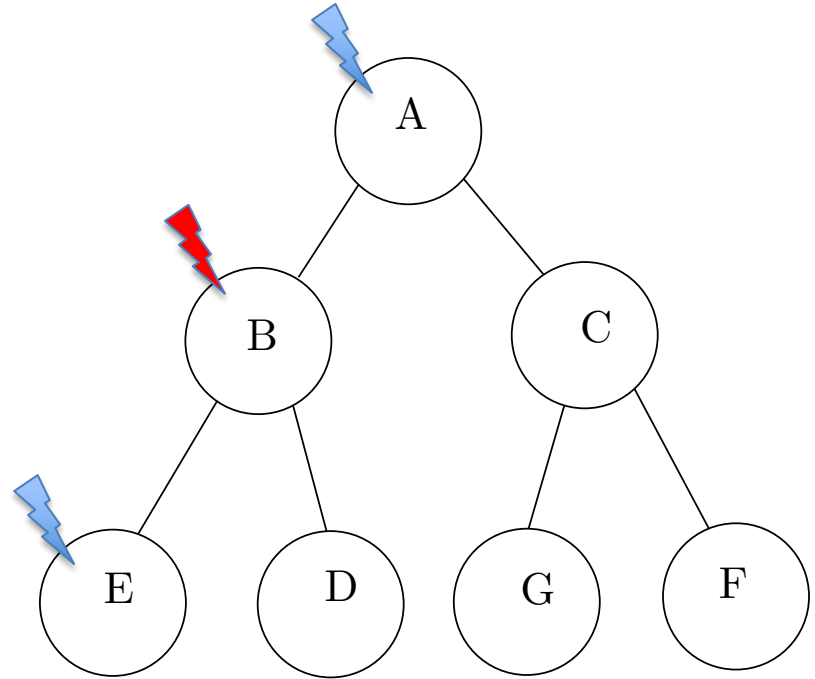
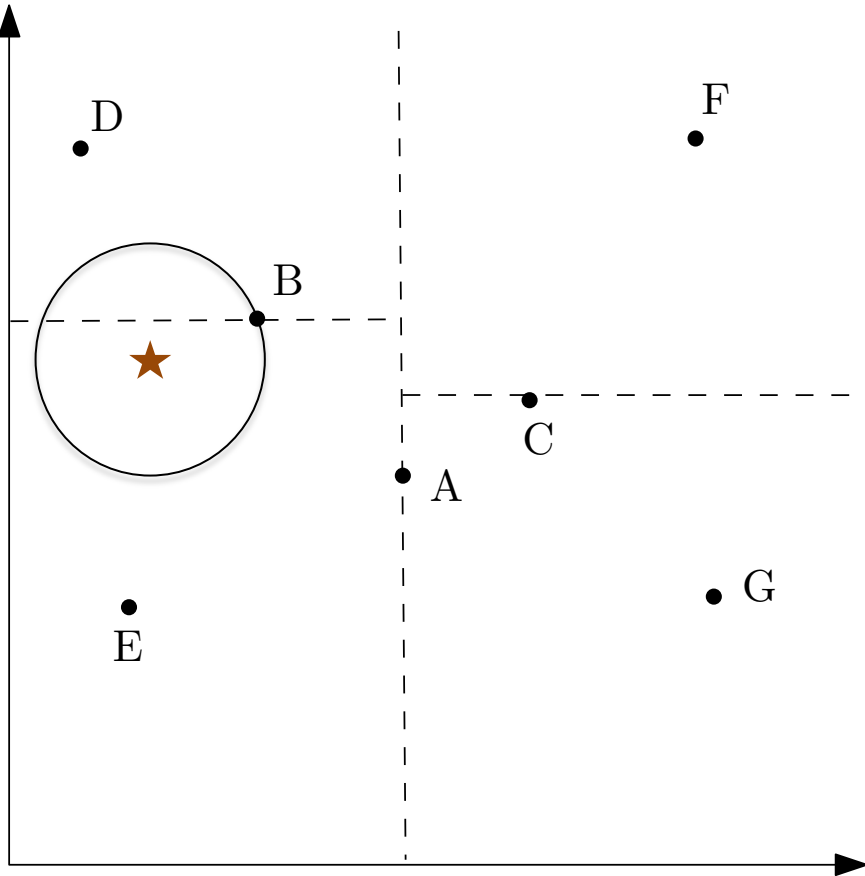
Test/query point: start at the root



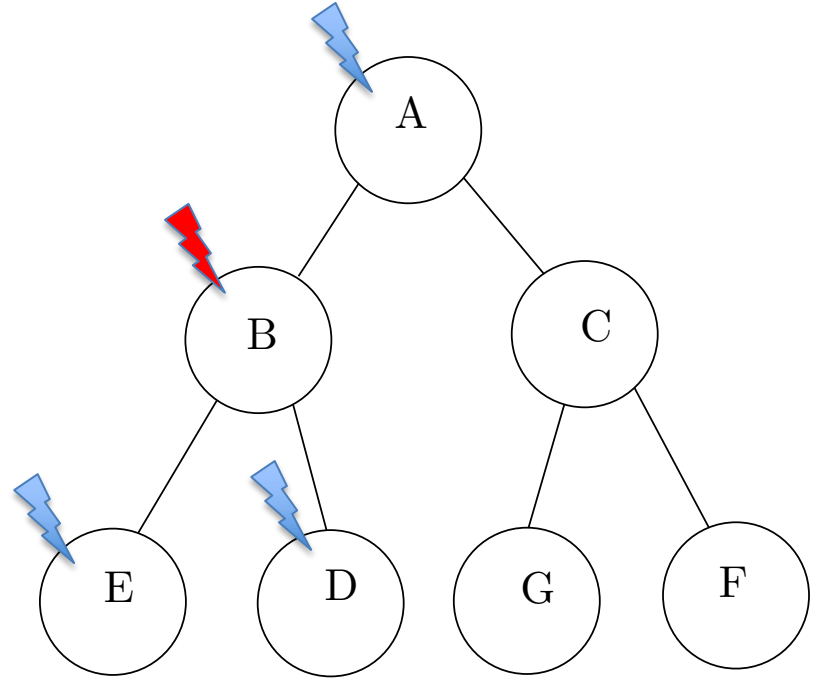
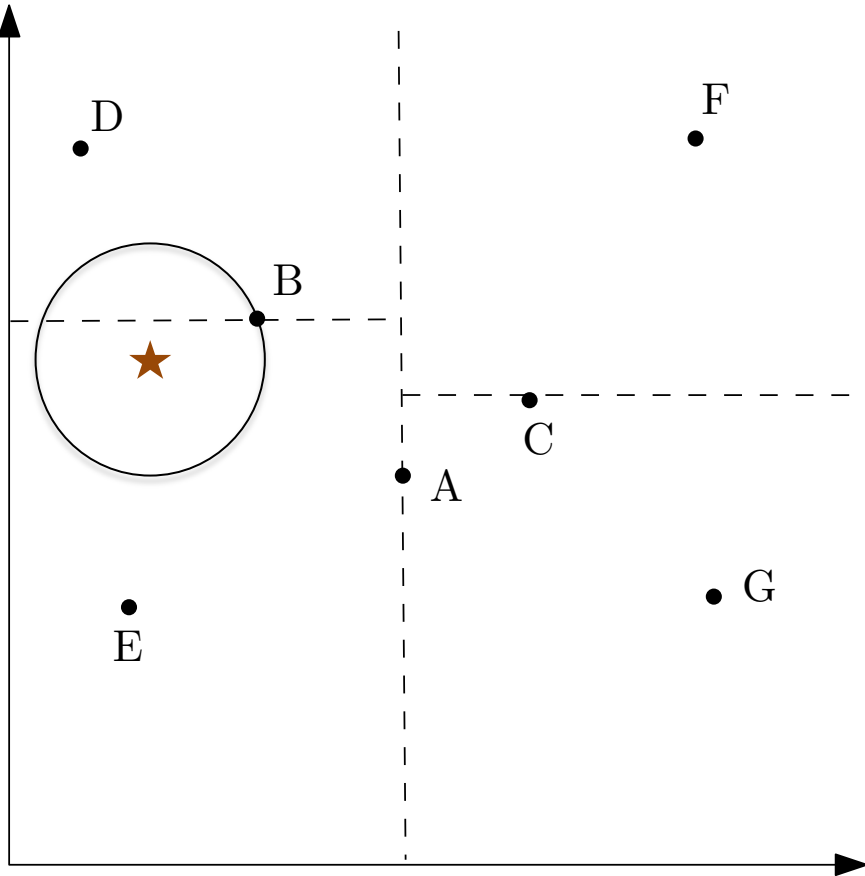
Handout 4, example 1



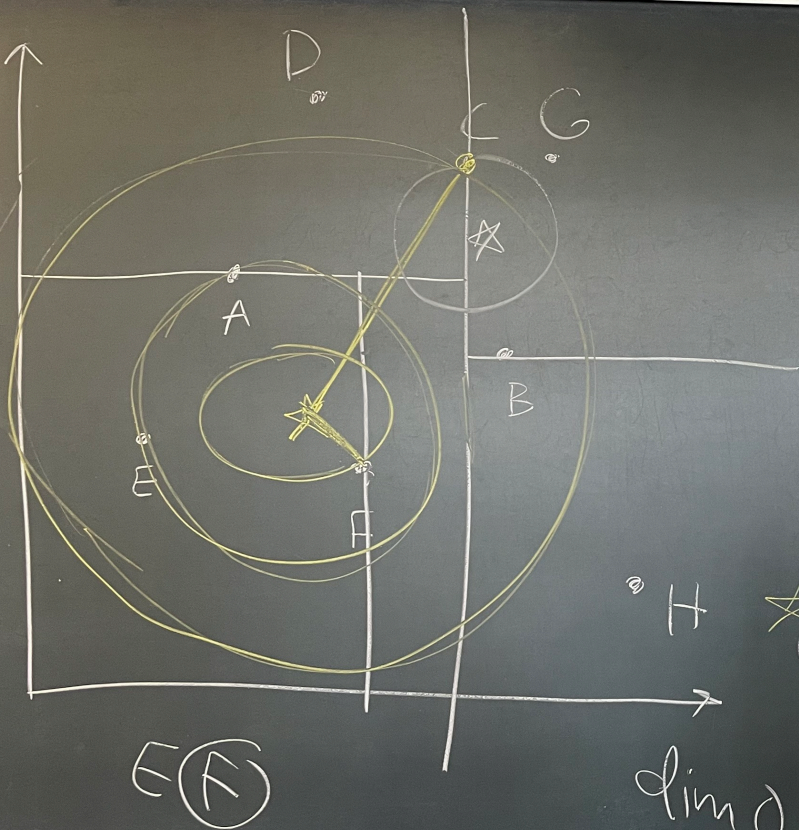
Handout 4, example 1



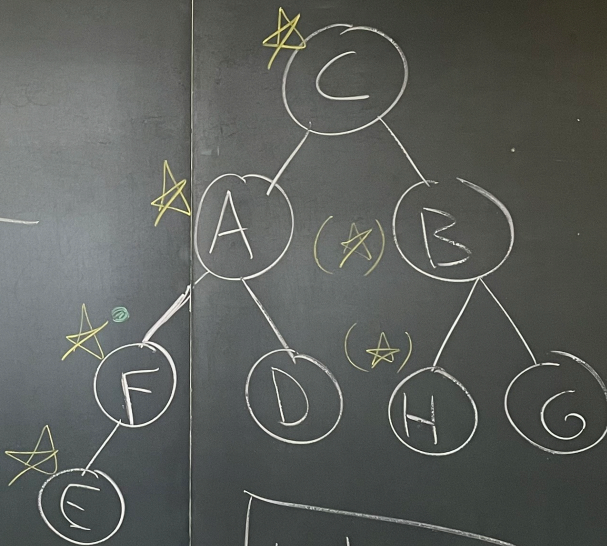
Handout 4, example 1



Handout 11: first page



① E A D F C B G H



dim 0

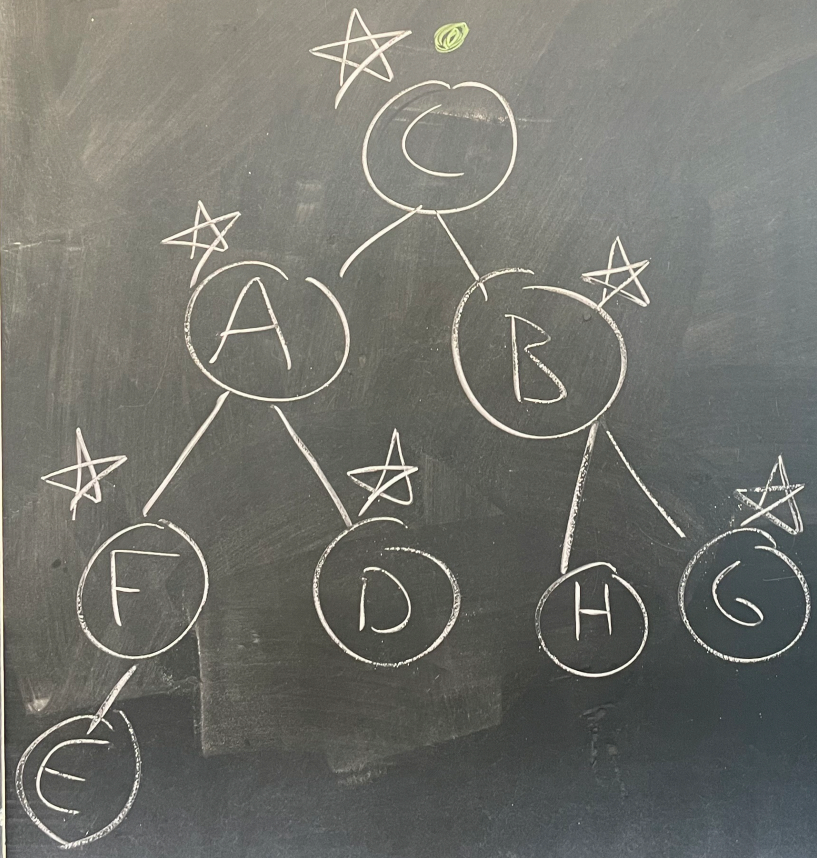
dim 1

dim 0

Balanced - depth $\log(n)$ dim 1

return F
+ min dist

dim



F E A D

regression: K neighbors
 \Rightarrow avg their y -values

classification: vote (including multi-class)

$K=5$

0 1 0 1
0 1

predict $\hat{y} = 0$

missing data: avg feature values of neighbors.

naive KNN

$O(nd)$
train \uparrow # dims \uparrow

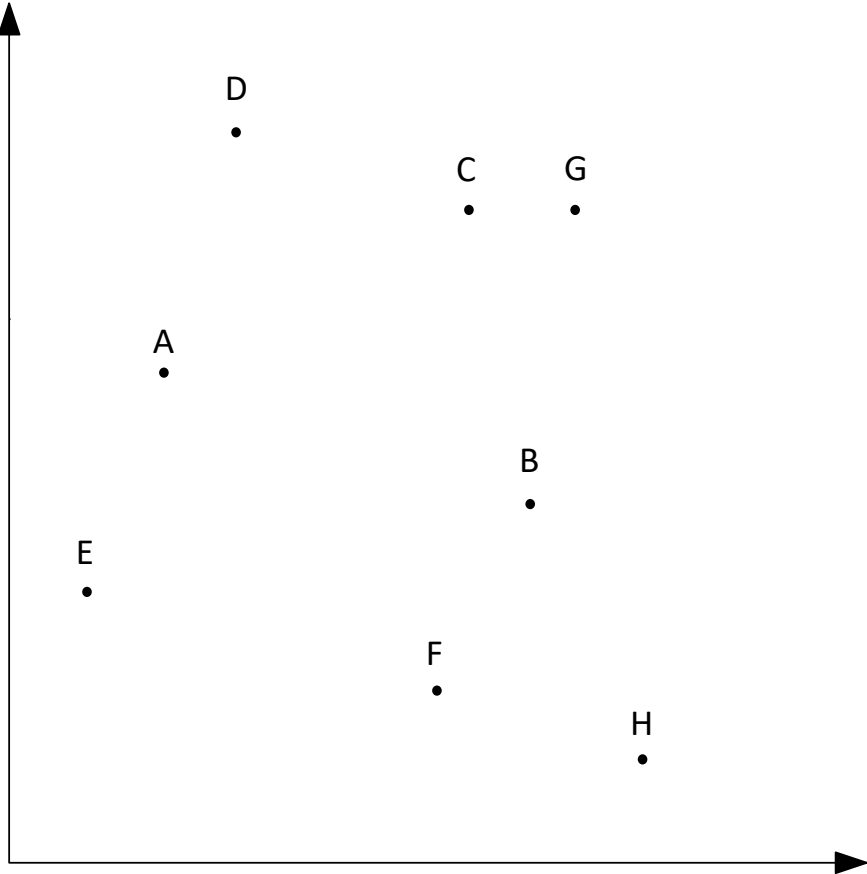
distance for one example = $O(d)$

\Rightarrow implementation Q?

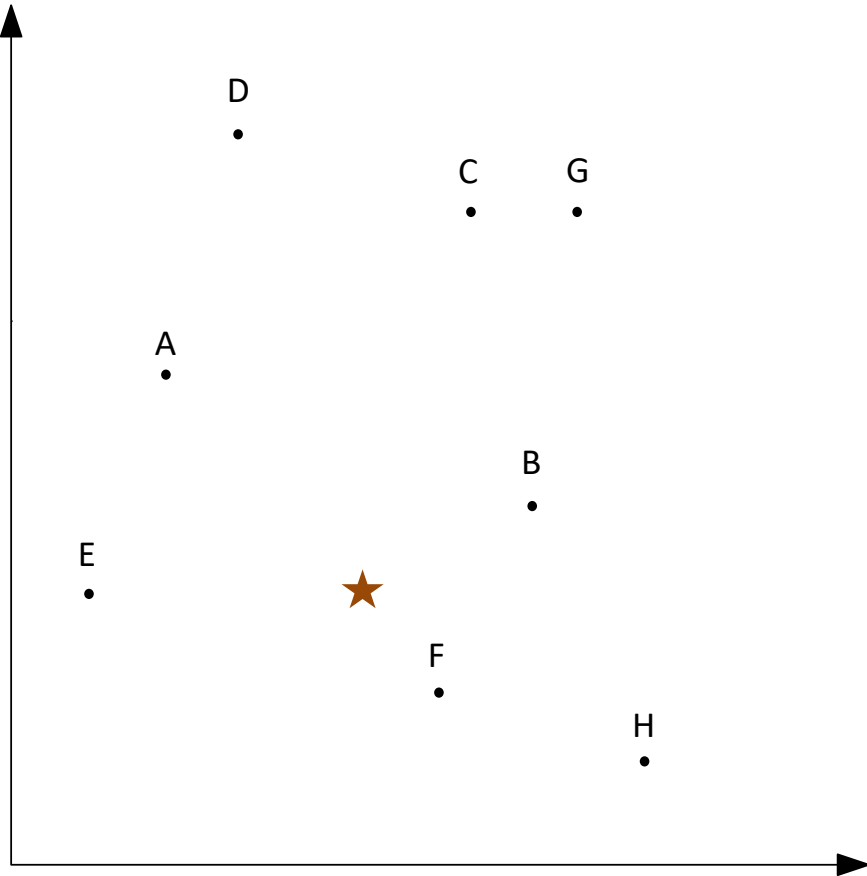
Sort vs. keep track

KD trees $O(d \log n)^*$

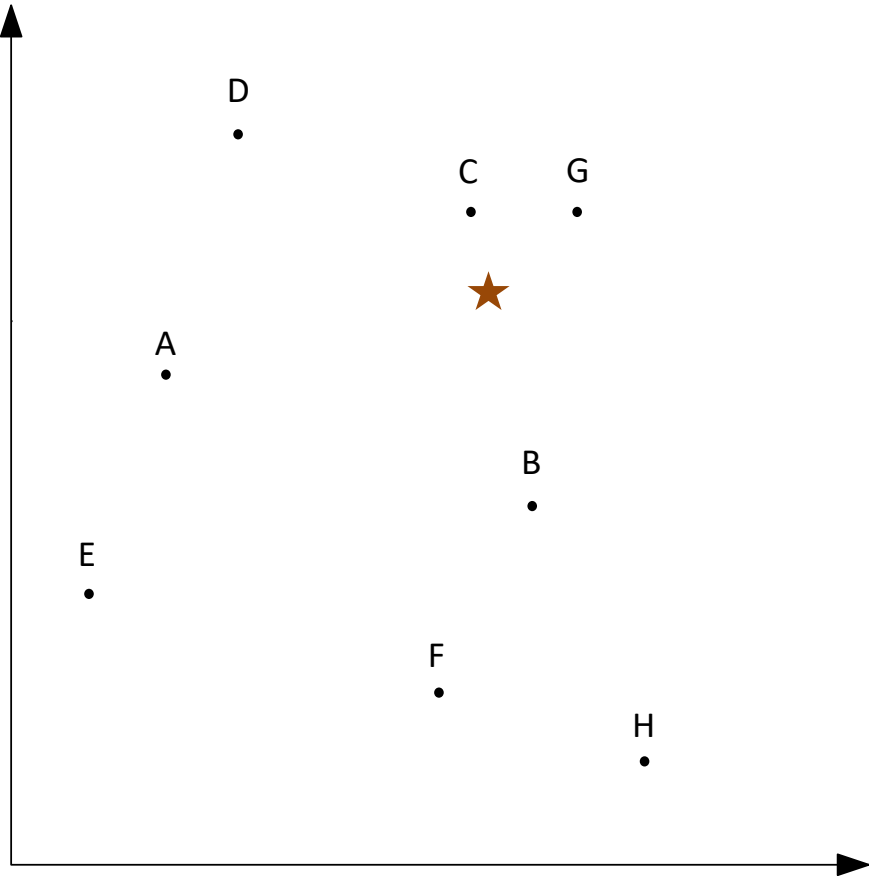
Handout 11: first page



Handout 11: query 1



Handout 11: query 2



Outline for Feb 27

- Correction for softmax
- Review K-nearest neighbors and KD trees
- **Review gradient descent and logistic regression**
- Thursday: decision trees, evaluation metrics, ML pipeline

$$\#2 \quad f(w) = w^2 - 6w + 9$$

$$f'(w) = 2w - 6 = 0$$

$$\alpha = 1$$

$$t = 1$$

$$w \leftarrow 0 - 1(2 \cdot 0 - 6)$$

$$w = 6$$

α derivative

$$t = 2$$

$$w \leftarrow 6 - 1(2 \cdot 6 - 6)$$

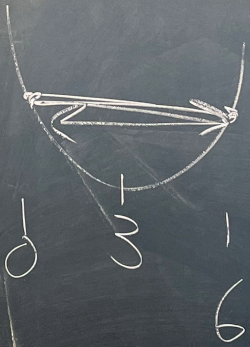
$$w \leftarrow 0$$

$$w = 3$$

$$w \leftarrow w - \alpha (\text{derivative})$$

(gradient)

$$\begin{bmatrix} -2 \\ 3 \end{bmatrix} - \alpha \begin{bmatrix} \nabla_{w_0} f \\ \nabla_{w_1} f \end{bmatrix}$$



$$\alpha = 0.1$$